CHILD DEVELOPMENT

Child Development, xxxx 2017, Volume 00, Number 0, Pages 1–15

Order Matters: Children's Evaluation of Underinformative Teachers Depends on Context

Hyowon Gweon and Mika Asaba
Stanford University

The ability to evaluate "sins of omission"—true but pragmatically misleading, underinformative pedagogy—is critical for learning. This study reveals a developmental change in children's evaluation of underinformative teachers and investigates the nature of their limitations. Participants rated a fully informative teacher and an underinformative teacher in two different orders. Six- and 7-year-olds (N = 28) successfully distinguished the teachers regardless of the order (Experiment 1), whereas 4- and 5-year-olds (N = 82) succeeded only when the fully informative teacher came first (Experiments 2 and 3). After seeing both teachers, 4-year-olds (N = 32) successfully preferred the fully informative teacher (Experiment 4). These results are discussed in light of developmental work in pragmatic implicature, suggesting that young children might struggle with spontaneously generating relevant alternatives for evaluating underinformative pedagogy.

Much of early learning unfolds in social contexts. Young children spend much of their time surrounded by others who constantly communicate with them and teach them about the world. Children also approach others as sources for learning, actively requesting information to learn from knowledgeable, helpful others (Goupil, Romand-Monnier, & Kouider, 2016; Gweon & Schulz, 2011). Therefore, it is critical for young learners to recognize and evaluate others as informants to decide whom to approach and trust for information, and whom to avoid or discredit.

Previous research shows that even young children recognize teachers who provide inaccurate information (e.g., providing wrong labels for common household items) and preferentially choose to learn from previously accurate informants (e.g., Birch, Vauthier, & Bloom, 2008; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004; Pasquini, Corriveau, Koenig, & Harris, 2007). However, real-world teachers and informants rarely tell blatant lies or provide information that is obviously false; their mistakes can be rather subtle, such as omitting relevant pieces of information. Recognizing and evaluating such underinformative pedagogy can be particularly challenging for young learners, because the information provided by the teacher is still true of the world.

In pedagogical contexts where teachers are expected to be knowledgeable and helpful, learners readily draw inferences that go beyond what was explicitly communicated. For instance, when a teacher demonstrated *one* interesting function of a novel toy, young children (even toddlers and children in cultures where explicit teaching is rare) inferred that the teacher intended to communicate *just one* aspect of the toy; children constrained their exploration to the demonstrated part, treating the absence of additional information as evidence for the absence of additional aspects to be learned about the toy (Bonawitz et al., 2011; Shneidman, Gweon, Schulz, & Woodward, 2016). This inference rests on the assumption that a helpful, knowledgeable teacher engages in *pedagogical sampling* to select information that is maximally helpful for the learner; if the toy had additional interesting parts, the teacher would have shown them, too (Shafto, Goodman, & Frank, 2012; Shafto, Goodman, & Griffiths, 2014). Thus, pedagogical contexts license inductive leaps from even omission of information, allowing learners to draw powerful inferences about the world.

Many thanks to Ilona Bass and Shivani Kaushal for help with data collection, Hannah Pelton and Cristina Spicer for help with filming stimuli, and Mike Frank, Laura Schulz, Tamar Kushnir, and Stella Christie for helpful comments and discussions. We are also grateful to Bing Nursery School, Palo Alto Junior Museum & Zoo, Boston Children's Museum, and the families who participated. This work was funded by Varieties of Understanding grant from the Fordham University and the John Templeton Foundation.

Correspondence concerning this article should be addressed to Hyowon Gweon, Department of Psychology, Stanford University, 450 Serra Mall, Jordan Hall (Bldg. 420), Stanford, CA 94305. Electronic mail may be sent to gweon@stanford.edu.

 0009-3920/2017/xxxx-xxxx
DOI: 10.1111/cdev.12825

However, when teachers violate such expectations, the power of pedagogical contexts can turn into a hazard. Imagine that a teacher demonstrated one function of a novel toy when the toy in fact had additional, undemonstrated functions; a learner who assumes that the teacher used pedagogical sampling would end up drawing an inaccurate inference about the toy. In this case, the teacher did not provide any false information about the toy. Instead, the teacher committed a *sin of omission*, misleading the learner by providing true but incomplete information.

A recent study has shown that 6- and 7-year-olds appropriately evaluate sins of omission in pedagogical contexts (Gweon, Pelton, Konopka, & Schulz, 2014). In this study, children observed and rated a puppet teacher who demonstrated one interesting function of a toy to a naïve learner. One group of children saw a fully informative teacher, because the toy had just one function. The other group of children saw an underinformative teacher, because the toy had three additional functions that were left undemonstrated. Even though the appearance of the toy and the teacher's behaviors were identical across conditions, children's ratings were significantly lower for the underinformative teacher than for the fully informative teacher. Furthermore, when the teacher later introduced a new toy and demonstrated one interesting function, children explored the new toy more broadly if the teacher had been previously underinformative than when he had been fully informative. These results suggest that by age 6, children explicitly evaluate sins of omission and even adjust their exploratory behaviors to compensate for potentially underinformative pedagogy.

Children's Evaluation of Underinformative Pedagogy

As noted earlier, however, prior work shows that sensitivity to pedagogical sampling emerges much earlier in childhood. Even toddlers treat the absence of additional information as meaningful in pedagogical contexts, drawing inferences not only from what was demonstrated but also from what was left undemonstrated (Bonawitz et al., 2011; Shneidman et al., 2016). This suggests that children must start facing the potential hazards of underinformative pedagogy well before age 6. Do these young children understand that a violation of pedagogical sampling can mislead the learner, and do they evaluate underinformative teachers accordingly? How does the sensitivity to sins of omission develop during early childhood, and what does it tell us about the cognitive mechanisms that underlie our evaluation of informativeness?

One possibility is that an understanding of pedagogical sampling is the necessary and sufficient prerequisite for the ability to accurately recognize and evaluate underinformative pedagogy. In other words, once children expect that a knowledgeable, helpful teacher would provide true and complete information and draw strong inferences based on these expectations, they might have no trouble evaluating those who violate such expectations. If so, children might show no difficulty evaluating sins of omission well before age 6.

An alternative possibility is that appropriate evaluation of underinformative pedagogy requires more than an understanding of pedagogical sampling. To provide a lower rating to an underinformative teacher than to a fully informative teacher, children should not only understand that the teacher's demonstration leads to an inaccurate inference (i.e., inferring that the toy has just one function when it in fact has four) but also recognize that the teacher could have done better by providing more relevant information and explicitly penalize the omission he could have avoided. Children younger than age 6 might experience difficulty with any of these aspects of informant evaluation.

Development of Children's Ability to Draw Pragmatic Implicature

We believe that useful insights can be gained from prior work on the development of pragmatic implicature, which suggests that even though young children readily draw pragmatic implicatures from verbal utterances, they still have trouble evaluating underinformative speakers. Almost a decade ago, Baldwin, Loucks, and Sabbagh (2008) made an intriguing theoretical conjecture that our reasoning about goal-directed, intentional actions might be subject to similar constraints that underlie our inferences from verbal communicative behaviors, guided by intuitions akin to Grice's (1975) cooperative principles. Following this argument, here we first provide a brief review of this literature and draw parallels between children's understanding of pedagogical sampling and children's understanding of pragmatic implicatures. We then use these connections to motivate our specific hypothesis about young children's ability to evaluate underinformative teachers in pedagogical contexts.

