UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Do You Know What I Know? Children Use Informants' Beliefs About Their Abilities to Calibrate Choices During Pedagogy

Permalink
https://escholarship.org/uc/item/2rm5z1fx

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Bass, Ilona
Mahaffey, Elise
Bonawitz, Elizabeth

Publication Date
2021

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Do You Know What I Know? Children Use Informants' Beliefs About Their Abilities to Calibrate Choices During Pedagogy

Ilona Bass (ilona.m.bass@rutgers.edu)[1,3], Elise Mahaffey (elise.mahaffey@rutgers.edu)[1], Elizabeth Bonawitz (elizabeth_bonawitz@gse.harvard.edu)[1,2]

[1] Department of Psychology, Rutgers University, Newark, NJ 07102

[2] Graduate School of Education, Harvard University, Cambridge, MA 02138

[3] Department of Psychology, Harvard University, Cambridge, MA 02138

Abstract

Models of pedagogy highlight the reciprocal reasoning underlying learner-teacher interactions, including that learners' inferences should be shaped by what they believe a teacher knows about them. Yet, little is known about how this influences learning, despite the fact that even young children make rapid inferences about teaching from sparse data. In the current work, six- to eight-year-olds' performance on a picture-matching game was either overestimated, underestimated, or accurately represented by a confederate (the "Teacher"), who then presented three new matching games of varying assessed difficulty (too easy, too hard, just right). A simple model of this problem predicts that while children should follow the recommendation of an accurate Teacher, learners should choose easier games when the Teacher overestimated their abilities, and harder games when she underestimated them. Results from our experiment support these predictions, providing insight into children's ability to consider teachers' knowledge when learning from pedagogy.

Keywords: Cognitive Development, Teaching, Evidence, Pedagogical Model

Introduction

From a young age, humans learn much about the world from the people around them (Csibra & Gergely, 2009; Koenig, Clément, & Harris, 2004). In order to communicate information effectively, informants must have a clear understanding of the learner's prior beliefs and abilities. To illustrate this point, imagine that the authors of this paper started by jumping straight to the results of our experiment, assuming that the reader already knew our research question and methods. This would render much of the resulting information difficult to interpret, and clearly inappropriate for this audience. On the other hand, it might be equally unhelpful for us to spend the majority of this introduction focusing on material that the reader likely already knows – for instance, providing a detailed review of classic fundamental work in psychology and cognitive science (e.g., see Bayes, 1763; Hume, 1748/1999; Karmiloff-Smith, 1996; Newell, Shaw, & Simon, 1958; Piaget, 1929). Without an accurate representation of the learner's prior knowledge and abilities, effective teaching is difficult. Here, we ask whether young learners share this intuition – and, if made aware that their teacher has inaccurate beliefs about their competence, whether they accordingly calibrate subsequent choices about attempted task difficulty.

Models of Pedagogy & Representing Others' Beliefs

To formally investigate the idea that good teachers should accurately represent their learner's prior knowledge, we appeal to computational models of cognition, which can help us characterize nuanced social reasoning processes both qualitatively and quantitatively. Particularly relevant here is the pedagogical model originally presented by Shafto and Goodman (2008). This account describes how recursive mental-state reasoning between a teacher and a learner results in the teacher sampling *pedagogically*, such that they select the set of evidence that should maximally increase the learner's belief in the target hypothesis. Specifically, the probability that the teacher should present the learner with a particular set of exemplars (or data, d), given that they are trying to communicate a target hypothesis h, is formalized as:

$$p(d|h)_{teacher} \propto (p(h|d)_{learner})^{\alpha} \quad (1)$$

where α controls the degree to which a teacher selects useful examples. Learners, in turn, update their beliefs following Bayesian inference, under the assumption that the data have been pedagogically sampled by a knowledgeable teacher:

$$p(h|d)_{learner} \propto p(d|h)_{teacher}\, p(h) \quad (2)$$

Under this model, then, pedagogy is understood as a set of recursive, mutually dependent inferences: The informant presents the evidence she believes will maximize the learner's belief in the target hypothesis; and the learner rationally updates her belief in that hypothesis, assuming that the informant sampled the evidence pedagogically. (See also Bonawitz & Shafto, 2016; Shafto, Goodman, & Frank, 2012; Shafto, Goodman, & Griffiths, 2014).

