Not by number alone: The effect of teachers' knowledge and its value in evaluating "sins of omission"

Ilona Bass (ibass@wesleyan.edu)[1], **Daniel Hawthorne-Madell (djthorne@stanford.edu)[2]**,
Noah D. Goodman (ngoodman@stanford.edu)[2], **Hyowon Gweon (hyo@stanford.edu)[2]**
[1] Department of Psychology, Wesleyan University, Middletown, CT 06459
[2] Department of Psychology, Stanford University, Stanford CA, USA

Abstract

What constitutes good teaching, and what factors do learners consider when evaluating teachers? Prior developmental work suggests that even young children accurately recognize and evaluate under-informativeness. Building on prior work, we propose a Bayesian model of teacher evaluation that infers the teacher's quality from how carefully he selected demonstrations given what he knew. We test the predictions of our model across 15 conditions in which participants saw a teacher who demonstrated all or a subset of functions of a novel device and rated his helpfulness. Our results suggest that human adults seamlessly integrate information about the number of functions taught, their values, as well as what the teacher knew, to make nuanced judgments about the quality of teaching; the quantitative pattern is well predicted by our model.

Keywords: pedagogy, Bayesian models, social learning, causal learning, pragmatics

Introduction

Learning from others is beneficial, but the benefits come with hazards. By learning from knowledgeable, helpful others, learners can draw powerful inferences that go beyond the face value of information. However, because its power hinges on the assumption that the teacher *is* knowledgeable and helpful, learning can go awry when teachers are inaccurate or misleading. Thus it is critical for learners to be sensitive to others' quality as teachers.

Reasoning in pedagogical contexts can be formally described as a set of mutually constraining inferences (Shafto, Goodman, & Griffiths, 2014). The teacher selects information that increases the learner's belief in the target hypothesis (pedagogical sampling), and the learner rationally updates his beliefs with the assumption that the information has been pedagogically sampled. This leads to a strong expectation that information provided by the teacher is not only true but also sufficient for the learner to draw accurate inferences. For instance, when a teacher pedagogically demonstrates just one function of a multi-function device, a naïve learner would (inaccurately) infer that the toy has one, and only one, function; if there were more, a knowledgeable and helpful informant would have shown them.

Prior developmental work suggests that young children draw strong inferences from pedagogically sampled information in ways that are consistent with the model predictions (Bonawitz et al., 2011; Xu & Tenenbaum, 2007), and appropriately respond to teachers who violate pedagogical sampling (Gweon, Pelton, Konopka, & Schulz, 2014). For instance, 6- and 7-year-olds rate a teacher as less helpful when he demonstrates one function of a multi-function toy (thus being under-informative), compared to when he demonstrates one function of a single-function (thus being fully informative) (Gweon, Pelton, et al., 2014), and given a binary choice, even four-year-olds favor a fully-informative teacher over an under-informative teacher (Gweon & Asaba, 2015; see also Barner, Brooks, & Bale, 2011 for similar inferences in linguistic communication). These results suggest that even early in life, human learners recognize and evaluate "sins of omission"–under-informative pedagogy that misleads the learner to draw an inaccurate inference.

Intuitively, not all sins of omissions are equally blameworthy. What affects our evaluations of under-informative teaching? One simple possibility is that the degree to which we penalize sins of omission is inversely related to the learner's degree of belief in the correct hypothesis. Thus the more information a teacher omitted, the more blame he would deserve. However, there are reasons to believe that people make more nuanced judgments about under-informative teaching.

First, going beyond the amount of omitted information, people might also consider the value of what was omitted. Because not all knowledge is equally useful or valuable, omission can have varying consequences for the learner depending on the utility of the resulting knowledge. Consider a gadget that has two buttons; one turns the gadget into an auto-navigating robot vacuum and the other activates a blinking light on the side. Because learning about the former would be more useful than the latter, showing just the vacuum function would be considered less blameworthy than showing just the blinking light.

