CHILD DEVELOPMENT

Child Development, January/February 2019, Volume 90, Number 1, Pages e148–e164

From Exploration to Instruction: Children Learn From Exploration and Tailor Their Demonstrations to Observers' Goals and Competence

Hyowon Gweon
Stanford University

Laura Schulz
Massachusetts Institute of Technology

This study investigated whether children learn from exploration and act as effective informants by providing informative demonstrations tailored to observers' goals and competence. Children (4.0–6.9 years, $N = 98$) explored a causally ambiguous toy to discover its causal structure and then demonstrated the toy to a naive observer. Children provided more costly and informative evidence when the observer wanted to learn about the toy than observe its effects (Experiment 1) and when the observer was ordinary than exceptionally intelligent (Experiment 2). Relative to the evidence they generated during exploration, children produced fewer, less costly actions when the observer wanted or needed less evidence. Children understand the difference between acting-to-learn and acting-to-inform; after learning from exploration, they consider others' goals and competence to provide "uninstructed instruction".

Considerable research suggests that both the ability to learn from spontaneous exploration (see Schulz, 2012, for a review) and the ability to inform others (see Kline, 2015; Strauss, Calero, & Sigman, 2014, for reviews) emerge early in development. However, acting as an effective informant may require learners to generate different evidence than they would in exploring the world for themselves. In fact, the ability to learn from self-guided exploration and socially transmit the acquired knowledge for others is critical for successful accumulation of cultural knowledge. The current work investigates whether preschool-aged children who learn from spontaneous exploration can take the information they gain to provide effective demonstrations for others. In particular, we look at whether children can take into account both individual differences in learners' goals and individual differences in their competence.

Children spontaneously explore the world starting in infancy (e.g., Adolph, Eppler, & Gibson, 1993; Kretch & Adolph, 2016; Needham, 2000); as early as 11 months, infants selectively explore when events violate their expectations about object properties (Stahl & Feigenson, 2015). By preschool, children also selectively explore when evidence violates their abstract beliefs about object kinds or causal relations (Legare, 2011; Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Schulz, Standing, & Bonawitz, 2008) and when evidence is confounded or fails to distinguish competing hypotheses (Cook, Goodman, & Schulz, 2011; Gweon & Schulz, 2008; van Schijndel, Visser, van Bers, & Raijmakers, 2015). Through free play, children can learn novel functions (Bonawitz et al., 2011; Shneidman, Gweon, Schulz, & Woodward, 2016), causal relationships (Gweon & Schulz, 2008; Sobel & Sommerville, 2010), and even abstract higher-order rules (Sim & Xu, 2017) from self-generated evidence.

Young children also spontaneously seek to inform and demonstrate things to others, taking into account observers' knowledge and goals. Even preverbal infants are more likely to point to inform adults more when the adults are ignorant than when they are knowledgeable (Liszkowski, Carpenter, & Tomasello, 2008), and toddlers readily override an adult's requests for help to provide better, alternative means to achieve the adult's goals (Martin & Olson, 2013). Yet, the relationship between children's ability to learn from self-generated evidence and the ability to provide useful evidence for others has been left relatively unexplored.

During preschool years, children begin to produce causal explanations both in physical and psychological domains and even "teach" their peers by

Thanks to Veronica Chu and Mika Asaba for help with data collection and coding. This work was funded by Varieties of Understanding grant from the Fordham University and the John Templeton Foundation to Hyowon Gweon and NSF STC awards (CCF-1231216) to Laura Schulz.

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.

 0009-3920/2019/9001-00036
DOI: 10.1111/cdev.13059

demonstrating causal interventions; the quality of their explanations and interventions increases with age (Ashley & Tomasello, 1998; Bass et al., 2017; Bensalah, Olivier, & Stefaniak, 2012; Rhodes, Bonawitz, Shafto, Chen, & Caglar, 2015; Strauss, Ziv, & Stein, 2002; Walker, Lombrozo, Legare, & Gopnik, 2014; Wellman & Lagattuta, 2004; Wood, Wood, Ainsworth, & O'Malley, 1995; see Strauss et al., 2014, for a review). By around 5 years of age, children generate different evidence to teach than to deceive (Rhodes et al., 2015), provide instructions that address the particular mistakes of the learner (Ronfard & Corriveau, 2016), prioritize transmitting information that is conventional and causally opaque (e.g., Clegg & Legare, 2016; Ronfard, Was, & Harris, 2016), and teach others what they were taught especially when they themselves had difficulty solving the problem (Ronfard et al., 2016).

