# Knowing what he could have shown: The role of alternatives in children's evaluation of under-informative teachers

**Hyowon Gweon** (gweon@stanford.edu) **Mika Asaba** (masaba@stanford.edu)
Department of Psychology, Stanford University

**Abstract**

What underlies young children's failure in evaluating under-informative teachers? We explore the hypothesis that children have difficulty representing relevant alternatives; knowing what the teacher could have done. Children rated two teachers who demonstrated toys to a naïve learner. One group first observed a fully informative teacher and then an under-informative teacher, while the other group saw the reverse order. Six- and seven-year-olds successfully rated the under-informative teacher lower than the fully-informative teacher regardless of the order (Exp.1). However, four- and five-year-olds showed this pattern only when they saw the fully informative teacher first (Exp.2). Given a binary choice after seeing both teachers, four-year-olds showed a preference for the fully informative teacher (Exp.3). We discuss these results in light of recent literature on children's understanding of pragmatic violations in linguistic communication; the contrast between the fully informative vs. under-informative teachers might help children understand what the teacher could have shown.

**Keywords:** cognitive development, pragmatics, scalar implicature, pedagogical reasoning, Theory of Mind

Nora arrived at the gym and noticed a shiny new device. Her trainer, Jim, proudly showed her how to measure weight on the device. Left unimpressed, she thought, "That is nothing but a fancy-looking scale!" Much to her surprise, she discovered that the device was a state-of-the-art body composition analyzer with many useful functions. Later, as Nora walked over to the treadmill area, Jim called out, "Some of the treadmills are broken!" She kept walking over, intending to use a machine that wasn't broken, only to find out that in fact, all of the treadmills were broken. Nora shook her head and fumed in frustration. Where did Jim go wrong?

Jim's first mistake was a violation of pedagogical sampling. In pedagogical contexts, learners expect a knowledgeable, helpful teacher to select and provide the evidence that is likely to increase the learner's belief in the correct hypothesis (Shafto, Goodman, & Frank, 2012). He failed to understand that demonstrating *one* function would imply that there's *only one* function, because Nora considers him to be a knowledgeable, helpful trainer. By showing how to measure weight but omitting other functions, he provided under-informative pedagogy that misled Nora to believe that the new device is just a scale. In other words, Jim committed a "sin of omission".

Jim's second mistake was a violation of pragmatic implicature; he failed to understand that saying "*some* of the treadmills" implies "*not all* of the treadmills". Because the use of an intermediate item on an implied scale (e.g., some treadmills) suggests that a stronger alternative (e.g., all treadmills) does not apply, his statement was logically true but nonetheless misleading.

At a quick glance, these two mistakes might seem rather different; one is about what he did and did not show, and the other is about what he did and did not say. However, in both cases, Jim failed to conform to Grice's cooperative principle, particularly the *maxim of quantity* (Grice, 1975); see also Horn, 1972, 1984). While Nora reasonably expected Jim to understand the basics of pragmatics in communication, he did not provide sufficient or strong enough information for Nora to draw accurate conclusions in both cases. Given these two transgressions, she should be rightly irritated with Jim, because as adults, we expect others to provide information that is requisite to the communicative context.

When learning from people who are expected to be knowledgeable and helpful, even very young children readily draw inferences that go beyond the face value of the provided evidence. For instance, when a teacher demonstrates *one* interesting function of a novel toy, preschoolers infer that the toy has *only one* function(Bonawitz et al., 2011). This inference rests on the assumption that the data were selected to be maximally helpful; if the toy had additional functions, the teacher would have chosen to show those, too. A more recent study has shown that six- and seven-year-old children are sensitive to teachers who violate pedagogical sampling and provide under-informative evidence (Gweon, Pelton, Konopka, & Schulz, 2014). In this study, older children observed a puppet teacher demonstrate one interesting function of a toy for a naïve learner. One group of children (Teach 1/1) saw a fully informative teacher, because the toy actually had just one function. Another group (Teach 1/4) saw the same teacher who demonstrated the same function, but his teaching was under-informative because the toy in fact had three additional functions. Critically, the appearance of the toy and the teacher's behaviors were identical across conditions. However, children's rating for the under-informative (Teach 1/4) teacher was significantly lower than children's rating for the fully informative teacher (Teach 1/1). Thus these results suggest that by around age 6, children appropriately evaluate Jim's presentation of the fancy new device; although his demonstration was true of the device, it was incomplete and misleading.