Furthermore, people might be sensitive to the reason behind the omission (i.e., the teacher's intentions). In evaluating harmful actions, both children and adults are sensitive to the intent of a perpetrator and appropriately exonerate those who caused accidental harms due to their ignorance or false beliefs (Hebble, 1971; Yuill & Perner, 1988; Young, Cushman, Hauser, & Saxe, 2007). In linguistic communication, listeners flexibly adjust their interpretation of an utterance depending on the speaker's knowledge; if the speaker didn't have the knowledge to justify the stronger alternative, listeners don't draw implicatures (Goodman & Stuhlmüller, 2013). Thus even in the absence of explicit information about others' intentions, people spontaneously use information about the others' epistemic states to infer their intent. Similarly, in evaluating sins of omission, people might care about whether or not the teacher knowingly omitted something. Consider someone who demonstrated just one function of a four-function gadget, simply because he was unaware of the other three functions. Although he might be

independently blamed for his ignorance, we might exonerate him from being guilty of a "sin of omission" because his ignorance suggests that the omission was unintended.

In this study, we investigate whether people consider the amount and the value of information as well as the teacher's knowledge to evaluate under-informative pedagogy. Previous Bayesian models of pedagogical reasoning have been extremely useful in formalizing our intuitions about what an ideal learner might infer from pedagogically sampled data (Shafto et al., 2014) and in how learners might choose between different teachers. For instance, Shafto, Gweon, Fargen, and Schulz (2012) looked at people's evaluations of efficient vs. inefficient teaching, and modeled people's relative preferences between two teachers as a ratio of two likelihoods. However, these models assume that teachers are always fully knowledgeable and that all facts that could be taught are equally valuable. Furthermore, they do not provide explicit, systematic predictions about people's evaluations of teachers based on their knowledge.

Building on prior work, here we provide a computational model of teacher evaluation. The model captures the idea that good teachers will work harder to teach well, where "teaching well" is specified by the pedagogical sampling assumption. We extend the underlying pedagogical model to account for incomplete knowledge of the teacher as well as the value of information taught. To test the model predictions, we designed a behavioral task to get parametric evaluations of teachers' ability in which we manipulated the knowledge of the teacher, the value of the knowledge, and the amount of knowledge communicated to the learner.

Model

Reasoning about a pedagogical situation requires a notion of the causal relations between the world and the teacher's actions—an intuitive theory of pedagogy that captures the teacher's goal of informing the learner, and specifies what examples should be given to teach a particular fact or concept. This theory can be used in learning *from* teachers (as in Shafto et al., 2014), but also learning *about* the teachers themselves—the importance they place on informing the learner and their ability to do so by choosing effective examples of things that will be useful to the learner. To formalize a theory of pedagogy, we first describe how a teacher might choose what demonstrations to give in different situations. We do so with the standard softmax decision rule:

$$p(d \mid f, \alpha) \propto e^{\alpha U(d) - C(d)}, \tag{1}$$

where f is the set of functions known to the teacher, and d is the set of demonstrations he provides to the learner. $C(d)$, which we set equal to the number of demonstrations provided, corresponds to the cost of a given set of demonstrations, and $U(d)$ is the utility a learner is expected to accrue from the demonstrations (Eq. 2). The parameter α controls the degree to which the teacher chooses to maximize the pedagogical utility of his demonstration. Thus α can be interpreted as the teacher's "quality." An indifferent teacher

($\alpha = 0$) would choose demonstrations at random, preferring less costly demonstrations; as α increases, a teacher would be more likely to select demonstrations that have high pedagogical utility, as good teachers would.