Toward the end of preschool years, children develop an explicit understanding of teaching as a process that causes changes in others' knowledge (Sobel & Letourneau, 2016; Ziv & Frye, 2004), suggesting that children's concept of teaching is already quite adult-like. Recent work suggests that children readily adjust the *amount* of information communicated depending on what the learner already knows (Gweon, Shafto, & Schulz, 2014, 2018); for instance, 5- and 6-year-old children provide efficient demonstrations (e.g., pushing only the three working buttons on a 20-button toy) when the learner and the child both know that functional buttons are rare, but provide more costly demonstrations (e.g., pushing all 20 buttons) when failing to do so might mislead a naive learner. Children this age also appropriately sample exemplars that support accurate learning (Rhodes, Gelman, & Brickman, 2010), and even selectively teach things that maximize the learner's benefits and minimize the learner's costs of exploration (Bridgers, Jara-Ettinger, & Gweon, 2016). Such sensitivity to others' knowledge and the ability to select appropriate evidence based on the costs and benefits of teaching are consistent with prior work on theory of mind (Wellman, Cross, & Watson, 2001) and a recent computational proposal for an early-emerging foundation for utility-based social reasoning (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). Collectively, prior work suggests that children's understanding of teaching develops rapidly during preschool years, raising important questions about how preschool-aged children learn to teach others with limited exposure to formal schooling.

Critically however, in most previous work on children's teaching, participants have themselves been explicitly taught before being asked to teach another learner: An adult instructor demonstrated the correct solution to a problem, showed how the toys worked, or communicated the rules of a game (e.g., Ronfard & Corriveau, 2016; Strauss et al., 2002; Wood et al., 1995). Thus, it is possible that children in these studies merely mimicked or reproduced the adults' instructions when they were asked to teach a learner, rather than generating informative evidence based on their knowledge. In the few studies where children were not explicitly taught, the experimental designs and stimuli themselves constrained children's teaching behaviors to simple directives (e.g., how to operate an apparatus for which there was a single solution; Ashley & Tomasello, 1998), choice between two sets of exemplars to communicate a concept (Rhodes et al., 2010), or repetitions of simple actions such as pressing buttons on a toy (Gweon et al., 2014, 2018). Therefore, although these studies suggest that children can act as helpful teachers in simple contexts where there is little ambiguity about what data to provide, they leave open the question of how effectively children teach when they must select and structure the data themselves. In the current study, we sought to examine children's ability to provide "uninstructed instruction" – generating useful evidence as teachers without the prior experience of being taught – with little constraints on what to demonstrate or how to structure the evidence.

Studying children's ability to generate evidence as teachers is important for understanding the origins of cumulative culture. Successful accumulation of knowledge not only occurs through successive social transmission of information or iterated learning (Kaush, Griffiths, & Lewandowsky, 2007; Kirby, Cornish, & Smith, 2008) but also through individuals discovering new knowledge from self-guided exploration and sharing that knowledge with others. Thus, the effectiveness of social learning may depend on learners' ability to transition from learning from noisy self-generated data to selectively reproducing data for others; such transition are most effective when they fit the learners' goals and needs. This ability is critical for the "ratchet effect" (Tomasello, Kruger, & Ratner, 1993), as it allows new discoveries and modifications to be incorporated into the repertoire of socially transmitted knowledge. More generally, teaching in humans means more than simply directing others what to do, re-enacting what others showed us, or reproducing our own past causal interventions (Kline, 2015). Here, we suggest that the power and flexibility of human teaching comes from the ability to generate informative evidence from one's own understanding of the world, in ways that are tailored for the observers' goals and abilities, regardless of how such knowledge was acquired in the first place.

