Adults and school-aged children accurately evaluate sins of omission in pedagogical contexts

Hyowon Gweon (hyora@mit.edu), Hannah Pelton (hannahp@mit.edu), Laura E. Schulz (lschulz@mit.edu)
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract

Recent formal models of pedagogy (Shafto & Goodman, 2008) assume that teachers provide evidence likely to increase the learner's belief in a target hypothesis. Thus in pedagogical contexts, the learner can infer that evidence is not merely true of the concept but representative of it. If for instance, a teacher demonstrates a single function of a toy, the learner should assume that only that function exists. (If there were more, the teacher should have provided evidence accordingly). What happens when a teacher violates these pedagogical sampling assumptions (e.g., showing only one function of a toy with many functions)? If the learner discovers that the evidence is incomplete, does the learner evaluate the teacher accordingly? Here we show that, much as learners are sensitive to cases when informants are inaccurate (sins of commission), both adults and children are sensitive to sins of omission and penalize teachers who provide information that is accurate but incomplete.

Keywords: cognitive development, pedagogy, Bayesian model, causal learning, social evaluation, pragmatics

A recent computational approach to understanding human cognition has formalized pedagogical learning as a set of paired inferences based on the following assumptions: (1) that learners rationally update their beliefs given data and (2) that teachers choose the data most likely to increase the learners' belief in the correct hypothesis (Shafto & Goodman, 2008). This account suggests that in pedagogical contexts, data provided by a teacher is assumed to be not only true of a concept but also representative of it. This assumption constrains the hypothesis space and helps the learner converge on the correct hypothesis given sparse data.

Consider, for example, a toy with four different functions. If a naïve person accidentally discovers one function of the toy, this provides evidence that the observed function is true of the toy but little beyond that: the toy might have just that function, or there might be one, two, or n, more undiscovered functions. However, when a knowledgeable, helpful teacher demonstrates one function of the toy, the learner should infer that the toy has just one function. This inference is warranted by the assumption that the evidence was chosen to maximize the learner's belief in the correct hypothesis. Thus the learner can infer not only that the observed function is present but also that other functions are not. That is, in pedagogical contexts, absence of evidence for additional functions is evidence for their absence.

Previous research (Bonawitz, Shafto, Gweon, Goodman, Spelke, & Schulz, 2010) formalized these intuitions and found that pedagogical sampling affected children's exploration and learning. When preschoolers were given a pedagogical demonstration of one of four functions of a novel toy, children constrained their exploration almost exclusively to the demonstrated function and discovered few other properties of the toy. When the demonstration was accidental or when no demonstration was provided, children explored the toy more broadly and consequently discovered much more about it.

What happens when teachers provide only partial information about what is to be learned? If children are in a position to discover that the teachers' information was incomplete, do they recognize that the teachers' failure to provide all the relevant evidence was indeed a failure? That is, do children recognize violations of pedagogical assumptions as sins of omission?

Previous research suggests that, by the age of four, children are sensitive to the epistemic status of others and prefer to learn from teachers who provide accurate rather than inaccurate information. For instance, Sabbagh & Baldwin (2001) found that four-year-olds preferentially learned the meaning of novel words from a speaker who said she knew the meaning of the words rather than one who professed ignorance. Furthermore, children treat reliability as a relatively stable trait, preferring to learn novel words from teachers who were previously accurate in labeling familiar objects rather than from teachers who were previously inaccurate (Koenig & Harris, 2005).

However, previous work on reliability looked at whether learners are sensitive to informants' tendency to provide false information: a sin of commission. To our knowledge, there have been no studies looking at whether learners are sensitive to informants' tendency to provide incomplete information: a sin of omission.