Recent work on children's understanding of scalar implicature finds that even preschoolers readily go beyond the literal meaning of speakers' utterances to infer their intended meanings. For instance, when asked to find "a friend with glasses," 3-year-olds choose a face with *just* glasses over a face with glasses *and* a top hat, even though both choices are logically consistent with the utterance (Stiller, Goodman, & Frank, 2015). Note that these results parallel children's inferences about a toy's function in pedagogical contexts; just as the teacher would have demonstrated more if the toy had additional functions, the speaker would have mentioned the top hat if he really meant the other face. In both cases, children went beyond the literal interpretation of the utterance or demonstration to infer what the informant intends to communicate, with the expectation that the communicator is knowledgeable, helpful, and informative.

However, despite the early emergence of such pragmatic competence, a large body of work suggests that children under 6 years of age have difficulty evaluating underinformative utterances that violate this expectation. For instance, a speaker who says "the boy ate *some* cookies" when he ate *all* the cookies is underinformative; he used a weak scalar term "some" when a stronger term "all" applies, thus providing a logically true yet pragmatically infelicitous statement that can mislead the listener to infer that the boy did *not eat all* the cookies. Many studies have found that children under age 6 consider such underinformative sentences as acceptable (Huang & Snedeker, 2009; Noveck, 2001; Papafragou, 2006; Papafragou & Musolino, 2003). These failures are not limited to scalar expressions, as 4-year-olds also fail to reject "the cat and the cow are sleeping" given a picture that shows three sleeping animals (a cat, a cow, and a dog; Barner, Brooks, & Bale, 2011). These results suggest that children's failure to penalize logically true but underinformative utterances is not necessarily due to their pragmatic incompetence and that the ability to draw pragmatic inferences alone is insufficient for accurately evaluating underinformative speakers.

Recent research offers at least two possible explanations for this failure. One possibility is that children might have difficulty in generating and representing alternative lexical items for a given scale (i.e., understanding that the speaker could have used "all" instead of the logically correct but pragmatically infelicitous "some"). For instance, given a picture of three sleeping animals (a cat, a cow, and a dog), children fail to reject sentences "some of the animals are sleeping" or "the cat and the cow are sleeping," but accurately reject "*only* the cat and the cow are sleeping" (Barner et al., 2011). This suggests that children are sensitive to underinformativeness, but it is revealed only when the alternatives are made clear and salient in context. Furthermore, exposing 5-year-olds to relevant uses of the scalar item "all" helps them successfully reject the underinformative uses of the term "some" (Skordos & Papafragou, 2016). Another possibility is that children are in fact sensitive to informativeness but fail to explicitly penalize underinformative speakers, because they are more forgiving of speakers who use an intermediate item (e.g., some) on the scale when a stronger item (e.g., all) should be used. For instance, Katsos and Bishop (2011) found that although 5- and 6-year-olds fail to reject underinformative utterances (e.g., "the boy ate some of the cookies" when he in fact ate all of the cookies) given binary choice (e.g., yes or no), they distinguish them from fully informative ones on a 3-point scale.

From Pragmatic Implicature to Evaluation of "Sins of Omission"

Drawing from these related findings, we might expect that the ability to draw strong inferences in pedagogical contexts is not necessarily sufficient for accurately evaluating underinformative pedagogy. Even though young children make inferences that are consistent with the use of pedagogical sampling (Bonawitz et al., 2011; Shneidman et al., 2016), they might still experience difficulty with evaluating sins of omission. As noted earlier, the difficulty might arise in at least two different ways. First, young children might fail to penalize sins of omission because they have trouble understanding that the teacher *could have* avoided misleading the learner by providing more information. In this case, we might see different patterns depending on which teacher was evaluated first; more specifically, seeing the fully informative teacher might help with their subsequent evaluation of the underinformative teacher. Second, young children might fail because they are more tolerant of underinformative pedagogy, in which case they would show generous evaluations regardless of the order.

In light of these possibilities, the current study investigates the developmental trajectory of children's ability to evaluate underinformative pedagogy. We designed a task similar to that used in Gweon et al. (2014) in which children first discovered all functions of a novel toy and then observed a puppet teacher teach another puppet learner about the toy. Critically, we adopted a within-subjects paradigm in which children observed and rated two teachers sequentially: a fully informative teacher and an underinformative teacher, in two different orders. Given the effect of prior exposure to better alternative scalar items in evaluating pragmatically underinformative utterances (Skordos & Papafragou, 2016), this within-subjects design allowed us to explore the possibility that young children's evaluations of an underinformative teacher are affected by their prior experiences with a fully informative teacher. Acknowledging limitations of using binary measures in assessing children's pragmatic competence (Katsos & Bishop, 2011), we used a continuous, multipoint scale following Gweon et al. (2014) so that children could provide graded evaluations for each teacher.

In Experiment 1, we first replicated 6- and 7-year-olds' ability to evaluate underinformative teachers (Gweon et al., 2014) using this computer-based, within-subjects task. In Experiment 2, we asked whether 4- and 5-year-olds show difficulty evaluating underinformative teachers and whether their evaluations are affected by order of teachers. In Experiment 3, we further explore the nature of young children's difficulty by manipulating the kind of teacher children observed first before observing the underinformative teacher. In Experiment 4, we presented 4-year-olds with a direct contrast between the two teachers in a binary choice paradigm.

Experiment 1

In Experiment 1, we replicated Gweon et al. (2014) with the same age group (6- and 7-year-olds) using a novel, computer-based task. Our task closely mirrored the one used in Gweon et al. (2014) except that children watched video clips of human actors discover all functions of the toys (rather than exploring the toys themselves) and saw videos of puppet teachers on stage (rather than seeing the experimenter impersonate a puppet teacher). This computer-based task used pre recorded videos to help minimize potential variability in children's interactions with the toys and the puppets' actions, and shortened the overall task to allow multiple teachers to be observed and evaluated sequentially. Each child provided a pair of ratings: one for a teacher who demonstrated one function of a single-function toy (*informative teacher*; equivalent to the teacher in Teach 1/1 condition in Gweon et al., 2014), and another teacher who demonstrated only one function of a four-function toy (*underinformative teacher*; equivalent to the teacher in Teach 1/4 condition in Gweon et al., 2014). By manipulating the order of teachers across conditions while controlling for the number of demonstrations, we asked whether children's evaluation of the underinformative teacher is influenced by their previous observation and evaluation of the informative teacher (Figure 1).