Eq. 1 thus implies that good informants should select the evidence that would be *most likely* to lead the learner to infer the correct hypothesis. But what that "optimal" set of exemplars actually is will critically depend on what a particular learner does or does not already know coming into a learning problem. To illustrate, consider one of the tasks used by Bass et al. (2019). Here, participants selected evidence to correct different learners' false beliefs about how a toy worked. The toy actually activated when a red block was placed on top of it; but one learner falsely believed that square blocks made the toy work, while another thought that circular blocks made it go. The evidence that would be most effective in communicating how the toy really worked would be different for the first learner (who would need to see red, non-square blocks) than it would be for the second learner (who would need to see red, non-circular blocks). Indeed, Bass et al. (2019) found that preschool-aged children are capable of selecting the particular evidence that would correct different learners' false

beliefs. Thus, young children appear to be at least tacitly aware that learners' prior knowledge affects the evidence that would be most likely to help them infer the target hypothesis.

Conversely, then, if an informant *misrepresents* what the learner's prior knowledge actually is, they may select evidence that is sub-optimal for that particular learner. While many have investigated the nature of children's evaluations and assumptions in pedagogical reasoning (e.g., Bass, Bonawitz, & Gweon, 2017; Bonawitz et al., 2011; Gweon, Pelton, Konopka, & Schulz, 2014; Yu, Landrum, Bonawitz, & Shafto, 2018), much of this past work has been predicated on the default assumptions of the pedagogical model – namely, that the teacher accurately represents the learner's prior knowledge. What are the consequences for learning when this assumption is violated? Recent work begins to shed light on this question: Yang and Shafto (2017) computationally compared the efficacy of pedagogy (i.e., a knowledgeable teacher presenting helpful evidence) and active learning (i.e., a naïve learner exploring the evidence that would maximize information gain) under circumstances in which the teacher and learner were *conceptually misaligned* to varying degrees. Here, conceptual misalignment reflected circumstances in which either the learner's or the teacher's hypothesis space did not reflect the true state of the world (i.e., $H_W = H_T \neq H_L$, or $H_W = H_L \neq H_T$). They found that when there was no conceptual misalignment ($H_W = H_T = H_L$), learning from pedagogy was generally superior to active learning. However, this benefit diminished when the learner's concept space was incorrect – which necessarily involved the teacher generating examples that would have been useful for a learner with an accurate hypothesis space. But because the learner's hypothesis space was *not* accurate, these examples were less helpful in leading that learner to the correct solution. This past work thus suggests that learning may indeed be hindered when the evidence selector has misrepresented the learner's prior knowledge. Critically, Yang and Shafto (2017) examined cases in which the teacher and learner simply had different hypothesis spaces with respect to the true state of the world; whether the teacher's beliefs about the learner's knowledge might shift the learner's expectations about how that teacher will sample data for them remains an open question.

Preliminary Evidence from Prior Work

Misrepresenting a learner's prior knowledge could thus impede effective pedagogy. However, prior work also tells us that young learners are quite flexible, and may engage in compensatory actions to maximize learning despite sub-optimal teaching (e.g., Gweon et al., 2014). If made aware of a conceptual misalignment, might learners adjust their own actions taken during subsequent exploration, compensating for pedagogy that was generated using an inaccurate representation of their knowledge and abilities? Bonawitz et al. (2011) indirectly tested this question. Here, preschool-aged children observed an experimenter pedagogically demonstrate one function on a novel toy (which actually had four functions) in one of three conditions: they were either directly shown this demonstration by the experimenter, or they *observed* the experimenter provide this demonstration to another child, or to an adult. Children then explored the toy themselves. When they were the direct recipients of pedagogy, children were more likely to restrict their exploration to the demonstrated function – which is in line with the predictions of the pedagogical model. Patterns of play were similar when children observed pedagogy provided to another child; but when children observed pedagogy provided to an adult, they were more likely to explore non-demonstrated features of the toy (Bonawitz et al., 2011).