Prior developmental work on pragmatic implicature suggests that 6- and 7-year-olds can also evaluate Jim's statement about treadmills as misleading. However, a number of studies showed that children under 6 years of age have difficulty recognizing such pragmatic violations. For instance, young children consider the sentence "the boy ate *some* cookies" as acceptable even when in fact, he ate *all* the cookies (Noveck, 2001; Papafragou & Musolino, 2003; Huang & Snedeker, 2009; Papafragou, 2006), failing to reject logically true but pragmatically infelicitous statements that are analogous to

Jim's comment about treadmills. More recent studies have suggested that younger children's (4- and 5-year-olds) failures may arise from difficulties aside from a genuine inability to draw pragmatic implicature (Barner, Brooks, & Bale, 2011; Skordos & Papafragou, 2014; Katsos & Bishop, 2011).

If the ability to evaluate Jim's two mistakes - violations of pedagogical sampling and violations of pragmatic implicature - depends on the same cognitive capacities, we might expect that children under age 6 would also experience difficulty recognizing differentially evaluating teachers based on informativeness. In the current study, we investigate young children's ability to evaluate under-informative pedagogy, or "sins of omission".

In light of this prior work on children's successes and failures in pragmatic implicature tasks, we can consider a few possibilities. First, young children might have no trouble recognizing and evaluating under-informative pedagogy. Once they expect teachers to engage in pedagogical sampling and constrain their inferences accordingly (Bonawitz et al., 2011), the ability to evaluate those who violate this expectation might come for free. This would suggest little commonality between children's ability to detect under-informative pedagogy and pragmatically infelicitous utterances.

Second, young children might fail to accurately evaluate sins of omission because they are simply more tolerant of under-informative pedagogy. Katsos and Bishop (2011) showed that even though 5- and 6-year-olds fail to reject pragmatically infelicitous speakers (e.g., "the boy ate some of the cookies" when he in fact ate all the cookies), they reward such speakers less than fully informative ones given a 3-point scale. Thus children who observe under-informative pedagogy might feel less inclined to penalize the teacher, even though they notice the violation of pedagogical sampling.

Finally, we consider the possibility that young children fail to evaluate under-informative pedagogy because they have difficulty representing relevant alternative actions; that is, understanding what else the teacher could have done. A related body of work comes from recent developmental literature on children's sensitivity to under-informative utterances, showing that even 4- and 5-year-olds successfully reject pragmatically infelicitous utterances when the relevant alternatives are made salient in context (e.g., the speaker says, "only the cat and the dog are sleeping" rather than "some of the animals are sleeping" given a picture of a cow, a cat, and a dog sleeping; Barner et al., 2011, see also Skordos & Papafragou, 2014). Similarly, in order to recognize sins of omission in pedagogical contexts, young children might require a clear representation of what the teacher could have shown in addition to his under-informative demonstration.

In order to explore these hypotheses, we designed a task in which children observed a fully-informative teacher and an under-informative teacher and rated them sequentially. In Experiment 1, we first replicated 6- and 7-year-olds' ability to evaluate under-informative teachers (Gweon et al., 2014) using a novel within-subjects paradigm. In Experiment 2, we used this task with two separate groups of 4- and 5-year-olds to see if younger children's evaluations are affected by order. If sensitivity to under-informative pedagogy is independent of the ability to detect other violations of pragmatic implicature, younger children might succeed on this task regardless of the order. If younger children simply show higher tolerance to under-informativeness, children would fail to differentiate the two teachers regardless of the order. However, if knowing the alternative helps, the order in which children observed the two teachers would have a significant impact on children's performance. That is, children would successfully provide lower ratings for the under-informative teacher if they had seen the fully informative teacher first (because this would help them represent what the under-informative teacher could have done), but not when they see the under-informative teacher first. To further explore the role of contrastive alternatives in children's performance, in Experiment 3 we presented 4-year-olds with a direct contrast between the two teachers by using a binary choice paradigm.