Method

Subjects

Twenty-eight 6- and 7-year-olds were recruited from a local museum (17 girls; $M_{age}(SD)$ = 7.05 (0.54), range = 6.07–7.90), and were randomly assigned to the informative first condition (N = 14) or the underinformative first condition (N = 14). This sample size was larger than the minimum

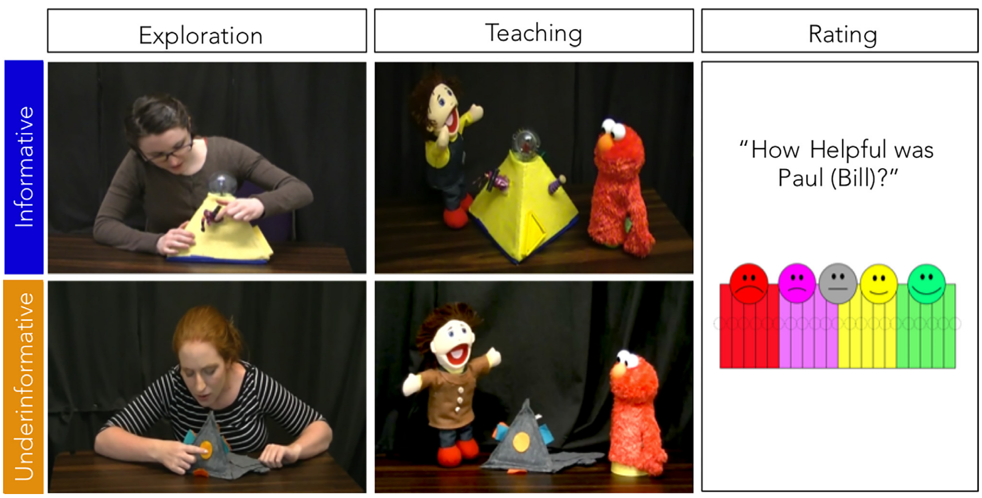

Figure 1. Procedures in Experiments 1 and 2, and the rating scale.

($N = 10$) required for 80% power in a within-subjects design given the large effect size ($d = 1.1$) in Gweon et al. (2014), Experiment 1 (data collection period: August 2013–April 2014; three additional children were tested in March 2015). An additional six children were dropped from analysis due to experimental error ($N = 2$), failure to pass the predefined inclusion criteria of correctly reporting the number of functions of the toy ($N = 2$), or rating the incorrect teacher higher than the correct teacher ($N = 2$; see Procedure). The demographics of participants were representative of a typical urban middle-class neighborhood.

Materials

Stimuli were presented as videos on a 13-in. MacBook Pro using MATLAB and PsychToolBox. Children viewed four trials total: two critical trials in which a puppet teacher demonstrated a toy (toy teacher trials: informative teacher and underinformative teacher) and two additional trials in which a puppet teacher provided labels for familiar objects (label teacher trials: correct teacher and incorrect teacher). Two custom-made toys, one yellow (Toy A) and one gray (Toy B), were used in the toy teacher trials. Toy A, which was used in Gweon et al. (2014), had four causal affordances: twisting a purple knob activated a wind-up mechanism, pressing a yellow button activated LED lights, pressing a green button activated a spinning light, and pressing an orange button played music tunes. Toy B also had four causal affordances: pressing a purple tab made a beeping sound, pressing a gray tab produced a buzzing sound, pulling down a flap on one side revealed a hidden mirror, and pulling down a flap on another side revealed a hidden embroidered duck. All of these were nonobvious causal affordances; thus, each toy could be presented as a toy with four functions (in the underinformative teacher trial) or presented as if it just had a single function (in the informative teacher trial). The type of toy was counterbalanced throughout; half the children saw Toy A in the informative teacher trial and Toy B in the underinformative teacher trial, whereas the other half saw the reverse. For label teacher trials, four common household objects were used: a stuffed carrot, a toy plane, a ball, and a stuffed tiger. Four hand puppets were used as the toy teachers (Paul and Bill for informative and underinformative trials) and label teachers (Sally and Laura for correct and incorrect trials). An Elmo puppet was used as a naïve learner. Children used a rating scale with tick marks (0–20) and a magnetic marker to evaluate each teacher. The scale was split into four different colored sections, along with faces that varied from frowny to smiley to serve as anchor points between the sections.

Procedure

Participants were tested in a quiet room inside the museum. All participants first received a brief training with the rating scale. The experimenter said, “We will watch some teachers who will teach Elmo about their toys. Then we will tell them how helpful they were in teaching Elmo so that they can do a better job next time.” She briefly explained the rating scale, and asked children to indicate where they would place the marker on the scale if the teacher did a “very good job,” “just okay,” and “not a good job” teaching Elmo about the toy. Children then underwent two toy teacher trials followed by two label teacher trials (correct and incorrect). Across two conditions, we varied the order of the toy teacher trials. In the informative first condition, the informative teacher trial was presented first; in the underinformative first condition, the underinformative teacher trial was presented first. Each toy teacher trial consisted of three phases: exploration, teaching, and rating (see Figure 1).

Exploration. Children first watched a video of a naïve adult exploring the toy. Two different adults explored Toy A and Toy B, respectively. In the informative teacher trial, the adult said, “I wonder what this toy does!” and discovered one function of the toy (wind-up mechanism on the Toy A; beep on Toy B) while acknowledging that other parts do not do anything. At the end of the video, she exclaimed, “This toy does one thing!” In the underinformative teacher trial, the adult sequentially discovered four functions of the toy, and said, “This toy does four things!” We ensured that the adult did not deliver any pedagogical cues; they initially claimed to be naïve about the toy, their utterances were self-directed, and they never directly addressed the child or made eye contact with the camera.

Children were then asked how many things the toy does. If the child could not answer or gave an incorrect answer, the experimenter replayed the video and prompted the child again. We dropped and replaced two children who were unable to report the correct number of functions after the second viewing.

Teaching. Children then watched a video of a toy teacher (Paul or Bill) teaching Elmo about the same toy they saw in the exploration phase. The toy teacher said, "Hi, I'm Paul (Bill), and I know all about this toy. I'm going to show you how it works!" Critically, in both the informative teacher and underinformative teacher trials, children watched the toy teacher demonstrate just one function. Thus, the toy teacher was fully informative in the informative teacher trial and underinformative in the underinformative teacher trial.

Rating. The experimenter then brought out the scale and asked the child, "How helpful was Paul (Bill)? How good of a job did he do teaching Elmo about the toy?" The participant indicated his or her response by placing a small marker on the rating scale. Then, the same procedure (exploration, teaching, and rating) was repeated with the other trial.

After rating two toy teachers, children rated two label teachers who taught Elmo the names of familiar objects. The correct teacher correctly referred to a toy carrot as a "carrot" and a toy plane as a "plane"; the incorrect teacher incorrectly labeled a ball as a "cup" and a stuffed tiger as a "cow." After each teacher provided names for the objects, the experimenter brought out the same scale and asked the child to rate the teacher. The order of the correct and incorrect teachers was counterbalanced. As prior work suggests that even preschoolers reliably distinguish informants who provide correct and incorrect names of familiar objects (e.g., Koenig & Harris, 2005), these ratings were mainly collected to identify children who did not yet understand how to use the rating scale.

Pilot data on an earlier version of the study suggested that younger children are less likely to provide clear justifications for their ratings; due to the length of the task and multiple rating trials, we collected explanations only as an optional, exploratory measure (mainly from older participants in Experiment 1, and about half of the participants in Experiment 2 who seemed comfortable verbally interacting with the experimenter) by asking them to justify their response after each rating.

Results and Discussion

First, we asked whether our results replicated Gweon et al. (2014) by comparing just the first trial of each participant, as if the two groups of children rated either the informative teacher or the underinformative teacher. Consistent with previous findings, 6- and 7-year-olds who rated the informative teacher on their first trial gave a higher rating than those who rated the underinformative teacher on their first trial (informative teacher M(SD) = M (SD) =14.75(4.18) versus underinformative teacher M(SD) = 8.57(5.45), $t(24) = 3.37$, $p = .003$).