The problem of choosing what to teach would be simple if the learners could identify and ask for exactly the information they need. Unfortunately, learners often do not know what they do not know. Even when the learner can specify a learning goal (e.g., "What is that?", "I want to know how it works"), the requests do not specify the set of evidence that would satisfy that goal. Although some instances of teaching may involve simply reproducing a known skill or procedure for a novice learner (e.g., showing how to operate a simple causal affordance) or providing factual knowledge (e.g., the name of an object), generating the right kind and the amount of data to satisfy an abstract learning goal can be arbitrarily complex; there may be many possible sets of demonstrations or instructions that the teacher could provide. Furthermore, because learners might have limited information processing capacities or lack relevant expertise, informants need to know how to adjust their communication (e.g., by providing multiple repeated demonstrations, speaking slowly, and showing the most informative data first) to ensure accurate learning. These challenges suggest that effective teaching requires an understanding of the link between *evidence* and *inference*; by understanding how observed evidence might influence the learner's inferences (and change or update the learner's beliefs), the teacher can flexibly select the evidence that is useful, relevant, and necessary for the learner, while avoiding evidence that is unnecessary and costly.

There are reasons to believe that this understanding may be within the capabilities of preschool-aged children. First, children between ages 3–5 show remarkable improvements on ToM tasks (e.g., false belief tasks; Wellman et al., 2001), which require a basic understanding of others' observations and beliefs. Notably, a recent study suggests a relationship between children's ToM performance and pedagogical selection of evidence (Bass et al., 2017). Second, children are sensitive to the quality of information provided by others, preferentially learning from accurate informants (e.g., Sabbagh & Baldwin, 2001; Koenig & Harris, 2005; Birch, Vauthier, & Bloom, 2008) and informants who provide more informative evidence (Gweon, Pelton, Konopka, & Schulz, 2014; Gweon & Asaba, 2017; see Heyman & Legare, 2013; Sobel & Kushnir, 2013 for reviews). Third, given a sequence of causal demonstrations, children selectively imitate causal actions that are likely to be efficacious based on statistical information and pedagogical cues (Buchsbaum, Gopnik, Griffiths, & Shafto, 2011) or the efficacy of their own prior actions (Williamson, Meltzoff, & Markman, 2008), suggesting that children can select evidence that is relevant to achieving their own goals. Furthermore, older preschool-aged children readily integrate knowledge about learners' prior beliefs, observed data, and the sampling process to decide when learners will retain or revise their prior beliefs (Magid, Yan, Siegel, Tenenbaum, & Schulz, 2017), and 6-year-olds preferentially provide a diverse sample (e.g., a Dalmatian, a Collie, and a Basset Hound) than a nondiverse sample (e.g., three Dalmatians) to teach a concept "dog" to a learner (Rhodes et al., 2010). Collectively, these studies suggest that by late preschool years, children have a sophisticated understanding of the relationship between evidence and inference. A critical question, then, is whether such understanding is sufficient for children to adapt the kinds of information they acquire from self-guided exploration to the kinds of information appropriate to other learners' goals and abilities.

To investigate this, we first let children learn about a causal mechanism from self-guided exploration (Exploration phase) and then asked them to demonstrate the mechanism to a naive learner (Demonstration phase). We compare the evidence children generate in the Exploration phase and the and Demonstration phase, and ask whether they provide different evidence in the Demonstration phase depending on the goals and abilities of the observers. We use a paradigm in which two potential causal variables (blocks and mats) can each take on one of two values (blocks: blue or yellow, mats: black or white) to generate one of two potential effects: A blue block on either a black or white mat turns on a red light, and a yellow block on either a black or white mat turns on a green light (Figure 1). The value of one variable (the color of the block) determines which effect will occur. However, the value of the other variable (the color of the mat) is irrelevant; either state will generate the outcome. We also manipulate the relative costs of changing the values of these variables; picking up either block is easy, but the mats are on opposite side of the room. Thus, changing the value of the block is relatively low cost and changing the value of the mat is relatively high cost. This setup allowed us to ask whether children selectively perform costly actions when such actions are specifically helpful to the observer. Furthermore, because children were allowed to freely demonstrate any aspect of the causal mechanism (referred to as a "toy" in the experiment), we were able to measure the effectiveness of children's teaching without imposing constraints on their behaviors or asking them to choose between prearranged samples of evidence.