Why might we see this selective interpretation of pedagogical sampling? One explanation is that children understood that the evidence the teacher would select for another child might be similar to what she would select for the participant; however, this assumption would not extend to sampling for an adult, given the vast differences in their prior knowledge. Therefore, children may attend to whether a teacher's demonstration is likely to be relevant to them. However, these effects raise questions about how exactly children's inferences are shaped by what they believe a teacher knows about them – which has yet to be directly empirically tested.

Current Work

The current work asks whether young learners use information about the correctness of their teacher's beliefs about them to contextualize pedagogical actions and optimize their own choices. In particular, here we investigate how a teacher's representation of the learner's *competence* could shape learners' decisions about attempted task difficulty when learning from that teacher in the future. This is in some ways different from the concept and causal learning research in which investigations of the pedagogical model are often situated. However, notions of competence and knowledge may be tightly related in the context of pedagogy. For instance, past work has found that when acting as teachers themselves, preschoolers calibrate their own pedagogical demonstrations to their learner's competence (Gweon & Schulz, 2019), similarly to how they do so in light of a learner's beliefs (Bass et al., 2019). Further, just as instruction can constrain exploration in the context of novel toy exploration tasks (Bonawitz et al., 2011), 18-month-olds have been shown to rationally integrate information about their own abilities with parental feedback about task difficulty, constraining their motor actions accordingly (Tamis-LeMonda et al., 2008).

We suggest that if a teacher's representation *underestimates* a rational learner's competence, that learner will independently seek out additional challenge in a subsequent learning task. Conversely, this kind of challenge-seeking might be curbed if a teacher instead *overestimates* that learner's abilities. We present a causal model that formalizes these intuitions. In our behavioral experiment, we test our predictions with a group of six- to eight-year-old children. It is possible that children as young as four might be sensitive to subtle cues about others' knowledge in pedagogy (Bass et al., 2019; Bonawitz et al., 2011). However, our task

may prove difficult for younger children for two main reasons: First, explicit second-order false belief reasoning abilities – which may be recruited when considering another's false belief about one's own mental state – are still developing through the later preschool years (Wellman, Cross, & Watson, 2001); second, the added attentional demands associated with running this task remotely via Zoom due to the COVID-19 pandemic might make this task more difficult for a younger sample.

Method

In order to investigate how children may use the accuracy of an informant's beliefs about their abilities to calibrate learning from that informant's future demonstrations, we designed a task in which a confederate (the "Teacher") overestimates, underestimates, or accurately represents participants' performance on a picture-matching game (between-subjects). Using their prior "knowledge" of the participant's ability, this Teacher then presents three new matching games and evaluates them as much too difficult, not difficult enough, or just right for the participant; children then rank their preferences for which of these new games they want to play.

Participants

Methods, participant Ns, and analyses were pre-registered.[1] We recruited 20 participants per condition such that our final sample included $N = 60$ six- to eight-year-olds ($M(SD)_{age}$ = $88.1(9.58)$ months, range $= 72 - 108$ months; $N = 29$ female). An additional 3 children were dropped and replaced due to failure to pass comprehension check questions ($N = 2$; see Procedure), or asking to terminate the study early ($N = 1$).

Procedure

See Figure 1 for a schematic of our procedure.

Introduction to Teacher This task was administered over Zoom, with stimuli displayed to participants using Open Broadcaster Software. After signing onto the call with the Experimenter, children were introduced to the "Teacher" – who, children were told, would be joining the call periodically throughout the activity. Unbeknownst to participants, the Teacher was actually a pre-recorded video. To explain away the Teacher's inability to interact with the participant, it was established that she was experiencing a technical glitch such that she could hear the Experimenter, but not the participant. The Teacher then (temporarily) signed off the call.