## Experiment 1

In Experiment 1, we used a modified version of Gweon et al. (2014)'s task to (a) replicate its findings with the same age group, 6- and 7-year-old children, and (b) test whether there is an effect of teacher order. This modified version allowed us to get a pair of ratings from each child, one for a teacher who commits a sin of omission by demonstrating just one of four functions of a toy, and another for a teacher who still shows just one function of a different toy that only has one function. By presenting the two teachers in different orders, we asked whether children's relative ratings of the two teachers are influenced by the order in which they were presented.

### Methods

**Subjects** Twenty-eight six- and seven-year-olds were recruited from a local museum, $M_{Age}$(SD): 7.05(.54), range: 6.07 - 7.90, and were randomly assigned to the Teach 1/1 First condition (N = 14) or Teach 1/4 First condition (N = 14). An additional six children were dropped from analysis due to experimental error (N = 2), failure to report the number of functions on the toy (N = 2), or failure to rate the Incorrect Teacher lower than the Correct Teacher (N = 2).

**Materials** Stimuli were presented as videos on a 13-inch Macbook Pro using MATLAB and PsychToolBox. Two custom-made toys, one yellow and one gray, were used in the videos. The yellow toy 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. The gray toy also had four causal affordances; pressing a purple tab made a beeping sound, pressing a grey 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. Although both toys had four functions, each toy could be presented as

a toy with four functions (Four-Function Toy in the Teach 1/4 trial) or presented as if it just had a single function (One-Function Toy in the Teach 1/1 trial). The type of toy was counterbalanced throughout; half the children saw the yellow toy in the Teach 1/1 trial and the gray toy in the Teach 1/4 trial, while the other half saw the reverse. Four hand puppets were used as the Toy Teachers (Paul and Bill) and Incorrect and Correct Teachers (Sally and Laura). 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, and along with faces that varied from frowny to smiley to serve as anchor points between the sections.

**Procedure** The procedure was similar to Experiment 1 in Gweon et al. (2014), except that the stimuli were presented as videos rather than live action, and that participants observed and rated two Toy Teachers sequentially. Participants were tested in a quiet room inside the museum. All participants received a brief training with the rating scale before the procedures began. Each child saw two trials, one Teach 1/1 trial, and one Teach 1/4 trial, in varying orders depending on the condition. Each trial consisted of three phases: Exploration, Teaching, and Rating (see Figure 1).

(1) Exploration: Children first watched a video of an adult exploring the toy. In the Teach 1/1 trial, the adult said,"I wonder what this toy does!" and discovered one function of the toy (wind-up mechanism on the yellow toy; beep on the gray toy), while acknowledging that other parts don't do anything. At the end of the video, she exclaimed, "This toy does one thing!" In the Teach 1/4 trial, the adult discovered all four functions of the toy, and said "This toy does four things!" [1]. 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 children (N = 2) who were unable to report the correct number of functions after the second viewing.

(2) Teaching: Children then watched a video of a Toy Teacher teaching Elmo about the toy from 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 Teach 1/1 and Teach 1/4 trials, children watched the Toy Teacher demonstrate just one function. Thus the Toy Teacher was fully informative in the Teach 1/1 trials and under-informative in the Teach 1/4 trials.

(3) Rating: The experimenter then brought out the scale and asked the child,"How helpful was Paul (Bill) in teaching Elmo?" The participant indicated his or her response by placing the marker on the rating scale. Then the same procedure (Exploration, Teaching and Rating) was repeated with the other trial. The type of toy and the trial type was counterbalanced throughout. One toy was always taught by Paul, and the other toy was always taught by Bill. Children in the Teach 1/1 First condition saw the Teach 1/1 trial first and then saw the Teach 1/4 trial; Children in the Teach 1/4 First condition saw the reverse order.