There is reason to believe that detecting sins of omission involves a more sophisticated inference than detecting sins of commission. To detect a sin of commission, the learner only needs to recognize whether presented information is true or false. In sins of omission, the information provided is true but nonetheless increases the learner's belief in a false hypothesis (e.g., that the demonstrated function of a toy is the only function that the toy has). What the teacher violates is not the learner's knowledge about the world, but the learner's implicit assumptions about what constitutes helpful teaching; a truly helpful teacher should provide information that increases the learner's belief in the correct hypothesis. Although past research suggests that children are astute social learners who are sensitive to informants' reliability with respect to accuracy from very early on, we do not know when the ability to make these more subtle social evaluations develops.

Indeed, we do not know to what extent even adults are sensitive to teachers' provision of partial information. Such sins of omission are closely related to the Gricean Maxim of Quantity: the idea that a speaker should provide all relevant information in communicative contexts (Grice, 1975). Thus for instance, a speaker is guilty of violating this maxim if she (accurately) communicates that she has one sister when in fact she has two. The computational analysis of pedagogy can be thought of as providing a formalization of this kind of pragmatic inference. Although to our knowledge, no previous research has explicitly linked pedagogy and pragmatics, there are grounds for believing that adults should mistrust teachers who omit relevant information in pedagogical contexts.

Across two experiments, we asked whether both adults (Experiment 1) and school-aged children (6 – 7 yrs, Experiment 2) detect sins of omission in pedagogical contexts. Specifically, we modified the toy-teaching paradigm from previous work (Bonawitz et al., 2010) and asked whether (a) learners understand that absence of evidence for additional functions constitutes evidence for their absence (consistent with previous research) and (b) whether they can successfully incorporate this understanding into their evaluation of how helpful a teacher is in teaching about a toy.

Previous work established not only that instruction constrains exploration, but that children make the same inferences from vicarious instruction that they make from direct instruction: if children observe another child being taught a single function of a toy, children who only overhear the instruction are as likely to infer that the toy has only one function as children who are taught directly. That is, although ostensive cues directed to the learner may be helpful in signaling a situation as pedagogical (Csibra, & Gergely, 2009), such cues are not necessary for pedagogical sampling assumptions to obtain. They hold in any situation where a knowledgeable teacher communicates information to a presumably naïve (or epistemically comparable) learner overhearing the communicative exchange (Bonawitz et al., 2010). In the current study, we exploited this fact to create a task in which people first explored the toy themselves to discover different function(s) of the toy, and then observed a teacher teach a naïve learner about the toy. This design helped minimize the possibility that people's interest in exploring the toy would affect their ratings of the teacher (a concern particularly for the children).

Experiment 1: Adults

In Experiment 1, adult participants explored one of two perceptually identical toys that differed only in whether they had only a single function (Single-Function Toy condition) or multiple (four) functions (Multi-Function Toy condition).

Figure 1: Stimuli used in Experiments 1 and 2. Left: the Single-Function Toy and the Multi-Function Toy. In both conditions, only the wind-up mechanism (yellow circle) was demonstrated. Right: the Toy Teacher (left), the Student (right), and the rating scale (front).

They then observed a teacher who demonstrated just one function to a naïve learner. We measured participants' ratings of how helpful the teacher was in teaching the learner and compared the average rating between the two conditions.

Methods

Participants

Twenty-one college students (mean age (SD): 19.57 (1.4), range: 18 – 22 years, 10 males) were randomly assigned to either a Single-Function Toy (N=11) condition or a Multi-Function Toy (N=10) condition. Two participants (one in each condition) were dropped from analysis because they did not meet the two inclusion criteria (see Results for details).

Stimuli

Photos of stimuli actually used in the experiment are shown in Figure 1. Two similar novel toys were constructed out of foamboard, wooden sticks, and electronic circuits and parts taken out from commercially available toys. Both toys were yellow four-sided pyramids and on the apex was a transparent plastic globe. Each face of the pyramid contained one or more potential affordances: a green button, a yellow button, a purple wooden stick, an orange button, and a purple and black wind-up knob. On the Multi-Function Toy, all affordances except the purple wooden stick (which was inert in both toys) were functional. First, when the wind-up knob was twisted, the part displayed a flapping motion. When the green button was pressed, a light-up and spinning mechanism activated inside the transparent globe. The yellow button played music, and the orange button activated two LED lights. These effects continued as long as the button was held down and stopped when the button was released. On the Single-Function Toy, the only functional affordance was the wind-up toy.