In Experiment 1, we ask whether children selectively provide different demonstrations depending on the observer's goals (i.e., seeing a toy's effects or learning how a toy works), focusing on children's distinction between mere "showing" versus "teaching." In Experiment 2, we ask whether children selectively provide different demonstrations depending on the learner's abilities (i.e., more or less competent) when they are teaching the learner. Prior work suggests that children understand how people can vary in their physical abilities (e.g., being able to jump) or intelligence (e.g., being smart) and use this to predict their behaviors or choose whom to trust (e.g., Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015; Lane, Wellman, & Gelman, 2013). Yet, whether children incorporate this understanding *as teachers* to tailor their selection of evidence is a question that has not been investigated. If children possess a genuine understanding of the relationship between evidence and inference, it is possible that children can also modulate their teaching behaviors depending on the learner's competence.

Given previous studies on children's understanding of teaching (e.g., Ziv & Frye, 2004; Ronfard & Corriveau, 2016 Sobel & Letourneau, 2016) and their ability to consider the costs of actions in teaching and social reasoning more generally (Bridgers et al., 2016; Gweon et al., 2014, 2018), we tested children between ages 4 and 6; by recruiting children between preschool and early school years, we sought to identify early competence as well as potential developmental change.

Experiment 1

In Experiment 1, we present a novel causal apparatus as described earlier and let children explore the apparatus to learn that one causal variable (the color of the blocks) is relevant to the particular outcome and the other (the color of the mats) is irrelevant to which effect (red vs. green lights) occurs. We then ask children to introduce the toy to a naive observer. Across two conditions, we manipulated whether the naive agent wanted to see the effect generated by the toy (Show Lights condition) or understand how the toy works (Teach Toy condition). Varying the mats is not necessary for the observer who wants to just see different effects, but it is important for the one who wants to know how the toy works. Critically, one of the mats was placed near the child (Near Mat), while the other mat was placed on the other side of the room (Far Mat), making it costly to provide evidence that the particular mat is irrelevant to the causal effects. We looked at whether children in the Teach Toy condition would be more likely than children in the Show Lights condition to (a) generate more evidence overall, (b) generate the costly informative evidence that required changing the mats, and (c) generate evidence comparable to the Exploration phase (when they also had to disambiguate the causal structure of the toy). Specifically in the Show Lights condition, children should preferentially use the blocks on the Near Mat since there is no need for children to move to the other side of the room to use the Far Mat. By contrast, in the Teach Toy condition, children should demonstrate both that the different blocks activated different lights and that the different mats behaved identically; this would involve showing more actions, more frequent changes between the mats, and more frequent uses of Far Mat, and result in overall more causally informative demonstration than in the Show Lights condition.

Method

Subjects

We recruited 48 children (N = 24/condition), sufficient for a power of .7 assuming reasonably large effect size (Cohen's d = .7). Children between ages 4.0–6.9 (M_{age} = 5.7 years) were recruited from a local children's museum (October 2012 to June 2013) and were randomly assigned to one of two conditions, N = 24/condition; mean age in months: 68.5 (Show Lights) versus 69.6 (Teach Toy); no. of boys: 14 (Show Lights) versus 11 (Teach Toy). Eight children were excluded from analyses due to parental interference (N = 2), not completing the procedure (N = 2), or experimental error (N = 4). The demographics of participants were representative of a typical urban middle-class neighborhood.

Materials

An Elmo hand puppet was used as the naive observer. The toy was a modified version of the causal toy used in Gweon and Schulz (2008), which consisted of three components: two mats (black and white), two blocks (Blue and Yellow), and two boxes (red and green). The boxes lit up when the blocks were placed on the mat (see Figure 1 for the experimental setup). Each mat was made of 30 × 30 (cm) foam boards, covered with wire mesh with distinct patterns. One was colored in black and the other in white. One mat was placed right next to where the child sat in the beginning of the experiment (henceforth Near Mat). It was set vertically against a wall in the testing room, approximately 0.4 m from the floor. The other mat (Far Mat) was placed vertically near the other side of the wall, on a table approximately 2 m from the other mat, and 0.8 m from the floor, so that the child had to walk over to the other side and go around the table (or climb on the table) to use the mat. Each block was made of acrylic boards, approximately 5 × 5 × 2.5 (cm) with a small knob on top. One was colored in blue and the other in yellow. Each light box was approximately 20 × 20 × 10 (cm), placed side by side in front of the participant. One was covered with red felt and one with green felt. The boxes contained light bulbs visible through a transparent window in front of the box. Importantly, both mats were magnetic, and each block contained a magnetic sensor wirelessly connected to each box. Thus, when the blue block contacted either the white or the black mat, the red light box lit up and stayed lit as long as the block remained in contact with the mat; similarly, the yellow block on either mat activated the green light box. Thus, although the mats were necessary for the activation of the lights, the distinction between the two mats was only perceptual and not functionally involved in activating different light boxes.