Given that we implemented a within-subjects design in which each child rated both teachers, we also collapsed the informative first and underinformative first conditions to compare the average ratings for informative and underinformative teachers. Again, 6- and 7-year-olds gave a higher rating for the toy teacher in the informative teacher trials than in the underinformative teacher trials (informative teacher M(SD) = 15.30(4.73) versus underinformative teacher M(SD) = 8.25(5.78), $t(27) = 5.85$, $p < .001$).

To ask whether the order of the toy teachers affected ratings, we performed a 2 (trial: informative teacher, underinformative teacher) $\times$ 2 (condition: informative first, underinformative first) mixed analysis of variance (ANOVA) with *trial* as a within-subjects factor and *condition* as a between-subjects factor. The results revealed a significant effect of trial, $F(1, 26) = 33.01$, $p < .001$, $\eta^2 = .318$, no main effect of condition, $F(1, 26) = .29$, $p = .592$, and no interaction between condition and trial, $F(1, 26) = .04$, $p = .851$. Planned t tests confirmed that children in both conditions rated the informative teacher higher than the underinformative teacher (informative first condition: $M(SD) = 14.75(4.18)$ versus 7.93 (6.28), $t(13) = 4.34$, $p < .001$; underinformative first condition: $M(SD) = 15.86(5.33)$ versus 8.57(5.45), $t(13) = 3.86$, $p = .002$). Including children who failed to correctly evaluate the correct and incorrect teachers ($N = 2$) did not change these results. Informative first condition: (SD) = 14.70 (4.03) versus 8.6(6.59), $t(14) = 3.74$, $p = .002$; underinformative first condition: $M(SD) = 16.13(5.25)$ versus 9.33(6.02), $t(14) = 3.73$, $p = .002$.

Finally, age did not affect the extent to which children penalized the underinformative teacher relative to the informative teacher; there was no correlation between age and each participant's difference in ratings of the two teachers (Age $\times$ Difference in Ratings between the informative and underinformative teachers: $r = -.004$, $p = .983$).

Children's explanations further suggested that their low ratings for the underinformative teacher are due to his "sin of omission"; most children (9 of 12 and 8 of 10 in the informative first and underinformative first conditions, respectively) explicitly appealed to the number of functions taught or completeness of teaching (e.g., "he only showed him one thing but there were four things," "he didn't teach all of them").

These results replicate the findings from Gweon et al. (2014), showing that by age 6, children reliably detect sins of omission in pedagogical contexts.

Even though the two teachers each demonstrated one function of an interesting toy, they appropriately distinguished them in their ratings by giving a lower rating for an underinformative teacher than for a fully informative teacher. By using a within-subjects paradigm, we confirmed that this relative penalty for the underinformative teacher does not change with the order of teachers. As predicted by previously reported success in a between-subjects paradigm (Gweon et al., 2014), children rated the underinformative teacher poorly even when this teacher was the very first teacher they rated (Figure 2).

Experiment 2

In Experiment 2, we examined younger children's ability to evaluate sins of omission. More specifically, we explored two different ways in which 4- and 5-year-olds have difficulty evaluating underinformative teachers; they might either fail to recognize sins of omission across the board, or their sensitivity might be influenced by whether they observed and evaluated a fully informative teacher before the underinformative teacher.

Method

Subjects

Given the possibility for an age-related difference between ages 4 and 5, we planned for a larger sample size compared to Experiment 1, ensuring equal numbers of 4- and 5-year-olds. We thus recruited thirty-two 5-year-olds (16 girls; $M_{age}(SD)$ = 5.46(0.30), range = 5.00–5.97, N = 16 per condition) and thirty-two 4-year-olds(16 girls; $M_{age}(SD)$ = 4.51(0.30), range = 4.07–4.99, N = 16 per condition; data collection: August 2013–May 2014; October 2014–March 2015), from a local museum or a university-affiliated nursery school. Although many families were from middle-class families, their socioeconomic, cultural, and ethnic backgrounds were diverse and representative of the local population. Following Gweon et al. (2014) and Experiment 1, we excluded an additional twenty-two 4-year-olds and twelve 5-year-olds from analysis because they rated the incorrect teacher the same as or higher than the correct teacher. Given the high exclusion rate, we also report results including all children.

Materials

The stimuli were identical as in Experiment 1.

Procedure

All children were assigned either to the informative first or the underinformative first condition, and the procedure was identical to Experiment 1. Because pilot data suggested that 4-year-olds often get confused during the training, we removed the smaller tick marks between the main anchor points for 4-year-olds. This effectively converted the 21-point scale to a 5-point scale, but the scores were converted back to 21 points for comparisons with other data (mapping from 21- to 5-point scale: 0 to 1, 5 to 2, 10 to 3, 15 to 4, 20 to 5). All children received training almost identical to Experiment 1 except that the experimenter mapped all five points on the scale explicitly stating the meaning of each

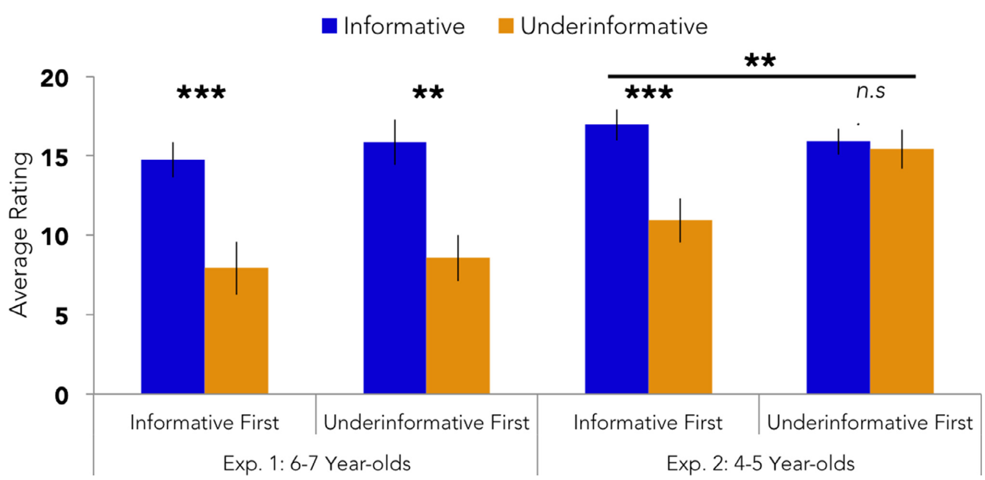

Figure 2. Results from Experiments 1 and 2. *p < .05. **p < .01. ***p < .001.

point (*very good job teaching*, *kind of good*, *just okay*, *kind of bad*, *really bad*) to ensure that children understood the scale and the purpose of the rating task.

Results and Discussion

We first compared the ratings between the informative and underinformative teachers using just the first trials, as if the children were run in a between-subjects task. We found that when children rated either the informative or underinformative teacher first, they did not distinguish between the two teachers (informative teacher $M(SD)$ = 16.95(5.55) versus underinformative teacher M(SD) = 15.44(7.00), $t(59) = 0.96$, $p = .341$). We found the same pattern in the older half of the participants 5-year-olds: informative teacher $M(SD)$ = 16.28(5.94) versus underinformative teacher M(SD) = 15.88(7.14), $t(29) = 0.17$, $p = .862$), as well as in the younger half (4-year-olds: informative teacher $M(SD)$ = 17.63(5.23) versus underinformative teacher M(SD) = 15.00(7.07), $t(27) = 1.19$, $p = .243$).