For our model to predict how people *evaluate* the teacher, we must specify the utility of the teacher: what counts as successful teaching? Shato et al. suggested that the teacher's utility should be proportional to the log-probability that the learner will guess the correct hypothesis (2014). To apply this idea to cases where a device has different functions that vary in value, we must generalize this utility to capture the fact that there are different aspects of the device that the teacher could convey, each of which has its own utility to the learner. Thus we can define the pedagogical utility of a particular set of demonstrations (d) as:

$$U(d) = \sum_i V(f_i) \ln p_L(f_i \mid d_i) \tag{2}$$

where d_i is 1 if the ith function was presented and 0 otherwise, $p_L(f_i \mid d_i)$ is the teacher's model of the learner's beliefs about function i after demonstration d_i, and $V(f_i)$ is the value of function i. This pedagogical utility function is a generalization of previous accounts which examined the special case where all functions are equally valuable (effectively dropping $V(f_i)$ from the equation). Assuming that a demonstration of each function provides an equal amount of information to the learner, and that *a priori* belief in each function is the same, Eq. (2) becomes (up to an additive constant which does not impact decisions):

$$U(d) = \sum_i k d_i V(f_i), \tag{3}$$

where k is a constant reflecting the change in the learner's belief that a function exists, resulting from a demonstration (this constant will be absorbed into the overall calibration of utility below).

The demonstrations that a teacher selects using Eq. (1), which uses this pedagogical utility (Eq. 3), depends on the precise balance between the cost of demonstrations, the teacher's quality, and the value of each function. The demonstration cost pulls against the tendency of high-quality teachers to teach everything they know, such that a high-quality teacher with high communication costs may only teach the high value functions.

To use this theory of pedagogy to evaluate (rather than learn from) a teacher, consider a person who knows that a teacher knows functions f and observes the teacher's d demonstration(s). This observer can use Bayes' rule to invert their model of how teacher's select demonstration (Eq. (1)) to infer the teacher's quality:

$$p(\alpha \mid f, d) \propto p(d \mid f, \alpha) p(\alpha), \tag{4}$$

where $p(\alpha)$ represents the person's prior beliefs of the teacher's quality (we assume $p(\alpha) \sim \text{Uniform}(0, 1)$). The resulting estimates of teacher quality are sensitive to the teacher's epistemic state and the value of the functions

demonstrated. For example, the quality estimate of a teacher who demonstrates two low-value functions but also knew of a high value function will be lower than someone who just knew the two low-value functions he taught. Similarly, a teacher who knows both a high- and a low- value function would get the highest rating for showing both, a lower rating for omitting the low-value function, and an even lower rating for omitting the high-value function.

Experiment

Methods

Participants All participants were recruited from Amazon Mechanical Turk, and were paid $0.50 for compensation. We excluded a total of 462 participants from analyses based on a priori criteria (responses to check questions; see Procedure) and 71 for participating more than once. The final sample consisted of 1024 participants (Age: $\mu(\sigma) = 36.2(12.6)$, range: 18 - 72 yrs; 575 female).

Stimuli We generated cartoon scenarios in which one character (e.g., Paul) encounters a novel device and discovers all four, or a subset (one or two) of its functions, and then shows another character (e.g., Laura) all or a subset of its functions. The device had four distinctive buttons that corresponded to each function. Pressing a red circular button made the device verbally report the current time; the purple lightning-shaped button reported the local weather; the orange crescent-shaped button made the device say "hello!"; and the green square button generated a beep sound. The relative values of these functions were validated by a separate group of 52 mTurk participants who rank-ordered the four functions by their usefulness. Weather and Time ($\mu(\sigma) = 3.4(0.29)$) ranked higher than Beep and Hello ($\mu(\sigma) = 1.6(0.29); t(51) = 23.3 p < .001$), with no difference within the high-value pair or the low-value pair (p's$> .298$).

Design We varied the number of functions the teacher discovered (1, 2, or 4), the utility of the functions he discovered (H for high, L for low), and the number of functions he demonstrated to the learner (1, 2, or 4). Crossing these variables while excluding impossible cases yielded 15 conditions (see Figure 2A for a full list of conditions). For instance, in the KA_TA condition the teacher Knew All and Taught All; in KHL_TH, he knew two functions (one high-value and one low-value) but taught only one high-value function. The exact function taught among the two equally valued functions (e.g. Weather vs. Time) was counterbalanced throughout.