Procedure

The experiment took place in a quiet room in the museum. Once the child sat down in front of the red and green light boxes (Figure 1), the experimenter asked the child to point to each of the light boxes, the yellow and the blue blocks, and the white and the black mats. These initial check questions ensured that the participant was able to answer simple questions in English and identify different colors and the components of the toy. Then, the experimenter pointed to the blue block and the Near Mat, and said "Hmm, why don't you try putting this blue block on this mat, and see what happens?" Once the participant saw that one of the light boxes lit up, the experimenter said, "I have to go write something down, so why don't you go ahead and play with the toy?" and walked out from the child's line of sight (Exploration phase). Children were allowed to freely explore the toy for

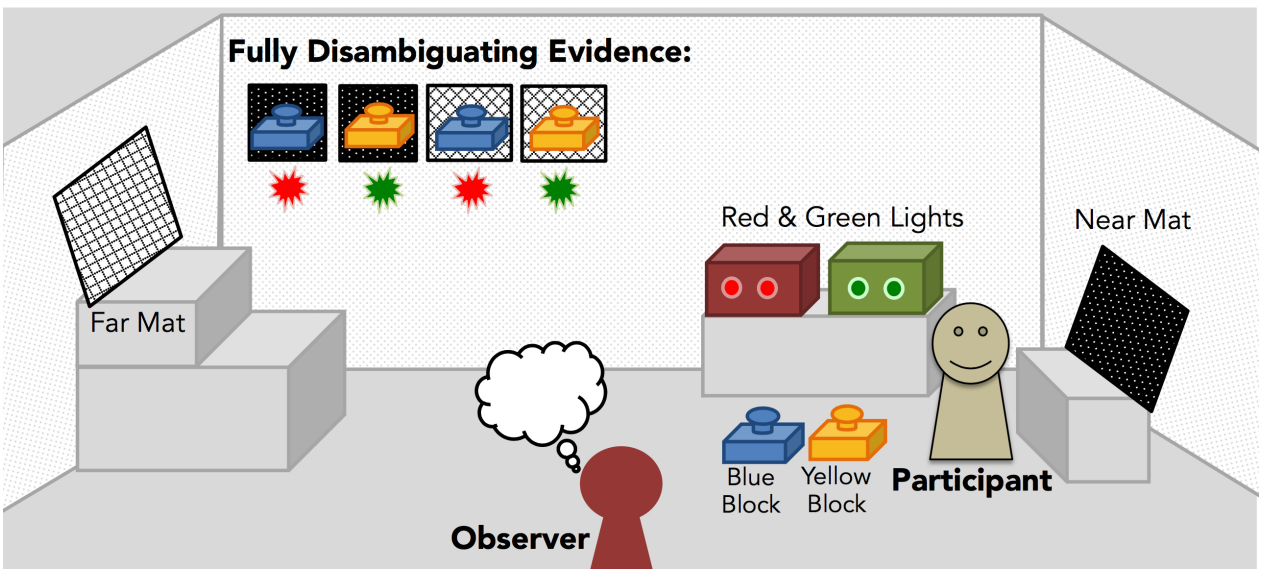

Figure 1. Schematic of the experimental setup in Experiments 1 and 2. One (black) mat is easy to reach and placed closer to the child (Near Mat), and the other (white) mat is costly to reach (Far Mat). Two potential causal variables (blocks and mats) could each take on one of two values (blocks: blue or yellow, mats: black or white) to generate one of two potential effects (lights: red or green). In fact, placing the Blue block on a mat activates the red light, and placing the Yellow block on a mat activates the green light, regardless of the mat color; in order to fully disambiguate the causal structure, at least three of the four possible actions are required.