Given that each child rated both teachers, we then used all trials to compare ratings for the toy teachers by collapsing across conditions. We found that children did rate the informative teacher higher than the underinformative teacher, informative teacher $M(SD)$ = 16.42(5.09) versus underinformative teacher M(SD) = 13.19(7.72), $t(63) = 3.60$, $p < .001$, and this was true not only in the older half of the group, 5-year-olds: informative teacher $M(SD)$ = 15.63(5.21) versus underinformative teacher teacher M(SD) = 13.09(7.66), $t(31) = 2.20$, $p = .036$, but also in the younger half, 4-year-olds: informative teacher $M(SD)$ = 17.22(4.93) versus underinformative teacher M(SD) = 13.28(7.89), $t(31) = 2.85$, $p = .008$.

Our main question was whether children's ratings showed different patterns across conditions (order of trials). Collapsing across 4- and 5-year-olds, we found that a 2 (trial: informative teacher vs. underinformative teacher, within-subjects factor) × 2 (condition: informative first vs. underinformative first; between-subjects factor) mixed ANOVA revealed a significant effect of trial, $F(1, 62) = 15.07$, $p < .001$, $\eta^2 = .062$, and an interaction between trial and condition, $F(1, 62) = 11.14$, $p = .001$, $\eta^2 = .047$. Planned t tests confirmed that children in the informative first condition rated the informative teacher higher than the underinformative teacher, and the pattern was clear in both age groups (5-year-olds: informative teacher $M(SD)$ = 16.28(5.94) versus underinformative teacher M(SD) = 10.31(7.34), $t(15) = 4.22$, $p < .001$; 4-year-olds: informative teacher $M(SD)$ = 17.63(5.23) versus underinformative teacher M(SD) = 11.56 (8.51), $t(15) = 2.692$, $p = .017$). However, children showed a marked failure in the underinformative first condition (5-year-olds: informative teacher $M(SD)$ = 14.97(4.45) versus underinformative teacher M(SD) = 15.88(7.14), $t(15) = 0.66$, $p = .522$; 4-year-olds: informative teacher $M(SD)$ = 16.81(4.75) versus underinformative teacher M(SD) = 15.00 (7.07), $t(15) = 1.21$, $p = .245$).

We then asked whether children's relative penalty for the underinformative teacher changes with age. No correlation was found between age and the difference between their ratings for the informative and underinformative teachers in either condition (Age × Difference in Ratings: informative first: $N = 32$, $r = -.08$, $p = .675$; underinformative first: $N = 32$, $r = -.13$, $p = .486$).

Many children in this younger age group failed to pass the inclusion criteria: twenty-two 4-year-olds and twelve 5-year-olds failed to rate the correct teacher higher than the incorrect teacher. Given the high exclusion rate, we ran the analyses including these 34 children (total $N = 98$; $M_{age}(SD) = 4.91(0.55)$, range = 4.06–5.97), to confirm that all results remained significant. A 2 (trial) × 2 (order) ANOVA revealed a significant effect of trial, $F(1, 96) = 14.60$, $p < .001$, and an interaction between trial and order, $F(1, 96) = 12.62$, $p < .001$. Planned t tests in each condition for each age group also showed the same pattern, showing robust success in the informative first condition (5-year-olds: informative teacher M(SD) = 17.28(5.18) versus underinformative teacher M(SD) = 12.04(7.65), $t(22) = 4.05$, $p < .001$; 4-year-olds: informative teacher M(SD) = 18.48(4.09) versus underinformative teacher M(SD) = 13.28(7.56), $t(28) = 3.59$, $p = .001$), and a marked failure in the underinformative first condition (5-year-olds: informative teacher M(SD) = 15.21(5.57) versus underinformative teacher M(SD) = 15.24(6.91), $t(20) = -0.02$, $p = .986$; 4-year-olds: informative teacher M(SD) = 14.88(6.60) versus underinformative teacher M(SD) = 14.80(7.14), $t(24) = 0.047$, $p = .963$).

These results suggest that 4- and 5-year-olds show some sensitivity to sins of omission, but they do so only under specific conditions. Children's ratings were highly influenced by the order in which they saw different trials; children were able to provide a lower rating for the underinformative teacher (underinformative teacher) only if they had already rated a fully informative teacher (informative teacher). When they saw the underinformative teacher first, children rated this teacher just as highly as the fully informative teacher.

It is possible that these results emerged from a simple bias to provide a generous rating for *anyone* they evaluate first. If the baseline rating is higher in younger children, then even though they are sensitive to sins of omission, children in the underinformative first condition might have been unable to rate the informative teacher higher than the underinformative teacher simply because they had already given a very generous rating to the underinformative teacher. If this is the case, then children might provide a similarly generous rating even for a teacher who provides obviously false information. To address this possibility, we recruited a separate group of 4- and 5-year-olds (N = 12, 7 girls; M_{age} (SD) = 5.22(0.52), range = 4.37–5.97) years, and asked children to rate the incorrect teacher who provided inaccurate labels for familiar objects (we used the same video clip used in Experiments 1 and 2). The average rating, $M(SD)$ = 9.58(8.38), was just as low as the rating for the incorrect teacher in the main experiment (M(SD) = 4.02 (4.88), p = .196), and significantly lower than the average rating for the underinformative teacher in the underinformative first condition who was also rated first (incorrect teacher $M(SD)$ = 9.58(8.38) versus incorrect teacher M(SD) = 15.44(7.00), $t(17)$ = 2.15, p = .046). These data suggest that our 4- and 5-year-olds' failure to evaluate underinformative teachers in the underinformative first condition is not simply due to a high baseline rating.

As expected, the majority of the younger children were unable to verbally justify their ratings or simply restated the meaning of the scale (e.g., "because he did a good job"). However, 8 of 18 in the informative first condition explicitly mentioned the number of functions or the completeness of the demonstration while only 3 of 19 in the underinformative first condition did so (p = .08, two-tailed Fisher's exact test); although these measures were optional and the results are only suggestive, this provides further support that children in the informative first conditions were more likely to recognize omission as a "sin" and appropriately rate the teacher's underinformative demonstration.

In sum, young children's evaluation of sins of omission was highly influenced by the order of teachers. When 4- and 5-year-olds saw the fully informative teacher before the underinformative teacher, children's performance was indistinguishable from those of 6- and 7-year-olds in Experiment 1. Furthermore, the relative penalty for the underinformative teacher did not show any age-related change. However, when children saw the underinformative teacher before seeing the fully informative teacher, they failed to rate the two teachers differently.

Experiment 3

Results from Experiment 2 show that young children's ratings were clearly influenced by the order, suggesting that prior experience with a fully informative teacher is critical for successfully evaluating an underinformative teacher.

Why did young children succeed in the informative first condition but not in the underinformative first condition? One possibility is that the mere process of explicitly evaluating a teacher as "helpful" led children to be more vigilant, increasing their sensitivity for sins of omission. If so, we might expect any experience of evaluating a teacher as helpful would lead to success, such as giving a high rating to a teacher who provided accurate information.