Three puppets were used as teachers, and one Elmo puppet on a stand was used as the naïve learner. Each teacher puppet was assigned a different role: one taught Elmo about the novel toy (Toy Teacher), one correctly named familiar objects (Correct Teacher), and one incorrectly named familiar objects (Incorrect Teacher). A plastic carrot, a duck, a small stuffed rabbit, and a plastic corn-on-the-cob were used familiar objects.

A rating scale was constructed with wooden bar and ceramic knob that sled along the scale. The scale had five anchor points: a very frowny face, a mildly frowny face, a neutral face, a mildly smiley face, and a very smiley face, from left to right of the sliding bar. Between each of the anchor points were four lines to allow for more accurate coding, thus the total range of scores was from 1 to 20.

Procedure

Participants were tested alone in a quiet room. The toys, puppets, and scale were initially put behind the experimenter covered with a cloth or in a bag.

To begin, the experimenter first explained that the task was designed to be appropriate for young children, and that they were going to watch different teachers teaching Elmo. Elmo was introduced as a "silly monster" who did not know much about toys. She said that the teachers would teach Elmo about their toys. They were told that some teachers might be helpful and good at teaching and that some might be not helpful and not good at teaching, and that their task was to rate the teachers' helpfulness in teaching Elmo.

After this introduction, Elmo was put back under the table, and the experimenter said "Ok, so before Elmo gets to see the teacher teach him about the toy, why don't you see if you can figure out how it works?" She gave the participants either the Single-Function Toy or Multi-Function Toy depending on the condition. Participants were allowed to explore the toy until they had tried every part of the toy. The functions (when present) were readily discoverable, thus all participants entered the study knowing whether there was only a single function of the toy or four functions of the toy.

Once the participant tried all parts of the toy, the experimenter placed Elmo on the table and brought out the Toy Teacher puppet. In both conditions, the Toy Teacher said, "I am going to teach you how my toy works", and demonstrated the wind-up part of the toy by turning on the knob and showing the flapping motion to the participant. Then the teacher said "See? This is how my toy works" and demonstrated the wind-up part again. After the demonstration, the participants rated the teacher on the sliding scale. After participants rated Toy Teacher, the experimenter asked them to show Elmo how the toy works.

Then, participants watched and rated the Correct Teacher who correctly named two familiar objects (a carrot and a duck) and finally the Incorrect teacher who named familiar objects incorrectly (by calling a stuffed rabbit a cow and the corn a cup). The order of teacher ratings was fixed so that any effects of anchoring on the rating scale could be held constant across participants and conditions.

Results and Discussion

We coded the ratings that participants gave to the three teachers by reading the mark on the sliding scale. One participant who gave a lower rating for the Correct Teacher than the Incorrect Teacher was excluded from analysis. Additionally, one participant who only gave extreme ratings to all three teachers (either 1 or 20) was excluded to ensure that only participants who understood the continuous range of the scale were included.

In the Single-Toy Function condition, the information provided by the toy teacher is both accurate and complete: the toy has just one function. However, identical information provided in the Multi-Function Toy condition is accurate but incomplete; the toy has three additional functions. If adults make the inference that the absence of evidence from a knowledgeable teacher constitutes evidence for absence, then they should be sensitive to teachers who commit the sin of omission. That is, those who saw the Toy Teacher demonstrate a single function on the Multi-Function Toy should rate the teacher lower than those who see the teacher demonstrate the same function on the Single-Function-Toy. Therefore, we predicted that the ratings of the Toy Teacher in the Multi-Function Toy condition should be relatively worse than the ratings of the Toy Teacher in the Single-Function Toy condition.