Novel Object Naming Because the task would later involve the Teacher either accurately or inaccurately representing the learner's abilities, we wanted to avoid the possibility that children would be more likely to follow the Teacher's recommendation in the Accurate condition simply because she was a more reliable informant; rather, we wanted to ensure that any observed differences between conditions could be more readily attributed to a specific difference in the Teacher's knowledge *about* the participant's competence. To this end, the Experimenter familiarized the participant with the names of two novel objects; then, she invited the Teacher back into the call, and asked her to name the objects. In the conditions in which the Teacher would later overestimate or underestimate the child's abilities, she correctly labeled one of the two novel objects. In the Accurate condition, she incorrectly labeled both objects. Thus, in all three conditions, the Teacher was correct on one of three total judgments throughout the task. We opted to control for overall accuracy using a novel object naming paradigm because a wealth of prior work suggests that children track this information, and use it to infer informant reliability (e.g., see Harris, Koenig, Corriveau, & Jaswal, 2018 for a review). This manipulation also happened to be particularly simple to implement: The pre-recorded video of the Teacher naming the objects was identical across conditions, and the Experimenter simply provided the child with different "correct" labels.

Experimenter's Matching Games With the Teacher off the call again, children were then shown how to play the matching game task. This was a simple memory matching game: Children were shown a target picture and an array of cards with different colored shapes on them. This array of cards was then briefly flipped over, revealing pictures on their undersides. After flipping back over, children indicated the color and shape of the card they thought the match was under. Once the child identified the correct location for the match, the experimenter moved the target picture to a "counter", which tracked the matches the child was able to complete.

After playing a practice game and displaying sufficient understanding of the task, children were told they would be playing the Experimenter's two matching games: first an easy set, and then a hard set. Their goal would be to find as many matches as possible before time ran out. In all conditions, the task was controlled such that all children finished the easy set, and no children completed any matches on the hard set.

Teacher Forms Representation The Experimenter then pulled up a display of participants' results on the screen. In the Accurate condition, this display correctly showed 1/2 matching sets completed, and the Experimenter reiterated the child's actual performance ("The screen shows that you were able to finish the easy set and not the hard set!"). In the Overestimate condition, the Experimenter "accidentally went to the answers screen", and the display showed 2/2 sets completed. In the Underestimate condition, the Experimenter "accidentally reset the counter", and the display showed 0/2 sets completed. The Teacher then re-entered the call and saw the display; she said she was about to pick out some new match-

[1] The pre-registration for this experiment can be found at: http://aspredicted.org/blind.php?x=bm68j7. In this pre-registration, we neglected to re-run our power analysis after changing our primary dependent measure from a categorical variable to a continuous variable. Therefore, although the pre-registration denotes our intention to run 30 participants in each condition, this was in error; 20 participants per condition should be sufficient for our anticipated effect sizes.

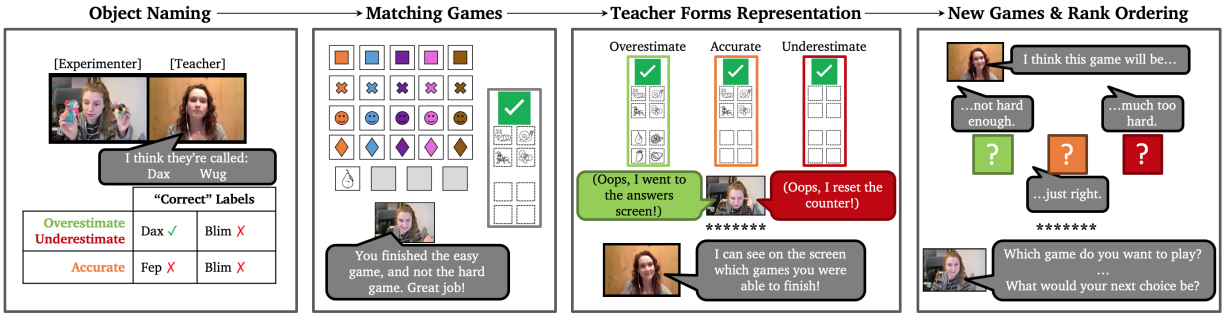

Figure 1: A schematic of the method used in our experiment.

ing games to play, and would keep the child's performance in mind when picking these new games. Due to the Teacher's previously established inability to hear the participant, children were not able to correct the Teacher if she saw an inaccurate display.