However, insights from the pragmatics literature suggest that seeing an example of a fully informative demonstration might have helped children generate the relevant scale for evaluating the subsequent underinformative demonstration. Note that children were simply asked to rate "how helpful the teacher was" in teaching Elmo without any specification of what "helpful teaching" entails. That is, a teacher could be regarded as helpful simply by being nice and friendly, showing something interesting, or providing accurate information. Thus, children who saw the underinformative teacher first might have had difficulty understanding the relevant dimension on which the teachers ought to be evaluated. By contrast, children who saw the informative teacher first might have more easily recognized that the next teacher failed to be maximally helpful and that he could have provided additional demonstrations to be fully informative.

In Experiment 3, we provide additional support to the hypothesis that young children's difficulty arises from their inability to spontaneously generate a relevant representation of a fully informative teacher. If children's success were due to the mere experience of rating a teacher as helpful, then rating a teacher who was helpful for a different reason (i.e., providing accurate labels for objects) would also lead children to successfully evaluate sins of omission. However, if understanding the relevant dimension for comparison (i.e., providing complete information) is critical for evaluating sins of omission, children would continue to fail even after providing high ratings for such a teacher.

Method

Subjects

Eighteen 4- and 5-year-olds were recruited from a university-affiliated nursery school and local museum. Most children were from middle-class families with diverse socioeconomic, cultural, and ethnic backgrounds that were representative of the local population (7 girls; $M_{\text{age}}(SD) = 5.15(0.58)$, range = 4.04–5.83; data collection: June 2015–August 2015). Following previous experiments, we excluded seven additional children for rating the incorrect teacher the same as or higher than the correct teacher.

Materials

Stimuli were identical as in Experiments 1 and 2.

Procedure

Procedures were similar to Experiments 1 and 2, except that the trial order was fixed. The trials were always shown in this order: correct teacher, underinformative teacher, informative teacher, and incorrect teacher.

Results and Discussion

Children in Experiment 3 first rated the correct teacher, followed by the underinformative teacher, and then the informative teacher. If the success in the informative first condition in Experiment 2 is simply because children gave a high rating for any teacher, then we would see successful distinction between the two toy teachers in Experiment 3. However, children failed to rate the underinformative teacher lower than the informative teacher (informative teacher $M(SD) = 11.67(8.22)$ versus underinformative teacher M(SD) = 13.61(7.43), $t(17) = 1.44$, $p = .167$) suggesting that simply giving a high rating to a teacher does not help children succeed.

Given the high exclusion rate, we conducted an additional exploratory analysis as in Experiments 1 and 2, including the seven children who failed to pass the inclusion criteria. In this larger data set, a trend that was insignificant in the main analysis became statistically significant; the average rating for the underinformative teacher was *higher* than the informative teacher (informative teacher $M(SD) = 14.2(7.0)$ versus underinformative teacher M(SD) = 11.8(7.6), $t(24) = 2.213$, $p = .037$). Although we did not predict this reversal prior to this analysis, this pattern is not inconsistent with our overall hypothesis. Note that children in Experiment 3 first rated the correct teacher (who simply named familiar objects) followed by the underinformative teacher (who demonstrated a toy that had four interesting functions). This might have highlighted the novelty of the toy and its interesting causal effects, arguably making the next teacher (informative teacher who demonstrated a toy with just one function) relatively less attractive or even considered "less helpful" for some children. Although explanation data could be potentially helpful, only a few offered justifications ($N = 3$) and thus were not analyzed. Therefore, this remains a speculation, and we remain cautious about interpreting the unpredicted effect in this exploratory analysis.

Most importantly, the failure in Experiment 3 (compared to their success in the informative first condition in Experiment 2) again reveals a striking limitation in young children's sensitivity to informativeness. Unlike older children, younger children (4- and 5-year-olds in our sample) failed to recognize and evaluate underinformative pedagogy without first seeing a fully informative demonstration (Figure 3).

Experiment 4

Results from Experiment 2 showed that young children successfully detect and evaluate sins of omission under certain conditions. More specifically, when they have had prior experience with a fully

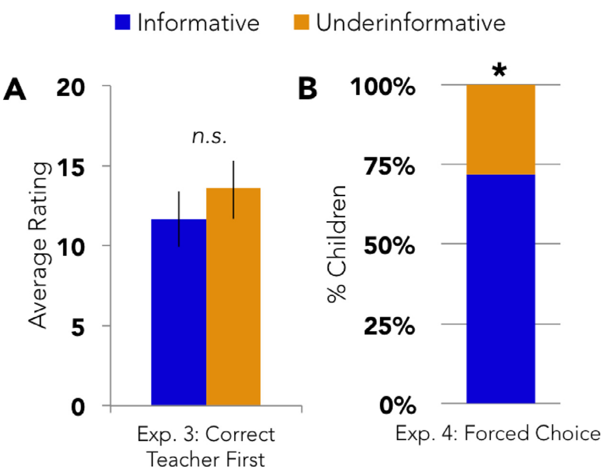

Figure 3. Results from Experiments 3 and 4. $*p < .05$. $**p < .01$. $***p < .001$. In Experiment 3, children rated the correct teacher first and then rated the underinformative teacher, followed by the informative teacher.

informative teacher, even 4-year-olds successfully penalized the underinformative teacher in their ratings. Furthermore, results from Experiment 3 provide suggestive evidence that it is the observation of a fully informative demonstration rather than the process of explicit rating that helps children succeed. If children's success depends on their understanding of what the teacher could have done, then presenting young children with binary choice between the informative and underinformative teachers would also lead to success even without the experience of explicitly rating the teachers. Observing two teachers back to back should create a clear contextual contrast, and children should succeed regardless of the teacher order. We test this hypothesis in Experiment 4. Furthermore, we limited our age range to just 4-year-olds to provide conclusive evidence that 4-year-olds can successfully evaluate sins of omission given enough contextual support.

Method

Subjects

Thirty-two 4-year-olds were recruited from a university-affiliated nursery school (18 girls; M_{age} (SD) = 4.43(0.27), range = 3.99–4.97; data collection: January 2015–September 2015). An additional two children were tested but were excluded due to technical errors (N = 1) or failing to report the correct number of functions for each toy (N = 1).

Materials

Stimuli were identical as in Experiments 1–3.

Procedure

Procedures were similar to Experiments 1–3, except that instead of rating after each trial, children first watched both trials (informative and underinformative, order counterbalanced) and then were asked to choose between the two teachers. The experimenter placed the toy teacher puppets on the table, equidistant from the child, and asked "Who did a better job teaching Elmo? Paul or Bill?" Children indicated their choice by pointing, touching, or saying the puppet's name.

Results and Discussion

When asked who was more helpful, 72% of children chose the informative teacher over the underinformative teacher (23 of 32; p = .020, binomial test). We found no difference between the children who saw the informative teacher first (10 of 15, 67.7%) and the children who saw the underinformative teacher first (13 of 17, 76.4%). Among those who made the correct choice, 14 provided verbal justifications and 6 of these children referred explicitly to the completeness of teaching or number of functions taught; among those who were inaccurate, only 1 of 9 children erroneously said the (chosen) teacher showed more things and the rest provided irrelevant answers (e.g., "I just know it"). These results suggest that given a clear contrast between fully versus underinformative teachers (i.e., having seen both teachers back to back), even 4-year-olds reliably preferred teachers who provide complete information.