First, we compared participants' ratings for the two control teachers. As predicted, adults in both conditions gave high ratings for the Correct Teacher (Single-Function Toy: M(SD)=15.55(3.6) vs. Multi-Function Toy: M(SD) = 13.66 (3.3); $z = 1.12$, $p = ns$, Mann-Whitney), and low ratings for the Incorrect Teacher (Single-Function Toy: M(SD)=3.65(4.3) vs. Multi-Function Toy: M(SD) = 3.61 (3.6); $z = -.08$, $p = ns$). Participants in both conditions rated the Correct Teacher higher than the Incorrect Teacher (Single-Function Toy: 15.6 (Correct) vs. 3.7 (Incorrect), $z = 2.81$, $p < 0.01$, Multi-Function Toy: 13.7 (Correct) vs. 3.6 (Incorrect), $z = 2.67$, $p < 0.01$; related-samples Wilcoxon Signed Rank Test).

We then compared the ratings people gave to the Toy Teacher[1] in the Multi-Function Toy condition and Single-Function Toy condition. As predicted, the average rating of the Toy Teacher in the Multi-Function Toy condition (M (SD) = 6.83 (4.5)) was significantly lower than the rating in

[1] The ratings for Toy Teacher and Correct Teacher reported in Experiment 1 pass the normality test (Kolmogorov-Smirnov, Shapiro-Wilk tests of normality), but the ratings for the Incorrect Teacher do not. In Experiment 2, children's ratings were not normally distributed in all measures. Therefore, we used nonparametric tests (Mann-Whitney U Test for between-subjects comparisons, and Wilcoxon Signed Rank Test for paired comparisons) throughout Experiments 1 and 2.

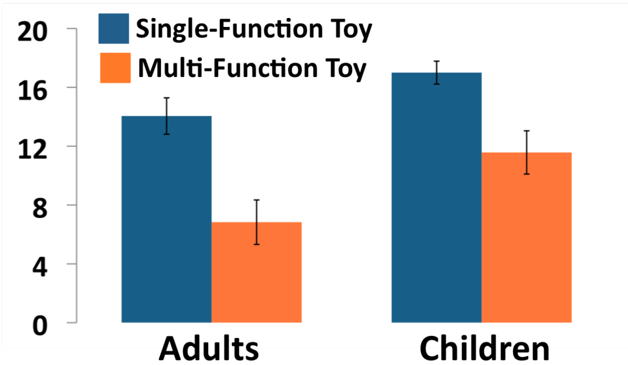

Figure 2: Average ratings for the Toy Teacher in Experiments 1 and 2.

the Single-Function Toy condition (M (SD) = 14.05(3.9), $z = 2.64$, $p < 0.005$).

Finally, participants in the Multi-Function Toy condition were much more likely to demonstrate multiple functions of the toy when they had a chance to teach Elmo rather than simply a single function (8 out of 9, $p < 0.05$ by binomial). Those in the Single-Function Toy condition were equally likely to show the functioning wind-up part or point out the non-functioning parts (5 out of 10, $p = ns$.)

These results suggest that adults indeed expect teachers to provide not just accurate, but also complete, information about what is to be learned. When the teacher demonstrated only one of four functions of a toy, they gave the teacher a lower rating than if the identical demonstration was provided for a toy that actually had just one function. In addition, almost all people in the Multi-Function Toy who observed a teacher that commits a sin of omission did not commit such sin themselves when they were given a chance to teach Elmo.

Experiment 2: Children

The results from Experiment 1 establish that adults are indeed able to detect "sins of omissions" in pedagogical contexts and can accurately incorporate such violations into their evaluations of the teachers. In our second experiment, we investigated whether 6 and 7-yr-old children exhibit a similar understanding. We tested children in this age range because studies of informant reliability with respect to sins of commission had investigated primarily four and five-year-olds (Sabbagh & Baldwin, 2001; Koenig & Harris, 2005). We believed that this task was slightly more difficult and thus early investigations should begin with slightly older children. Additionally, because young school-aged children are just beginning the process of formal pedagogy, their responses to information provided by teachers' are of ecological interest.