New Games & Rank Ordering The Teacher then returned with three new matching games, each of which would be worth a different number of points. Based on her purported understanding of the child's performance on the Experimenter's matching games, the Teacher thought one of these games would be "just right" for the child, and each correct match would be worth 4 points. Correct matches on the game that she thought "might not be hard enough" would be worth 3 points each; and for the game that she thought would be "much too hard", matches would be worth 5 points each. The participant was made aware that the Teacher would return later to see how many points they had earned. With the Teacher again off the call, the Experimenter then asked the child which game they wanted to play, given that they would only have time to play one of the Teacher's new games. After the child made their initial choice, the experimenter feigned an inability to open the game and asked participants what their next choice would be if she couldn't get their first choice to work. By doing this, the experimenter elicited a natural ranked choice of the participant's game preferences.

Comprehension Checks After playing their selected matching game (which was identical for all children, regardless of which game they chose) with the Experimenter, and then saying goodbye to the Teacher, participants were asked to recall: 1) how many of the Experimenter's matching sets the Teacher thought they were able to complete; and 2) how many of the novel objects the Teacher correctly named. Children who were not able to answer these questions correctly after two attempts and a forced-choice follow-up were dropped and replaced prior to analysis.

Coding

Game choices were assigned different values (too easy = 1, just right = 2, too hard = 3), and rankings were assigned different weights (first choice = 2, second choice = 1, third choice = 0). From this, we calculated a "difficulty score", which was simply the weighted rank-ordering of children's choices (e.g., a first choice of the "just right" game and a second choice of the "too hard" game implies a difficulty score of $(2(2)+1(3))/3 = 2.33$). We computed this continuous score as opposed to looking only at children's first choices to allow for more sensitivity in our measure; we suspected from past work on epistemic trust that children might have a bias to follow the teacher's recommendation across conditions (a pattern that was also supported by pilot data).

Model

In what follows, we present a simple causal framework for the problem participants might be solving in this task. Although we do not test all of the variables in this model here, this kind of formal causal framework is useful in formalizing our predictions, and could also serve to illuminate additional empirical questions that could be tested in future work.

When choosing to play one of the three presented games, participants might compute something akin to the *expected value* of playing each of the three games. This can be thought of simply as the maximum number of points (or reward, R) that could be earned by playing a particular game, weighted by the child's belief $p(S_g)$ that they will be able to successfully complete that particular matching game g:

$$\mathbb{E}(g) = R_g p(S_g) \tag{3}$$

Importantly, to calculate $p(S_g)$, the learner needs to estimate the actual objective difficulty of each game, which is unobserved. Therefore, this must be inferred from the available evidence – namely, the Teacher's statement about how difficult she thinks the game will be for the child, and the Teacher's beliefs about the learner's competence.[2] (See Figure 2 for a causal graphical model of this inference problem.)

Thus, the actual difficulty of each game (as known by the Teacher) and the Teacher's beliefs about the learner's competence will mutually inform the statement the Teacher makes

[2]In our model, $p(S_g)$ is also dependent on the Learner's beliefs about her *own* competence. We conceptualize this as the child's performance on the Experimenter's matching games, which was controlled across all participants in our task. Therefore, we do not consider this node further here – but see the Discussion for why it may be critical to probe this variable in future work.