1 min, or until the child stated that they were done playing, whichever came first. The experimenter then returned to the child and covered up the light boxes so that the child was unable to see which box lit up. She asked the child to turn the red light on and then to turn the green light on. Then, she asked two more test questions. First, she took whichever block the child had just used to turn the green light on, brought it near the opposite mat, and asked, “If I put this block on this mat, will it turn on the same green light, or the different red light?” Second, she took the other block to bring it over the same mat, and asked, “What if I put this block here? Will it turn on the same red (green) light, or the different green (red) light?” These questions were used to assess whether the child had learned the relevant and irrelevant variables during play. Importantly, the experimenter did not give feedback or teach children about the causal structure of the toy. If the child could not answer the question or explicitly said, “I don’t know,” the child was given another minute to play (7 of 48 children played for another minute). These children were asked the same questions after their second playtime.

Finally, in the Demonstration phase, the experimenter brought out the Elmo puppet and said, “Now my friend Elmo is going to come out! Elmo is a silly monster and he has never seen this toy before.” (Describing Elmo as “silly” emphasized his naive status relative to the children, making it less likely that they would treat him as an equally knowledgeable peer or an authority.) In the Show Lights condition, she said, “Elmo really likes to see red and green lights! Can you show Elmo red and green lights?” In the Teach Toy condition, the experimenter said, “Elmo really wants to learn how it works. Can you teach Elmo how the toy works?” Children demonstrated the toy to Elmo for as long as they wanted; when the child said, “I’m done,” or when they indicated that they were done by putting down the blocks or stopped to look at the experimenter, the experimenter asked “Are you done?” If the child said “No,” they were allowed to continue. If children were still demonstrating the toy after 90 s, the experimenter said, “Let me know when you’re done!” and repeated the prompt every 30 s.

Video Coding

Video recordings of the testing sessions were coded using a video annotation software (VCode; Hagedorn, Hailpern, & Karahalios, 2008) by a trained coder blind to condition manipulation. Its outputs were then analyzed using a custom script in separately for the initial play with the toy (Exploration phase) and during the child’s demonstration of the toy to Elmo (Demonstration phase). For both exploration and demonstration, we coded for each time the child placed a block (Blue, Yellow) on a mat (Near, Far) to turn on a light box; each of these instances was coded as an “Action” (e.g., if both blocks were placed on the same mat, they were coded as two actions). Note that the actions on the Far Mat were more costly to perform (see Stimuli) than actions on the Near Mat; we thus calculated the number of actions on the Far Mat separately from those on the Near Mat, and used Far Mat Actions as a measure of children’s generation of costly evidence. Because children could repeat several actions on the Far Mat once they have moved over to that side, we also coded each time the child moved from one end of the room to the other end to use a different mat (coded as a “Transition”) as a measure of their overall effort during the demonstration.

We measured the informativeness of the evidence children generated by looking at the first four actions they performed, and their actions overall. The child could perform one of four different causal actions: placing the yellow block on each mat, and the blue block on each mat. In principle, any three of the four actions would allow the observer to disambiguate the causal structure and infer that the blocks, rather than the mats, control which light is activated. For a naive observer, however, it is maximally helpful to see all four unique actions as soon as possible; this set of evidence eliminates the observer’s uncertainty about the causal structure, allowing the observer to quickly learn the correct causal structure. We thus analyzed how many of these unique actions the child produced during her *first four actions* in the Exploration phase and the Demonstration phase. The four unique actions did not necessarily have to be fully isolated actions. For instance, if a child placed both blocks on the mat but then lifted one of the blocks off the mat (such that only one block was placed on a mat to activate one light), this counted as an instance of one of four unique actions. We used these data to calculate an Informativeness Score of the child’s first four actions in each phase, ranging from 0 to 3 (0: fully confounded evidence, 1: trying one block on each mat or trying each block on one mat; 2: three of the four unique actions; 3: all four unique actions). This Informativeness Score served as a graded measure of children’s ability to effectively structure their demonstration by front-loading the most informative set of evidence early in the Demonstration phase. Finally, we also looked at whether the child produced at least three of the four actions at any point during the exploration and Demonstration phase.