After rating Toy Teachers, children watched two more teacher puppets teach Elmo the names of simple household objects. The Correct Teacher provided the correct names of two objects (referring to a stuffed carrot as a 'carrot' and a toy plane as a 'plane'). The Incorrect Teacher provided the incorrect names of two different objects (referring to a ball as a 'cup', and a stuffed tiger as '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, such that half of the children in both the Teach 1/1 First and Teach 1/4 First conditions first viewed the Correct teacher first and the other half viewed the Incorrect teacher first. These additional ratings were collected to identify children who did not understand the difference between true and false information or those who did not yet understand how to use the rating scale.

### Results and Discussion

We first asked whether our results replicated Gweon et al. (2014) by comparing the average ratings for the Teach 1/1 and Teach 1/4 trials, collapsing across conditions. In line with previous findings, 6- and 7-year-olds gave a higher rating for the Toy Teacher in the Teach 1/1 trials than the Teach 1/4 trials (Teach 1/1 vs. Teach 1/4: M(SD) = 15.30(4.73) vs. 8.25(5.77), $t(27) = 5.851, p < .001$).

To ask whether the order of the Toy Teachers affected ratings, we performed a 2(Trial: Teach 1/1, Teach 1/4) by 2(Condition: Teach 1/1 First, Teach 1/4 First) mixed 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) = 33.013, $p < .001$, $\eta_p^2$ = .56, no main effect of Condition, F(1,26) = .294, $p = ns$, and no interaction between Condition and Trial, F(1,26) = .036, $p = ns$.

These results replicate findings from Gweon et al. (2014), showing that that by age 6, children reliably detect sins of omission in pedagogical contexts and appropriately evaluate teachers depending on whether they provided fully informative or under-informative information.

## Experiment 2

In Experiment 2, we ask whether younger children (4- and 5-year-olds) also show the same pattern as older children, or whether their ratings are influenced by the order in which they observed and rated different teachers.

### Methods

**Subjects** Thirty-two 5-year-olds ($M_{Age}$(SD) = 5.45(.30), range = 5.00-5.95) and thirty-two 4-year-olds ($M_{Age}$(SD) = 4.51(.30), range = 4.07 - 4.99) were recruited from a local museum or a University-affiliated nursery school (N=16 in each condition). Following Gweon et al. (2014), an additional

[1] We ensured that the adults did not deliver any pedagogical cues; they initially claimed to be naïve about the toy, their utterances were self-directed, and they never made eye contact with the camera.

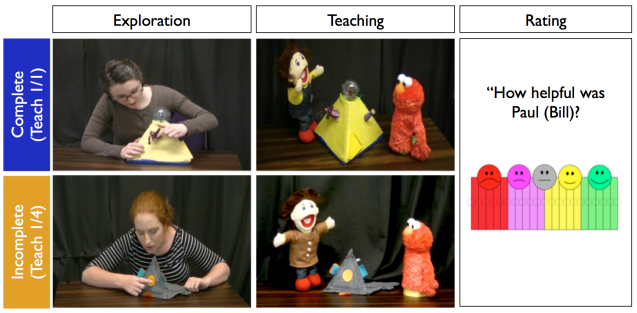


Figure 1: Procedures in Exp. 1 & 2 and the rating scale.

22 4-year-olds and 12 5-year-olds were excluded from analysis because they rated the Incorrect Teacher the same as or higher than the Correct teacher.

**Materials** The stimuli were identical as in Exp. 1.

**Procedure** Five-year-olds were trained on the same 21-point rating scale as in 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 that effectively converted the 21-point scale to a 5-point scale (for 4-year-olds only). These scores were then 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). The procedures were identical.

## Results and Discussion

First, we looked at ratings for the Toy Teachers by collapsing across conditions in each age group. 4-year-olds and 5-year-olds rated the Toy Teacher higher in the Teach 1/1 trials than in the Teach 1/4 trials (4-year-olds: Teach 1/1 vs. Teach 1/4: M(SD) = 16.91(5.37) vs. 12.97(7.81), $t(31) = 2.83, p = .007$; 5 yr olds: Teach 1/1 vs. Teach 1/4: M(SD) = 15.78(5.26) vs. 12.81(7.57), $t(31) = 2.61, p = .014$).