Participants Forty-one school-aged children (mean age (SD): 6.94 (0.63), range: 6.03 – 7.97) were recruited from a local children's museum and randomly assigned to either the Multi-Function Toy condition (N=21) or the Single-Function Toy condition.(N=20). No children were dropped due to parental interference or experimental error. We dropped and replaced four children who did not give a higher rating to the Correct Teacher than to the Incorrect Teacher, and an additional ten children who only gave extreme scores of 1 or 20 to all the teachers (i.e., whose use of the scale could not be distinguished from merely liking to slide the knob). This ensured that children who did not yet possess a good understanding of the task instruction and rating scales were not included in the final analysis.

Stimuli Stimuli used in Experiment 2 were identical to those described in Experiment 1.

Procedure All children were tested individually in a quiet room inside the museum. The child sat on a small stool in front of a round table, across from the experimenter. The parent was behind the child and out of the child's line of sight, and all the sessions were video-recorded.

The procedure was almost identical to that used in Experiment 1, except that after introducing the teacher puppets and Elmo (student) to the child, the experimenter said that the child is going to help them be a good teachers by playing a "rating game" to show how helpful the teachers were in teaching Elmo. Second, children were asked to demonstrate on the rating scale how they would rate a "good teacher", "bad teacher", or "teachers who are just okay". All children were able to do this.

Results and Discussion

We first compared children's ratings for the Correct and Incorrect Teachers between conditions. Like the adults, children in both conditions gave high ratings for the Correct Teacher (Single-Function Toy: M(SD)=14.76(3.5) vs. Multi-Function Toy: M(SD) = 15.0 (4.6); $z = -.19$, $p = ns$), and low ratings for the Incorrect Teacher (Single-Function Toy: M(SD)=2.69(2.7) vs. Multi-Function Toy: M(SD) = 4.10 (4.9); $z = -.07$, $p = ns$).

In both conditions, children's average rating for the Correct Teacher was significantly higher than the average rating for the Incorrect teacher (Single-Function Toy: 14.8 (Correct) vs. 2.7 (Incorrect), $z = 4.02$, $p < 0.001$, Multi-Function Toy: 15.0 (Correct) vs. 4.1 (Incorrect), $z = 3.93$, $p < 0.001$; related-samples Wilcoxon Signed Rank Test).

Next, we compared children's ratings for the Toy Teacher between the two conditions. As predicted, children in the Multi-Function Toy condition gave lower ratings to the Toy Teacher than those in the Single-Function Toy condition (Single-Function Toy: M(SD) = 17.00 (3.6) vs. Multi-Function Toy: M (SD) = 11.58(6.6), $z = 2.64$, $p < 0.01$).

Additionally, given a choice between just imitating the teacher or demonstrating additional functions, children in the Multi-Function Toy condition were much more likely to demonstrate multiple functions of the toy to Elmo than children in the Single-Function Toy condition (9 of 21 (Single-Function Toy) vs. 18 of 20 (Multi-Function Toy), $\chi 2 (1, N = 41) = 10.12$, $p < 0.005$).

These results suggest that children, like adults, understand that a teacher who demonstrates one function of a multi-function toy is less helpful than a teacher who demonstrates the same function on a toy that only has that function. Despite the fact that the teacher's demonstration was identical across conditions, children in the Multi-Function Toy condition gave lower ratings to the Toy Teacher than did children in the Single-Function Toy condition. Also, given an opportunity to teach Elmo themselves, both adults and children did not commit sins of omission (even when they had just observed a teacher who did): instead participants tried to provide complete information to the learner.