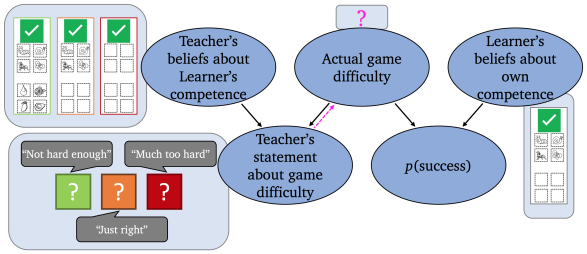

Figure 2: A causal graphical model of the inference problem in our task. In order to estimate the probability of success on a particular game, the learner must infer the true level of game difficulty that would be most likely to lead the Teacher's evaluation of how difficult that game would be for the learner, given her beliefs about the learner's competence.

about how difficult she believes each game will be for the learner. In particular, the model captures the intuition that if the actual game difficulty is approximately equal to the Teacher's subjective appraisal of the learner's abilities, the Teacher will assess that game as being "just right". On the other hand, if the actual game difficulty exceeds what the Teacher believes the learner is capable of, she will evaluate the game as being "much too difficult"; conversely, if the actual game difficulty level is lower than what the Teacher has represented as the learner's competence, she will instead judge that game to be "not hard enough".[3]

Results

In line with our pre-registered analysis plan, we first conducted a one-way ANOVA on the difficulty score for children's preferred games. This ANOVA was significant: $F(2,57) = 13.02$, $p < 0.0001$, $\eta^2 = 0.314$. Follow-up Bonferroni-corrected pairwise comparisons further revealed significant differences between the Overestimate condition and the other two conditions. Specifically, when the Teacher overestimated their competence, children selected easier games ($M = 1.58, SD = 0.40$) than when she had been accurate ($M = 1.97$, $SD = 0.42$; $p < 0.05$) or when she had underestimated their abilities ($M = 2.3$, $SD = 0.51$; $p < 0.01$). See Figure 3 for a summary of these results.

We also discretely analyzed children's first choice of game across conditions. A 3x3 Fisher's exact test on these choices was significant ($p < 0.0001$). To follow up on this result, we compared the modal response in each condition to chance (1/3) by exact binomial test. A majority of children in the Accurate condition chose the "just right" (16/20, one-tailed $p < 0.0001$). In the Overestimate condition, 10/20 children picked the "just right" game (one-tailed $p = 0.092$); this was closely followed by the predicted "too easy" game (9/20), although this latter proportion did not differ from chance. In contrast, the most common choice in the Underestimate condition was the "too hard" game (11/20, one-tailed $p = 0.038$). See Figure 4 for children's ranked choices across conditions.

[3]The specific quantitative formalization of this can be further specified, but the above details are sufficient to capture the pattern of qualitative results presented here.

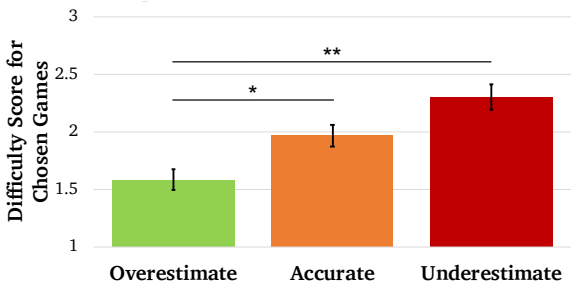

Figure 3: The average difficulty scores for children's selected games across conditions. Children preferred easier games when the Teacher overestimated their knowledge, and harder games when she underestimated their knowledge, while those in the Accurate condition fell in between. $*p < 0.05$; $**p < 0.01$. Error bars represent (+/-) the standard error.

Together, these results are broadly in line with our predictions, suggesting that children's choices were directionally informed by what they thought the informant believed about their abilities.

Discussion

We teach and learn from one another every day, yet this is no simple cognitive feat. If informants are to select helpful evidence, they must consider what exactly their learner already knows. And for learners to be able to interpret the meaning of presented evidence, they must consider whether their informant is sampling appropriately for *them*. Can children use a teacher's beliefs about their abilities to directionally adjust their learning from that teacher's future demonstrations? The current work suggests the answer to this question may be "yes". When a teacher correctly understood their competence, six- to eight-year-old children were happy to play a game that she thought would be about right for them. However, if she overestimated their knowledge, children compensated by choosing to play games that the teacher thought might not be challenging enough for them; in contrast, when she underestimated them, children instead preferred games that she thought would be too difficult. Thus, beyond a mere sensitivity to an informant's beliefs about them, children may use this information to rationally calibrate choices when learning from that informant in the future.