For each age group, we asked whether condition (order of trials) affected children's ratings. We performed a 2(Trial: Teach 1/1, Teach 1/4) by 2(Condition: Teach 1/1 First, Teach 1/4 First) mixed ANOVA with Trial as a within-subjects factor and Condition as a between-subjects factor. For 5-year-olds, the results revealed a significant effect of Trial, F(1,30) = 8.41, $p = .007$, $\eta_p^2$ = .22, no effect of Condition, F(1,30) = .969, $p = ns$, and a significant interaction between Condition and Trial, F(1,30) = 8.231, $p = .007$, $\eta_p^2$ = .22 . We saw the same pattern in 4-year-olds, with a significant effect of Trial, F(1,30) = 9.26, $p = .005$, $\eta_p^2$ = .24, no effect of Condition, F(1,30) = .122, $p = ns$, and a significant interaction between Condition and Trial, F(1,30) = 4.52, $p = .042$, $\eta_p^2$ = .13.

Planned comparisons revealed that 5-year-olds in the Teach 1/1 First condition gave higher ratings in the Teach 1/1 trial than the Teach 1/4 trial (Teach 1/1 vs. Teach 1/4: M(SD) = 16.28(5.94) vs. 10.31(7.33), $t(15) = 4.22, p = .001$). However, 5-year-olds in the Teach 1/4 First condition failed to show this distinction (Teach 1/1 vs. Teach 1/4: M(SD) = 15.28(4.62) vs. 15.25(7.19), $t(15) = .021, p = ns$). Similarly, 4-year-olds in the Teach 1/1 First condition gave higher ratings in the Teach 1/1 trials than Teach 1/4 trials (Teach 1/1 vs. Teach 1/4: M(SD) = 17.94(5.21) vs. 11.25(8.27), $t(15) = 3.14, p = .007$), whereas 4-year-olds in the Teach 1/4 First condition did not show this pattern (Teach 1/1 vs. Teach 1/4: 15.88(5.50) vs. 14.69(7.18), $t(15) = .81, p = ns$).

Finally, we compared the ratings between Teach 1/1 and Teach 1/4 teachers using just the first trials, as if the children were run in the original single-trial paradigm used in Gweon et al. (2014). For both 4-year-olds and 5-year-olds, we did not see a significant difference in children's ratings for fully vs. under-informative teachers (4-year-olds: Teach 1/1 vs. Teach 1/4: M(SD) = 17.94(5.21) vs. 14.69(7.18), $t(27) = 1.47, p = ns$, 5 yr olds: Teach 1/1 vs. Teach 1/4 : M(SD) = 16.28(5.94) vs. 15.25(7.19), $t(28) = .44, p = ns$).

These results suggest that 4- and 5-year-olds' ratings were highly influenced by the order in which they saw different trials. When children first saw a fully-informative Toy Teacher, children were able to give a lower rating for the under-informative Toy Teacher, but those who saw the reverse order failed to penalize the under-informative Toy Teacher.

One possible interpretation is that the results reflect 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 understood that the Teach 1/1 teacher was better than the Teach 1/4 teacher, they might have been unable to rate him higher simply because they had first given a very generous rating to the Teach 1/4 teacher. If this is the case, then children might be willing to provide a high rating even for a teacher who provides obviously false information. While unlikely, we addressed this alternative with an independent group of 4- and 5-year-olds ($M_{Age}$(SD) = 5.22(.52), range = 4.37 - 5.97) and asked children to rate a teacher who provided false information (Incorrect Teacher). We compared the average rating for this Incorrect Teacher to the average rating for the Teach 1/4 Toy Teacher in the Teach 1/4 First condition. Children who rated the Incorrect Teacher first gave a significantly lower rating compared to the children who rated the 1/4 Toy Teacher first (Incorrect vs. Teach 1/4: M(SD) = 9.58(8.38) vs. 16.62(6.39), $t(17) = 2.56, p = .02$). Furthermore, while 3 of 12 children gave the highest possible rating to the Incorrect Teacher, 17 out of 24 children in the combined 4- and 5-year-old group gave the highest rating for the Teach 1/4 Teacher (Fisher's exact, p = .01)[2].