General Discussion

Across four experiments, we tested children's ability to evaluate "sins of omission"; true but underinformative pedagogy that can mislead the learner. We asked whether children provide lower ratings to a teacher who demonstrates one function of a four-function toy (thus providing an underinformative demonstration) than to a teacher who demonstrates one function on a single-function toy (thus providing a fully informative demonstration). Our results (a) replicated earlier work (Gweon et al., 2014) showing that 6- and 7-year-olds successfully rate the underinformative teacher lower than the fully informative teacher using a new task, (b) revealed the same pattern of evaluation in 4- and 5-year-olds, but *only when* the fully informative teacher was evaluated before the underinformative teacher, and (c) provided further evidence with 4-year-olds, showing that they prefer the fully informative teacher over the underinformative teacher after sequentially observing the two teachers. Collectively, these results suggest that children as young as 4 are capable of evaluating sins of omission, although their competence is revealed only under certain conditions.

The striking effect of teacher order suggests that young children are not simply more tolerant of underinformativeness. We considered this possibility given recent work suggesting that 5- and 6-year-olds' fail to reject underinformative utterances because they are more generous toward pragmatic infelicities (Katsos & Bishop, 2011). If children were simply tolerant of pragmatic violations, their ratings would not have been influenced by order; instead, we would have seen a deficit in penalizing the underinformative teacher in both conditions. Although suggestive, children's explanations in Experiment 2 also provide further support that younger children's tendency to evaluate teachers with respect to their informativeness is influenced by the order of teachers. However, it remains possible that the reluctance to reject or penalize others further complicates their ability to evaluate sins of omission.

The pattern of our data is more consistent with the possibility that children's difficulty evaluating sins of omission is due to their difficulty understanding what it means for a teacher to be "helpful" in the context. Our hypothesis was motivated by recent work showing children's limited success in scalar implicature tasks, suggesting that children's failure to reject underinformative scalar expressions (e.g., use of "some" when "all" applies better) might come from their inability to spontaneously consider the relevant alternative term the speaker could have used (Barner et al., 2011; Foppolo, Guasti, & Chierchia, 2012; Skordos & Papafragou, 2016; Horowitz, Schneider, & Frank, in review). Possible reasons for this difficulty in these tasks include a lack of lexical knowledge as well as a general inferential failure to generate and consider relevant alternative representations.

In particular, a recent study showed that young children's sensitivity to pragmatically underinformative scalar expressions (e.g., "*Some* of the blickets have a crayon" when all of the blickets have a crayon) is modulated by their prior experience with relevant versus irrelevant uses of a stronger alternative item "all" (Skordos & Papafragou, 2016). More specifically, 5-year-olds were able to reject such underinformative sentences only when they have first rejected a sentence "*All* of the blickets have umbrellas" as a description of a picture in which only three of four blickets have an umbrella (use of "all" is wrong due to quantity) but not when they rejected the same sentence as a description of a picture in which all of the blickets had a shovel (use of "all" is wrong due to the kind of objects). Similarly, simply accepting a correct use of "all" did not facilitate subsequent rejection of an infelicitous "some," but evaluating various infelicitous uses of "all" and "some" improved children's later judgments (Foppolo et al., 2012). These results resonate with the strong effect of teacher order in our findings, adding weight to the idea that the relevant experience of a more appropriate alternative matters for children's sensitivity to pragmatically underinformative expressions, regardless of whether the inferences are drawn from verbal utterances or pedagogical demonstrations.

These parallels, although intriguing, open further questions about how children interpret and evaluate pedagogical actions. Why does prior experience with a fully informative teacher help children evaluate an underinformative teacher? First, seeing a teacher who demonstrates a toy's function might simply help children detect that another teacher left out some functions while demonstrating a different toy. This might have led children to recognize and penalize the incompleteness of the demonstration itself, in the absence of an understanding that incompleteness can be misleading in pedagogical contexts. Second, because there are many possible ways in which a teacher could be unhelpful (e.g., inaccurate, incomplete, uninteresting, clumsy, too slow, too fast, etc.), seeing an example of a teacher who provides complete information about one toy might help children understand that the currently relevant dimension for comparison or evaluation is the completeness of the demonstration. This might allow children to interpret the incomplete demonstration as potentially misleading and undesirable, and less helpful than what it could have been.

Although our study does not directly tease these possibilities apart, 4-year-olds' robust preference for the fully informative teacher (Experiment 4) provides indirect evidence that even the children in the youngest age group successfully encoded and retained the teachers' actions regardless of the order. Thus, the benefit of seeing the fully informative teacher might be less about helping children notice the omission per se but more about *recognizing and interpreting* the omission as a sin. In particular, having a relevant representation of a fully informative teacher might facilitate comparisons with other teachers on the dimension of informativeness, helping children evaluate the underinformative teacher's actions in light of possible alternatives that could have made his teaching more informative. A body of work on the development of higher order relational reasoning suggests that the process of comparing and categorizing various perceptual figures facilitates children's ability to detect and evaluate abstract, relational similarities between these figures (e.g., Christie & Gentner, 2010; Kotovsky & Gentner, 1996). Similarly, prior experience with a fully informative teacher might have allowed children to better evaluate the subsequent teacher in light of the fully informative one, leading them to successfully penalize his sin of omission. Children's explanations in both age groups mostly referred to the incompleteness of teaching (a few mentioned teacher knowledgeability), and almost none mentioned moral (e.g., "he lied") or trait (e.g., "he was mean") inferences that go beyond informativeness. However, the exact nature of the representations and the development of richer inferences about moral or dispositional traits of the teachers remain an interesting question for future work.

We note at least three different reasons for young children's difficulty representing relevant alternatives in our task. First, a general limitation in representational capacity might have prevented children from considering the full range of possible alternatives. Indeed, researchers have proposed theoretical connections between pragmatic competence and representational capacity (e.g., Foppolo et al., 2012; see also Gopnik & Rosati, 2001).

Second, the social-evaluative nature of our task speaks to the possible role of Theory of Mind and an abstract understanding of informative communication in generating relevant, alternative goal-directed actions. That is, children might have trouble understanding what other demonstrations the teacher could have provided given his intent and knowledge. For instance, if a teacher actually knew only one of the four functions of a toy, he could not have possibly provided more than one demonstration; recent work shows that adults, and even 4- to 6-year-olds exonerate teachers who provide underinformative demonstration due to such epistemic constraints (Bass, Hawthorne, Goodman, & Gweon, 2015; Bass, Bonawitz, & Gweon, 2017). However, in Bass et al. (2017), children had already observed and rated a fully knowledgeable and informative teacher; it would be interesting to examine the developmental trajectory of this "epistemic pardon", and whether children's performance would be related to other theory of mind measures.

Finally, young children might have struggled with a more local problem of generating what a "good" teacher is like in the specific context of the current task, rather than suffering from a general representational limit or a weak understanding of informative communication. Just as lexical knowledge play a key role in children's ability to compute scalar implicature (e.g., Barner et al., 2011, Horowitz et al., in review), knowledge of concrete action repertoires available within situational constraints might also be critically important for interpreting and evaluating others' actions.