Procedure Participants were randomly assigned to one of the 15 conditions and were shown the cartoon scenario that corresponded to that condition. The first phase of the cartoons introduced the device to ensure that the participants knew about all the functions; the second phase showed the teacher discovering some or all of the device's functions. In the third phase, the learner entered the room[1] and asked the

---

[1]In conditions where the teacher did not know all the functions, the learner's entrance interrupted the teacher's discovery; this was to minimize the possibility that the participants' ratings were af-

Figure 1: Schematic of procedure in the KLL_TL condition. The teacher discovers two low-value functions and teaches just one low-value function.

teacher to show her how the device works. In the fourth phase, the teacher demonstrated some, or all, of its functions (see Figure 1). Participants then answered the critical question: 1. Overall, how would you rate his teaching abilities? We also asked additional questions in fixed order (2. Which functions did he discover? 3. Which functions did he teach? 4. How well-intentioned do you think he was? 5. How nice do you think he is? 6. Given what he knew, how good a job did he do? 7. How willing would you be to learn from him?). Finally we asked the first question again (8. Overall, how would you rate his teaching abilities?). We used a seven-point Likert scale for questions 1 and 4-8. Questions 2 and 3 were used as check questions to exclude participants who did not pay close attention.

Results and Discussion

Participants' ratings of teacher ability in the first and the last questions (Q1 and Q8: the teacher's overall teaching abilities) were highly correlated ($r(1021) = 0.833$, $p < .001$). We therefore used the average of the two ratings in our analyses. See Figure 2A for mean ratings of teacher ability in

---

fected by the teacher's incompetence or failure to discover all functions.

all conditions. We conducted a linear regression to examine the effect of (1) the number of functions demonstrated by the teacher, (2) the value of these functions, and (3) number of functions the teacher knew on participants' estimate of his ability. We coded these three factors as continuous variables (to code the effect of the value of the demonstrated functions as a continuous variable, we quantified it as the proportion of the high valued functions known to the teacher that were demonstrated). All three factors were highly significant predictors. The higher the number of demonstrations the higher the teacher was rated ($\beta = 1.1, t(796) = 17, p < .001$); the higher the value of his demonstrations, the higher he was rated ($\beta = .66, t(796) = 4.8, p < .001$); and the more functions the teacher knew, the lower he was rated ($\beta = -.3, t(796) = 17, p < .001$); that is, when the teacher did not know as many functions, people were more likely to exonerate him for his limited knowledge. We further investigated these main effects with targeted analyses of a subset of the conditions that highlights the predicted effects of (a) naïve omission, (b) the teacher's prior knowledge, and (c) the effect of prompting intention and knowledge.

a) Naïve omission We first looked at conditions in which the informant discovered only a subset of the device's functions, but taught everything he knew (KL_TL, KH_TH, KLL_TLL, KHH_THH). A 2 (number: 1 or 2) x 2 (value: H or L) ANOVA found a significant main effect of number ($F(1, 297) = 21.1, p < .001$). We also observed a marginally significant main effect of value ($F(1, 297) = 2.9, p = .088$), and the interaction was not significant ($F(1, 297) = .27, p = .603$). Thus even when the teacher communicated all he knew, participants' ratings still reflected the number of functions taught.

b) Informant's prior knowledge We then looked at conditions in which the informant taught two functions but either taught all he knew (KHH_THH, KLL_TLL) or a subset of what he knew (KA_THH, KA_TLL). A 2 (prior knowledge: Knew Two (Taught All) or Knew All (Omitted)) x 2 (value: HH, LL) ANOVA found a significant effect of Prior Knowledge ($F(1, 285) = 30.9, p < .001$): Even though the teacher showed the same number of functions, participants took into account the teacher's knowledge when rating his teaching abilities. See Figure 2B. We also saw a significant main effect of value ($F(1, 285) = 12.0, p = .001$) as well as an interaction ($F(1, 295) = 5.6, p = .019$). More specifically, the value of functions taught affected ratings only when the teacher taught just a subset of what he knew ($t(132) = 3.8, p < .001$); if he taught all he knew, the value had no effect ($t(153) = .084, p = .40$).