This work probes a previously untested prediction that falls out of the original pedagogical model (Shafto & Goodman, 2008; Shafto et al., 2014), and provides insight into the processes that facilitate learning from sub-optimal evidence selections. While we propose a general model outlining these inferences, we have yet to mathematically specify the underlying distributions in this model, and there are many more specific predictions that could be tested. For instance, we suggest that learners may gauge the probability of succeeding at a task using both the difficulty of that task and their own

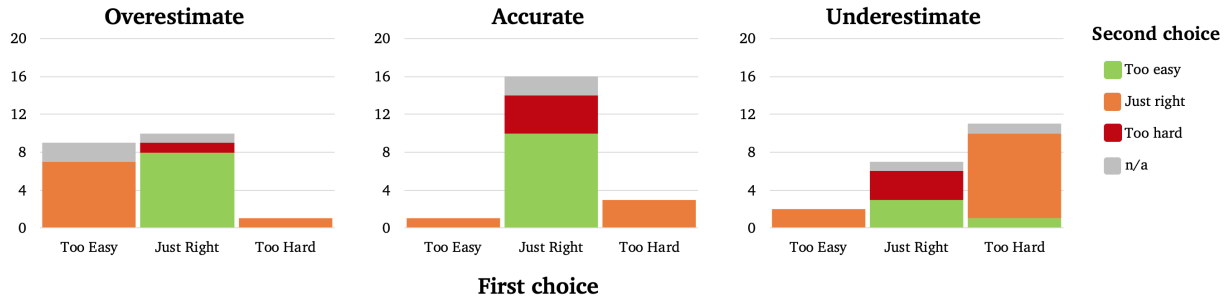

Figure 4: Children's first and second choices of games, split by experimental condition. The 3x3 Fisher's exact test on children's first choices was significant. Further, the proportions of children who chose the "just right" game in the Accurate condition, and the "too hard" game in the Underestimate condition, both significantly differed from chance.

perceived competence. We held this latter variable constant in the current work by controlling children's performance on the Experimenter's matching games; future work could instead test the converse by holding the Teacher's beliefs constant and manipulating the child's abilities on the task. This might illuminate the role that "confidence" plays in children's decisions about attempted task difficulty during pedagogy.

There were asymmetries in our data across conditions that could be worth exploring further. On the one hand, difficulty scores in the Accurate and Underestimate conditions did not significantly differ; but on the other hand, a majority of children's first choices in the Underestimate condition was the predicted ("too hard") game, while this was not the case in the Overestimate condition. Perhaps this reflects something about the trade-off between $p(S_G)$ and R_g in the computation of the expected value for each game across conditions, particularly with respect to the goals imposed by the experiment. Indeed, children were told that the Teacher was going to come back to see how many points they earned. This is distinct from how developmental learning tasks are typically construed, in which exploration is relatively consequence-free; thus, how the goals of our experiment might have asymmetrically incentivized choice patterns across conditions (e.g., see Figure 4) would be an exciting question for future work.

There are also open questions about what exactly children believed the Teacher was inferring about them. Did they think her judgment of their performance was limited to this particular matching game task? Or might they have taken this to be an evaluation of their competence in general? Answers to these questions could inform our understanding of how the current findings might extend to more formal pedagogical contexts, such as classroom settings. Indeed, *student-teacher rapport* is a key factor in predicting students' academic success (see Wentzel, 2009 for a review). For instance, Skinner and Belmont (1993) found that teachers' support of individual students' autonomy predicted children's motivation in grades three through five; and correlations have been found between student–instructor relationships and student achievement orientations even in samples of college students (Creasey, Jarvis, & Gadke, 2009). While rapport is a rich and complex construct that likely incorporates a multitude of mutually dependent social judgments that extend beyond the confines of one specific task (e.g., general liking of one another), it is possible that one component of rapport is this sense that a teacher is providing evidence at the appropriate level for the learner. If we are to ultimately connect the current line of work to the formal education literature, it will be critical to investigate how children believe teachers' judgments of their competence generalize across domains, tasks, and time.