These results suggest that the younger children were anchored well on the scale, and the failure to distinguish fully informative and under-informative teachers in the Teach 1/4 First group was not due to a higher baseline rating. Children were able to appropriately penalize teachers who provided false information, but they struggled to penalize those who provided incomplete information.

[2]There were no differences between ratings to the Incorrect Teacher when the child saw this teacher first or when observed in the Teach 1/4 First condition, p = .196.

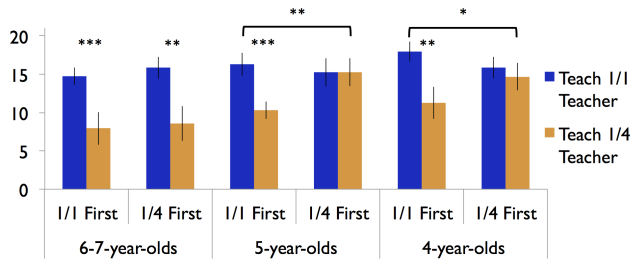


Figure 2: Results from Exp. 1 & 2 (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$).

## Experiment 3

Results from Experiment 2 suggest that although four- and five-year-olds appropriately evaluate sins of omission if, and only if, they saw the fully informative teacher first, and that this limitation is not due to a high baseline rating. Moreover, the data were inconsistent with the possibility that children's difficulty arise from their higher tolerance for under-informativeness; if this were true, children's ratings for the under-informative teacher would not have been affected by order. Our data are more consistent with the possibility that young children successfully evaluate under-informative pedagogy when the alternatives are made clear in the context (i.e., having seen a teacher who provides fully informative demonstration). In Experiment 3, we test this hypothesis with a more specific prediction. If alternatives matter, then presenting young children with binary choice between the teachers should also lead to success; watching two teachers back-to-back should create a clear contextual contrast, and children should succeed regardless of the teacher order.

### Methods

**Subjects** Twenty-three 4-year-olds were recruited from a University-affiliated nursery school, $M_{Age}$(SD) = 4.45 (.26), range: 4.00 - 4.97.

**Materials** Stimuli were identical as in Exp. 1 & 2.

**Procedure** Procedures were similar to Exp. 1 & 2, except that instead of rating after each trial, children first watched both trials (Teach 1/1 and Teach 1/4, order counterbalanced) and were asked to choose between the two teachers. The experimenter placed the actual Toy Teacher puppets on the table, equidistant from the child, and asked: "Who did a better job of teaching Elmo? Paul, or Bill?" Children indicated their choice by pointing, touching, saying the puppet's name.

### Results and Discussion

Children showed a clear preference for the Teach 1/1 teacher over the Teach 1/4 teacher (17 of 23; $p = .035$, by binomial test). Thus given a clear contrast between fully vs. under-informative teachers (i.e., having seen both teachers back-to-back), even four-year-olds reliably chose teachers who provide complete information.

## General Discussion

Across three experiments, our results show 4- and 5-year-olds' limited success in evaluating under-informative pedagogy. More specifically, children successfully evaluated under-informative teachers *only if* they had seen the fully-informative teacher first. These results are in stark contrast with older children's success (age 6 - 7) who reliably distinguished the two teachers regardless of order. However, given a clear contrast between the two teachers, even 4-year-olds showed a robust success in distinguishing the two.

These results are most consistent with the hypothesis that children's judgments of sins of omission are affected by the availability of alternatives (i.e., what the teacher could have shown), and suggests that the same constraints might underlie children's failures in evaluating violations of pragmatic implicature and their failures in evaluating under-informative teachers. Barner et al. (2011) suggests that children have difficulty with scalar implicature tasks because they do not know that the word *all* is an alternative to the word *some*; they cannot represent that the speaker could have said "all of the animals are sleeping" to describe a picture in which three of three animals are sleeping. Thus children even fail to reject false sentences like "*only* some of the animals are sleeping". However, when an equivalent sentence allows the relevant scale to be constructed from the contextual information, children reject the sentence "*only the dog and the cat* are sleeping". Similarly, our results suggest that seeing a teacher who provides a fully informative demonstration establishes a contrast that facilitates evaluation of the under-informative teacher; seeing the Teach 1/1 teacher first allows children to understand what the Teach 1/4 teacher could have done, making it easier for children to penalize him for his "sin of omission".