c) The effect of prompting intention and knowledge Our questionnaire had 8 questions, with the first and the last question serving as our main dependent measure. Between these two identical questions, people were asked a few check questions as well as questions that prompted them to think about the knowledge of the teacher and the intention of the teacher. Thus it is possible that people are more likely to consider others' knowledge after having been explicitly prompted to think about it; if so, answering these questions would lead to amplified differences in people's ratings of omissions. To test this idea we performed an exploratory analysis comparing conditions in which the informant always taught just one function, with his knowledge base ranging from one to four functions across conditions. We used the difference between the two "Overall" questions as the DV (as opposed to the mean) in a one-way ANOVA and found a significant main effect of prior knowledge ($F(2, 527) = 31.9, p < .001$). That is, the less the teacher knew, the more likely participants were to give a more generous rating for the teacher. When the teacher taught just one function but knew just one or two functions, people significantly increased their ratings relative to their initial rating (K1: $\mu_{diff}(\sigma) = 0.77(1.1), t(145) = 8.4, p < .001$; K2: $\mu_{diff}(\sigma) = 0.25(.95), t(256) = 4.1, p < 0.001$). Conversely, when the teacher knew all functions but just taught one function, participants gave a harsher rating in the final question relative to their initial rating ($\mu_{diff}(\sigma) = -.12(.63), t(126) = 2.1, p = .035$).

Model Fit The two free parameters in the model—the difference in the utility of knowing a high vs. a low value function, and the cost the teacher incurs by communicating a function—were fit to the mean of participants' estimates of the teacher's overall ability. As seen in Figure 2C, the resulting model well predicts these ability estimates (R^2=0.96).

General Discussion

Here we presented a formal Bayesian account of how a learner might evaluate a teacher. Our model considers the amount of information a teacher provided, the value of that information, and what he knew (but didn't show) to infer his "quality" as a teacher. Going beyond using the amount of communicated information as a proxy for evaluating teachers, our behavioral results showed that human learners spontaneously consider both the reason behind a teacher's omission (i.e., he didn't show because he didn't know) as well as the consequences for the learner (i.e., teaching X is more valuable for the learner than teaching Y). Participants appropriately penalized or exonerated a teacher, generating graded ratings that reflect the teacher's quality in ways that are consistent with our formal analysis.