It may be fruitful to expand the target age range of this study in both directions. Testing younger children on this task could shed light on the cognitive capacities that serve as prerequisites for this kind of nuanced social reasoning in pedagogy. While we suspect that false belief reasoning abilities might support performance on our task, we could directly test this idea by collecting data from toddlers and preschoolers in an adapted paradigm. On the other hand, it might also be worthwhile to investigate whether our findings extend into adulthood, and how individual differences in traits such as Need for Cognition (the degree to which an individual enjoys engaging in cognitively challenging tasks; Cacioppo & Petty, 1982) might influence choices on this task. This developmental approach would also allow us to further flesh out our model: Are these potential additional factors better thought of as moderating variables, or unique causal nodes? Answers to these questions could elucidate how these representations might shift throughout the lifespan.

The information provided by other people can help us navigate an uncertain world. But even well-intentioned teachers often have imperfect knowledge, both about the subject matter and about their learners. How does learning remain robust in the face of non-ideal teaching? Here, we show how even young learners may capitalize on subtle information about others' knowledge to make rational inferences during informal pedagogy. This work connects to a growing body of literature showing that from a young age, children deftly reason both about evidence, and the people providing that evidence. Although many open questions remain, we hope that we have neither overestimated nor underestimated the potential implications of the current work.

Acknowledgements

We thank Taisha Blanc and Sugandhi Yerram Reddy for assistance reliability coding. Thanks also to our participating families and to ChildrenHelpingScience.com, which allowed us to recruit families to participate online during the COVID-19 pandemic. This work was funded in part by an American Fellowship from AAUW to IB, a Henry Rutgers Presidential Fellowship to IB, and support from the McDonnell Foundation and the Jacobs Foundation to EB.

References

Bass, I., Bonawitz, E., & Gweon, H. (2017). Didn't know, or didn't show? Preschoolers consider epistemic state and degree of omission when evaluating teachers. *Proceedings, 39th Annual CogSci Conference*.

Bass, I., Gopnik, A., Hanson, M., Ramarajan, D., Shafto, P., Wellman, H., & Bonawitz, E. (2019). Children's developing theory of mind and pedagogical evidence selection. *Developmental Psychology*, *55*(2), 286–302.

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, *53*, 370-418.

Bonawitz, E., & Shafto, P. (2016). Computational models of development, social influences. *Current Opinion in Behavioral Sciences*, *7*, 95-100.

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.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *JPSP*, *42*(1), 116–131.

Creasey, G., Jarvis, P., & Gadke, D. (2009). Student attachment stances, instructor immediacy, and student-instructor relationships as predictors of achievement expectancies in college students. *Journal of College Student Development*, *50*(4), 353–372.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *TiCS*, *13*(4), 148–153.

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

Gweon, H., & Schulz, L. (2019). From exploration to instruction: Children learn from exploration and tailor their demonstrations to observers' goals and competence. *Child Development*, *90*(1), e148–e164.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273.

Hume, D. (1748/1999). *An enquiry concerning human understanding*. Oxford: Oxford University Press.

Karmiloff-Smith, A. (1996). *Beyond modularity: a developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*(3), 151-166.

Piaget, J. (1929). *The child's conception of the world*. New York, NY: Harcort, Brace.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. *Proceedings, 30th Annual CogSci Conference*.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351.

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.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581.

Tamis-LeMonda, C. S., Adolph, K. E., Lobo, S. A., Karasik, L. B., Ishak, S., & Dimitropoulou, K. A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Developmental Psychology*, *44*(3), 734—746.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wentzel, K. R. (2009). Students' relationships with teachers as motivational contexts. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 301–322). New York, NY: Routledge/Taylor & Francis Group.

Yang, S., & Shafto, P. (2017). Teaching versus active learning: A computational analysis of conditions that affect learning. *Proceedings, 39th Annual CogSci Conference*.

Yu, Y., Landrum, A. R., Bonawitz, E., & Shafto, P. (2018). Questioning supports effective transmission of knowledge and increased exploratory learning in pre-kindergarten children. *Developmental Science*, *21*(6), e12696.