The predicted results in Experiment 2 were comparable to those from adult participants in Experiment 1. The only difference between children and adults was that children were more likely than adults to rate the toy teacher favorably across the board. (Single-Function Toy condition: 17.00 (Children) vs. 14.05 (Adults) $z = 2.19$, $p < 0.05$); Multi-Function Toy condition (11.57 (Children) vs. 6.83 (Adults) $z = 1.84$, $p = 0.06$). Presumably, children were inclined to rate the teacher positively in part simply because she introduced a toy (rather than an object label). Interestingly, the children's rank order ratings of the three teachers varied by condition. Relative to the Multi-Function condition, more children in the Single-Function toy condition rated the toy teacher the best of the three (Single-Function Toy condition: 15 of 21 children, Multi-Function Toy condition: 3 of 20 children, $p < 0.001$ by Fisher's Exact). This result suggests that although children's ratings can be affected by factors irrelevant to the quality of the teaching (e.g., the content taught or the object shown), they are sensitive to sins of omission and judge teachers accordingly.

General Discussion

In the current study, we asked whether adults and school-aged children in a pedagogical context accurately distinguish a teacher who provides true but *incomplete* information from a teacher who provides both true and complete information. Our results were consistent with the rational analysis of pedagogy. Across two experiments, we showed that adults and children (a) understand that a knowledgeable and helpful teacher should provide evidence that increases a learner's belief in a correct hypothesis, and (b) incorporate this understanding into their social evaluation of knowledgeable others. Therefore, a teacher who taught one function on a four-function toy was rated lower than a teacher who taught the same function on a toy where only that function was present.

In addition, we found that learners themselves resist committing sins of omission: given a chance to teach someone else, people spontaneously gave information that was both accurate and complete. Not only adults but also young children in the Multi-Function Toy condition avoided imitating the behavior of the Toy Teacher and instead demonstrated more than one function of the toy for the student.

Taken together, these results suggest that by the age of six, children make rational inferences about what teachers intend to communicate, can accurately detect sins of omission, and can incorporate these inferences into to their evaluations. To our knowledge, this is the first demonstration that children and adults can successfully penalize others for the provision of partial information in communicative contexts.

Humans have always relied on social transmission of information to pass on knowledge from one generation to the next (Vygotsky, 1978; Tomasello, 1999). One factor that may contribute to the power of human learning is our ability to infer unseen evidence in pedagogical contexts; in particular, given pedagogical sampling assumptions, learners can infer the absence of functions merely from absence of evidence for them (also see other work on strong sampling and the ability to infer the extent of concepts from only positive evidence: Tenenbaum, 1999; Gweon, Tenenbaum, & Schulz, 2010). However, the efficacy of this powerful mechanism depends on the reliability of the teacher. If the teacher provides only partial information the learner might incorrectly infer that what has been provided is all there is to be learned.

The current research suggests that processes of social evaluation can act as a check on misleading pedagogy. Adults, and even children, successfully detect teachers who provide incomplete evidence in pedagogical contexts. Thus even in childhood, social evaluations depend not just on attractive, how friendly, or how powerful other agents are, but also on a rational analysis of how likely they are to provide information that supports accurate learning.

Acknowledgments

Thanks to Joseph Kim for help with data collection. This research was supported by an NSF Faculty Early Career Development Award, a John Templeton Foundation Award, and James S. McDonnell Foundation Collaborative Interdisciplinary Grant on Causal Reasoning to L.S.

References

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N., Spelke, E., & Schulz, L. (2010). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, doi:10.1016/j.cognition.2010.10.001*.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148-153.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics: Vol. 3, Speech Acts*. New York: Academic Press.

Gweon, H., Tenenbaum, J., & Schulz, L. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107*(20), 9066 - 9071.

Koenig, M., & Harris, P. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*(6), 1261-1277.

Sabbagh, M., & Baldwin, D. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*(4), 1054-1070.

Shafto, P., & Goodman, N.D. (2008). Teaching games: Statistical sampling assumptions for pedagogical situations. In Proceedings of the 30th annual conference of the cognitive science society (pp. 1632–1637).

Stiller, A., Goodman, N. D., & Frank, M. C. (2011). Ad-hoc scalar implicatures in adults and children. Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.

Tenenbaum, J. B. (1999). *A Bayesian framework for concept learning. Ph.D. Thesis.* Cambridge, MA: MIT.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Vygotsky, L. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.