Both knowledgeability and helpfulness are important traits of good teachers, but not everyone around us is equally knowledgeable and helpful. In fact, violation of pedagogical sampling occurs frequently in real-world learning contexts. For instance, a well-intended, helpful teacher might provide insufficient information because he didn't have all the relevant knowledge. A fully knowledgeable teacher might omit information because she thought it is not worth teaching. Furthermore, we have an intuitive sense that a "good" teacher is not just someone who provides everything he knows; it is someone who knows what's best for the learner and prioritizes teaching "what matters". Critically, what matters for the learner might not be the same as what matters for the teacher. Although our model does not yet capture all the
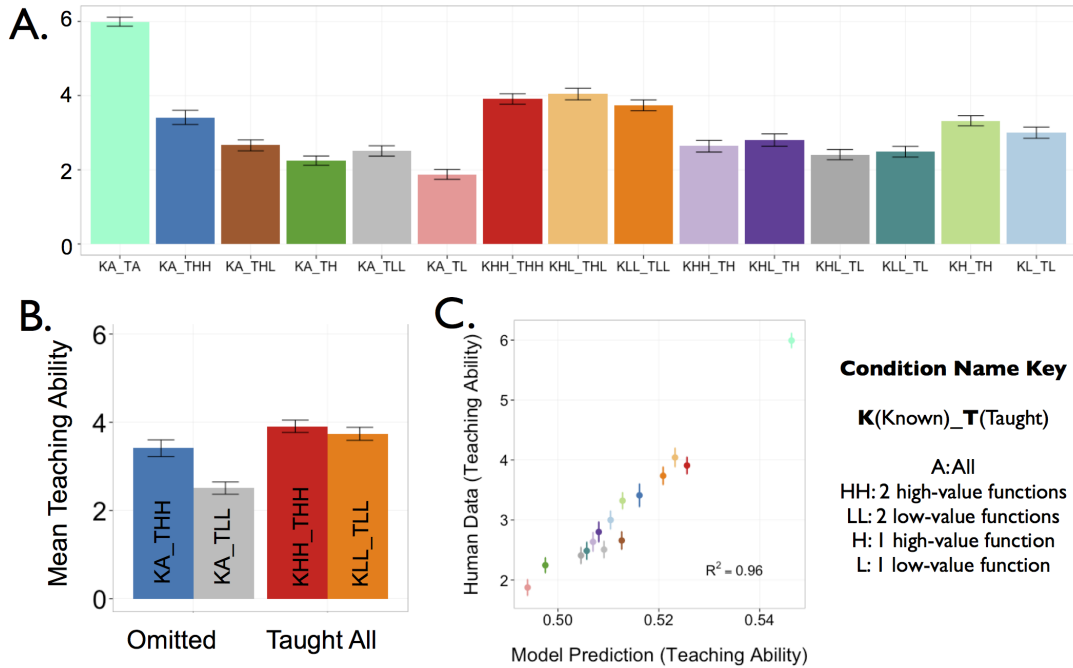
Figure 2: A. Mean teacher ability rating in all 15 conditions. B. The effect of value depends on whether or not the teacher knew more than he taught. C. Comparison of behavioral data and the model predictions. Each colored point corresponds to a bar in the same color in A. The fit between model and human data is $R^2 = 0.96$.

complexities and sophistication in gauging someone's quality as a teacher, it builds on, and importantly extends, prior computational work on pedagogical reasoning by assuming that human learners integrate different sources of information to distinguish potentially misleading teachers from those who are truly knowledgeable and helpful.

Our behavioral results were remarkably well predicted by the model predictions. When the teacher knew all functions of the device, people's evaluations were rationally affected by both the number and the value of the functions he taught; the more he showed, and the more valuable the functions that he showed, the higher people rated his teaching effectiveness. Furthermore, given two teachers who taught exactly the same functions, people's judgments reflected the knowledge of the teacher; the less he knew (thus the less he omitted), the more generous the ratings were. While we used the average of the first and the last questions for better reliability, these effects were robustly present even in people's responses to the first question ($R^2 = .93$). These results suggest that human adults spontaneously consider both the potential intent of the teacher's omission and its consequences for the learner to give nuanced, graded judgments about others' qualities as teachers even without being explicitly prompted about their intentions or knowledge.

One might wonder why we still observed an effect of number of demonstrated functions when the teacher taught all he knew (see Naïve Omission under Results). This is still reasonable, because the teacher's quality as judged by the

observer is still influenced by the pedagogical utility of the demonstration. Consistent with what is predicted by our model, in our behavioral task participants were asked to evaluate the teacher's overall helpfulness, rather than the blameworthiness of the omission per se.