Importantly, these possibilities are not necessarily mutually exclusive; any and all of these possibilities might have contributed to young children's limited ability to evaluate sins of omission. In light of these possibilities, we note one interesting aspect of our data; the effect of order appeared rather abruptly between ages 5 and 6, with little evidence for a gradual change between these ages. Collapsing across children who saw the fully informative teachers first in Experiments 1 and 2, the difference in ratings between the two teachers remained stable and robust across ages 4–7. This suggests a possibility that the sensitivity to underinformativeness itself is present even in the youngest children, but a sudden change between ages 5 and 6 facilitated children's teacher evaluation. One explanation is that experience with formal schooling experience helps children spontaneously understand that the most relevant and defining property of helpful pedagogy is informativeness (e.g., complete demonstration of a toy's functions). Although we did not collect information about children's formal schooling experience, further investigating this possibility would help explain the reason behind the sudden shift that occurs during this period. Indeed, further research is required to understand the exact nature of children's limitations. Nevertheless, our results provide meaningful support for the idea that pragmatic competence, in both verbal communication and in social inferences more generally, is rooted in both content-based knowledge of available words and actions and the ability to represent alternative states of others (both their observable behaviors and their unobservable mental states).

Evaluation of pragmatically infelicitous sentences has traditionally been a topic in linguistics; evaluation of "sins of omission," on the other hand, is considered a topic in social learning and pedagogical reasoning that started to gain interest more recently (Gweon et al., 2014). Although rather different at a superficial level, both can be construed as failures to conform to Grice's cooperative principles (particularly *maxim of quantity*, Grice, 1975; see also Horn, 1984). In other words, both are examples of underinformative, and therefore potentially misleading, behaviors that violate the expectation to be helpful for the partner in a given communicative context. Our data suggest that similar representational constraints might underlie children's difficulty in these two seemingly different domains, providing preliminary empirical support for the theoretical proposal made by Baldwin et al. (2008). More specifically, the inferences drawn from goal-directed, intentional actions might follow similar pragmatic principles that guide our inferences from verbal communicative behaviors. These results build upon recent discoveries of young children's ability to draw rich, sophisticated inferences from intentional, goal-directed actions of others (e.g., Bonawitz et al., 2011; Butler & Markman, 2012; Gweon, Tenenbaum, & Schulz, 2010; Shneidman et al., 2016) by showing that children evaluate others via the same principle they use for drawing such inferences.

By identifying constraints in children's evaluation that parallel those found in implicature tasks, the current work provides initial steps toward drawing more explicit connections between pragmatic inferences from verbal and nonverbal behaviors. In future work, it would be intriguing to find parallel successes and failures in scalar implicature tasks and our "sins of omissions" task within the same group of participants, as well as the role of representational capacity and theory of mind in children's performances. We look forward to a body of future work that aims to provide a unified view of pragmatic inferences, finding common social and cognitive capacities that support inferences from both verbal communication and goal-directed actions.

Finally, our results have implications for methodological limitations in typical developmental studies in detecting and interpreting developmental changes. In the current study, we modified our previous single-trial task (Gweon et al., 2014) to specifically test the hypothesis that accessing alternatives is requisite to younger children's appropriate evaluations of underinformative demonstrations. In the absence of this prediction, one could have used a between-subject paradigm to test this younger population and concluded that 4- and 5-year-olds "do not yet recognize sins of omission." We were able to detect this competence only by asking children to evaluate both teachers. However, repeated questioning comes with its own hazards (e.g., Gonzalez, Shafto, Bonawitz, & Gopnik, 2012; Poole & White, 1991), and it is not always the ideal design in many developmental studies. Our findings highlight the importance of carefully considering the limitations of experimental designs particularly when drawing conclusions about developmental trajectory.

Even from early in life, humans constantly communicate with and learn from each other. Although the format of information and the modalities by which we communicate might vary across contexts, we are always tuned to others' intentions and knowledge, and we attempt to infer what others mean by going beyond the evidence. From this perspective, violations of pedagogical sampling and violations of scalar implicature are both failures to conform to what we expect of others' behaviors. The current study provides an important first step toward providing empirical links between some of the most distinctively human behaviors: teaching, learning, and communication.

References

Baldwin, D., Loucks, J., & Sabbagh, M. (2008). Pragmatics of human action. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 96–129). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195188370.003.0006

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84–93. doi:10.1016/j.cognition.2010.10.010

Bass, I., Bonawitz, L. B., & Gweon, H. (under review). *Didn't know, or didn't show? Preschoolers consider epistemic state and degree of omission when evaluating teachers*. Proceedings of Cognitive Science Society.

Bass, I., Hawthorne, D., Goodman, N. D., & Gweon, H. (2015). Not by number alone: The effect of teacher's knowledge and its value in evaluating "sins of omission." In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 166–171). Austin, TX: Cognitive Science Society.

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*, 1018–1034. doi:10.1016/j.cognition.2007.12.008

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*, 322–330. doi:10.1016/j.cognition.2010.10.001

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*, 1416–1428. doi:10.1111/j.1467-8624.2012.01775

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*, 356–373. doi:10.1080/15248371003700015

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*, *8*, 365–394. doi:10.1080/15475441.2011.626386

Gonzalez, A., Shafto, P., Bonawitz, E., & Gopnik, A. (2012). Is that your final answer? The effects of neutral queries on children's choices. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1614–1619). Austin, TX: Cognitive Science Society.

Gopnik, A., & Rosati, A. (2001). Duck or rabbit? Reversing ambiguous figures and understanding ambiguous representations. *Developmental Science*, *4*, 175–183. doi:10.1111/1467-7687.00163

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 3492–3496. doi:10.1073/pnas.1515129113

Grice, H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics Vol. 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*, 335–341. doi:10.1016/j.cognition.2014.04.013

Gweon, H., & Schulz, L. E. (2011). 16-Month-olds rationally infer causes of failed actions. *Science*, *332*, 1524. doi:10.1126/science.1204493

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9066–9071. doi:10.1073/pnas.1003095107

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q- based and R-based implicature. In D. Schiffrin (Ed.). *Meaning, form, and use in context: Linguistic applications* (Vol. *11-42*). New York: Academic Press.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (under review). The trouble with quantifiers: Exploring children's deficits in scalar implicature.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*, 1723–1739. doi:10.1037/a0016704

Jaswal, V., & Neely, L. (2006). Adults don't always know best. *Psychological Science*, *17*, 757–758. doi:10.1111/j.1467-9280.2006.01778.x

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67–81. doi:10.1016/j.cognition.2011.02.015

Koenig, M., Clément, F., & Harris, P. (2004). Trust in testimony. *Psychological Science*, *15*, 694–698. doi:10.1111/j.0956-7976.2004.00742.x

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

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*, 2797–2822. doi:10.1111/j.1467-8624.1996.tb01889.x

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165–188. doi:10.1016/S0010-0277(00)00114-1

Papafragou, A. (2006). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language*, *33*, 721–757. doi:10.1017/S0305000906007550

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253–282. doi:10.1016/S0010-0277(02)00179-8

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226. doi:10.1037/0012-1649.43.5.1216

Poole, D., & White, L. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology*, *27*, 975–986. doi:10.1037/0012-1649.27.6.975

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*, 341–351. doi:10.1177/1745691612448481

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. https://doi.org/10.1016/j.cogpsych.2013.12.004

Shneidman, L., Gweon, H., Schulz, L. E., & Woodward, A. L. (2016). Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development*, *87*, 723–735. doi:10.1111/cdev.12502

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18. doi:10.1016/j.cognition.2016.04.006

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*, 176–190. doi:10.1080/15475441.2014.927328