Our results also suggest that young children are not simply more tolerant of under-informativeness. Recent data suggest that 5- and 6-year-olds fail to reject under-informative utterances but appropriately reward under-informative speakers less than fully-informative ones (Katsos & Bishop, 2011), suggesting that the gradedness of the measure could mask children's performance. However, high tolerance of pragmatic infelicity alone cannot account for our data. If children were simply generous with pragmatic violations, younger children should have shown similar ratings for the under-informative teacher regardless of the order. Furthermore, we would have seen a linear increase with age in the difference between Teach 1/1 and Teach 1/4 ratings. These patterns were not clear in our data. However, our data do not conclusively rule out the possibility that a reluctance to reject or penalize others might further complicate their ability to detect sins of omission (as well as other violations of pragmatic implicatures), and further studies should study the role of different factors that contribute to children's failure in both domains.

Why are alternatives important in interpreting others' actions? Although some of our goal-directed actions are executed in well-established patterns, there are often many degrees of freedom in how we behave in the world. For instance,

a teacher might be not very helpful because he didn't know much, his voice is too soft, or because he demonstrates things too quickly. In the context of our experiment, young children might have trouble understanding that the completeness of his demonstration is the relevant dimension for evaluation; however, upon observing someone who provides full information about a toy, children might understand that informativeness, or completeness, is the dimension on which they should evaluate the teacher. Thus one interesting prediction is that the type of alternatives matter; for instance, even though it is still a "good" teacher, seeing a teacher who was simply correct about a toy's function should not help children evaluate under-informative teacher (see also Skordos & Papafragou, 2014, for related data on scalar implicature). Testing this prediction will further clarify the role of contrastive context and alternative representations in children's social evaluations.

Relatedly, results from Exp. 3 suggest that even 4-year-olds can encode and retain the teachers' actions regardless of the order; given binary choice, children showed a robust preference for the more informative teacher. Therefore, the role of alternatives (i.e., seeing the Teach 1/1 teacher first) seems to be in *interpreting* observed actions of others, rather than in deciding what to attend to or what to remember. This points to a possible importance of Theory of Mind – the ability to explain and predict others' actions in terms of their unobservable mental states – in our ability to generate and represent alternatives actions of others. Indeed, a key factor in the ability to compute scalar implicature is lexical knowledge (e.g., Barner et al., 2011), and similarly, knowledge of available action repertoires would be important for interpreting others' actions. However, the ability to represent an unobservable state of the world or another person's mind might also be a key factor in understanding our pragmatic competence (Foppolo, Guasti, & Chierchia, 2012), and we look forward to future work that directly investigates this link.

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 under-informative demonstrations. In the absence of this prediction, one could have easily 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. 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. Even though 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 Grice's cooperative principle (Grice, 1975). The current study provides an important empirical link between some of the most distinctively human behaviors: teaching and communication.

## Acknowledgements

We thank Hannah Pelton & Cristina Spicer for help with filming stimulil, Shivani Kaushal & Ilona Bass for help with data collection, and Laura Schulz, Mike Frank, Tamar Kushnir, & Desmond Ong 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 John Templeton Foundation.

## References

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

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

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language learning and development*, *8*(4), 365–394.

Grice, H. (1975). *Logic and conversation'in p. cole and j. morgan (eds.) syntax and semantics volume 3: Speech acts.* Academic Press, New York.

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

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*(6), 1723–1739.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165-188.

Papafragou, A. (2006). From scalar semantics to implicature: children's interpretation of aspectuals. *J Child Lang*, *33*(4), 721-57.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, *86*(3), 253–282.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning From Others: The Consequences of Psychological Reasoning for Human Learning. *Perspectives on Psychological Science*, *7*(4), 341–351.

Skordos, D., & Papafragou, A. (2014). Scalar inferences in 5-year-olds: the role of alternatives. *Proceedings from the 38th Annual Boston University Conference on Language Development.*, Somerville, MA: Cascadilla Press.