In the current model, the teacher's quality was defined as the degree to which he considered pedagogical utility when choosing demonstrations, with the assumption that the learner would learn what was shown once. However, this global weighting represents just one of many dimensions on which teaching can be evaluated. Indeed, people's evaluations of teaching may be even more nuanced than our model presented here, and sensitive to factors that are left unexplored in our current work. For example, one of the hallmarks of a good teacher is sensitivity to the difficulty of the subject material for the learner (i.e., "this is a difficult concept, so I should show this example multiple times"), as well as the learner's epistemic states, learning style, and abilities (i.e., "she is quick so I shouldn't dwell on this", or "I'll skip what he already knows"). Thus in some cases, omissions might not only be forgivable but even desired, especially when transmitting information can be costly (see Gweon, Shafto, & Schulz, 2014 for children's sensitivity to learner's epistemic states and cost of information). These abilities correspond to rich knowledge of the subject domain and a more sophisticated model of learners' abilities and goals, both of which were simple in the current model.

Furthermore, we have assumed that both the learner and

the teacher have identical utility functions. Indeed, this is often not true in the real world, particularly in cases where young children or students learn from adult teachers; children are often taught what adults think is valuable for children, rather than what children they themselves think is valuable (e.g., the value of math or history classes). Even within learners and within teachers there are vast differences in how they value and prioritize different kinds of knowledge. These are subtle yet critically important considerations that go into designing curriculums in real-world educational settings, and where future computational and empirical work can be useful in formalizing what constitutes "good teaching."

In our current work, we have considered cases where a teacher omits functions with positive values. However, there are cases where omission leads to negative consequences (e.g., a building manager who forgot to announce 3-day water shutdown in your apartment). In fact, these are the kinds of omissions that can elicit the harshest evaluations, and perhaps the most indicative of informants who should be avoided in the future. Understanding the asymmetries in omitting positively-valued and negatively-valued information is an interesting topic for future work.

Recall that in our behavioral results, we observed a change in people's ratings between the first and the last questions (both of which were about his overall teaching abilities; see The effect of prompting intention and knowledge under Results). Between these two questions, people answered questions that prompted them to think about the teacher's intentions and his knowledge. Even though these prompts are not necessary for accurate evaluation, the effect of these prompts suggest that explicitly thinking about these factors can make the differences even more pronounced. One interesting possibility is that people's ability to think about others' beliefs (Theory of Mind) is directly related to the extent to which learners consider the teachers' knowledge and intentions. Thus one might predict that people with impaired or low Theory of Mind abilities might (a) show less consideration of the teacher's knowledge and intent, but (b) benefit more from explicit prompting of these factors. Follow-up research is under way to investigate these possibilities.

We note that although our behavioral experiments allowed a precise manipulation of what was taught and what the teacher knew, it leaves an open question about whether students in real-world pedagogical settings are also sensitive to these factors. Furthermore, our participants were third-party observers of a teacher-student interaction rather than the students themselves. An interesting extension of this work is to ask whether young learners, as students, also consider these factors to evaluate teachers in more realistic, live interactions with a teacher.

As learners in the real world, we often face a challenge of dealing with both uncertainty about the world as well as the uncertainty about others' qualities as teachers. How real-world learners exploit diverse sources of information, including their own exploration of the world, to simultaneously learn about both other people and about the world is

an incredibly rich and exciting area for future computational and empirical work. Our work provides an important step towards delineating the factors that we all consider in learning from whom to learn.

References

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, *5*(1), 173–184.

Gweon, H., & Asaba, M. (2015). Knowing what he could have shown: The role of alternatives in children's evaluation of under-informative teachers. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014, September). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*(3), 335–341.

Gweon, H., Shafto, P., & Schulz, L. E. (2014). Children consider prior knowledge and the cost of information both in learning from and teaching others. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*.

Hebble, P. W. (1971). The development of elementary school children's judgment of intent. *Child Development*, 1203–1215.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive psychology*, *71*, 55–89.

Shafto, P., Gweon, H., Fargen, C., & Schulz, L. (2012). Enough is enough: Inductive sufficiency guides learners ratings of informant helpfulness. In *Proceedings of the 34th annual conference of the cognitive science society*.

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in bayesian word learning. *Developmental science*, *10*(3), 288–297.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235.

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology*, *24*(3), 358.