Results

Children were given identical instructions and questions during the Exploration phase, and had limited time (up to 1 or 2 min for those who initially failed the check questions). Thus, we did not predict any differences in how children initially explored the toy; we expected that differences would emerge only in how they demonstrated the toy to the observer. We provide the full mixed-effects analysis of variance (ANOVA) results (condition as a between-subjects variable, phase as a within-subjects variable) in Supporting Information; here we focus on planned comparisons between conditions separately for the Exploration and Demonstration phases.

Given our use of children's action-based demonstrations as the primary measure of teaching, prior to our main analyses we first checked whether children relied on language as an alternative means to teach. Most children focused on demonstrating the toy rather than providing verbal instruction; 12.5% of children (3 of 24) in the Show Lights condition and 29.1% of children (7 of 24) in the Teach Toy condition produced a verbal utterance during the Demonstration phase (p = .29, Fisher's exact), and these verbalizations were always accompanied by demonstrations. Furthermore, their verbal utterances rarely contained causal information; only three of these children (all in the Teach Toy condition) explained the causal structure by explicitly mentioning the blocks and the mats. This allowed us to consider children's actions as an appropriate measure of what, and how much, evidence children generated for the observer. Unless otherwise noted, we used two-tailed between-subjects Welch's t-test for planned comparisons between conditions, and two-tailed paired-samples t-test for comparisons between phases within each condition (Exploration vs. Demonstration).

Evidence Generation During the Exploration Phase

As expected, during the Exploration phase, children in the Show Lights condition and Teach Toy condition did not differ in their total playtime, (Show Lights vs. Teach Toy: 76 vs. 68 s, $t(45.73)$ = 1.05, p = .30; see Figure 2 for results) Actions (10.29 vs. 12.29, $t(33.63)$ = 0.87, p = .38; Far Mat Actions, 2.88 vs. 3.29, $t(44.58)$ = 0.47, p = .64), Near Mat Actions (7.41 vs. 9.0, $t(36.54)$ = 0.73, p = .47), Transitions between the two mats (Show Lights vs. Teach Toy: 1.96 vs. 2.17, $t(45.96)$ = 0.34, p = .73) or the Informativeness Score (0.79 vs. 1.08, Z = 0.88, p = .38; Mann–Whitney U test). All children used both blocks at least once to activate the lights; 75% of children (Show Lights) and 79.1% of children (Teach Toy) used both mats at least once during the Exploration phase (p = 1.0, Fisher's exact). Children also learned the toy's causal structure from their exploration: 79.1% of the children (Show Lights) and 70.8% (Teach Toy) of children correctly answered that changing the block would activate a different light (p = .74), and 62.5% (Show Lights) and 66.7% (Teach Toy) of children correctly answered that the same block on a different mat would activate the same light (p = 1). Thus, during the Exploration phase, children did not show any difference across conditions; children explored the toy and learned from self-generated evidence regardless of condition.

Evidence Generation During the Demonstration Phase

Our primary question of interest was whether children's Actions, Far Mat Actions, Transitions, and Informativeness Score would differ between conditions during the Demonstration phase. The duration of the Demonstration Phase did not differ across conditions (Show Lights vs. Teach Toy: 40 vs. 57 sec., t(34.71) = 1.56, p = 0.13). However, during the Demonstration phase, children in the Teach Toy condition produced more Actions (Show Lights vs. Teach Toy: 5.17 vs. 11.58, $t(31.21)$ = 2.27, p = .030) specifically on the Far Mat (1.42 vs. 4.88, $t(27.62)$ = 2.42, p = .022) and produced more Transitions between the mats (1.29 vs. 3.25, $t(28.27)$ = 2.04, p = .050). Note, that we did not have an a priori hypothesis about the Near Mat; children in the teach condition might perform overall more actions on both mats, but it was also possible that the effect would be especially larger on the Far Mat (the presence of a condition effect on the Near Mat is independent of its presence on the Far Mat). An exploratory analysis showed that children's actions on the Near Mat did not differ significantly across conditions (3.75 vs. 6.37, $t(33.05)$ = 1.40, p = .17).

Comparison of the Informative Score also revealed that the first four actions of children in the Teach Toy condition were more informative than those of children in the Show Lights condition (Show Lights vs. Teach Toy: 0.67 vs. 1.54, Z = 2.58,

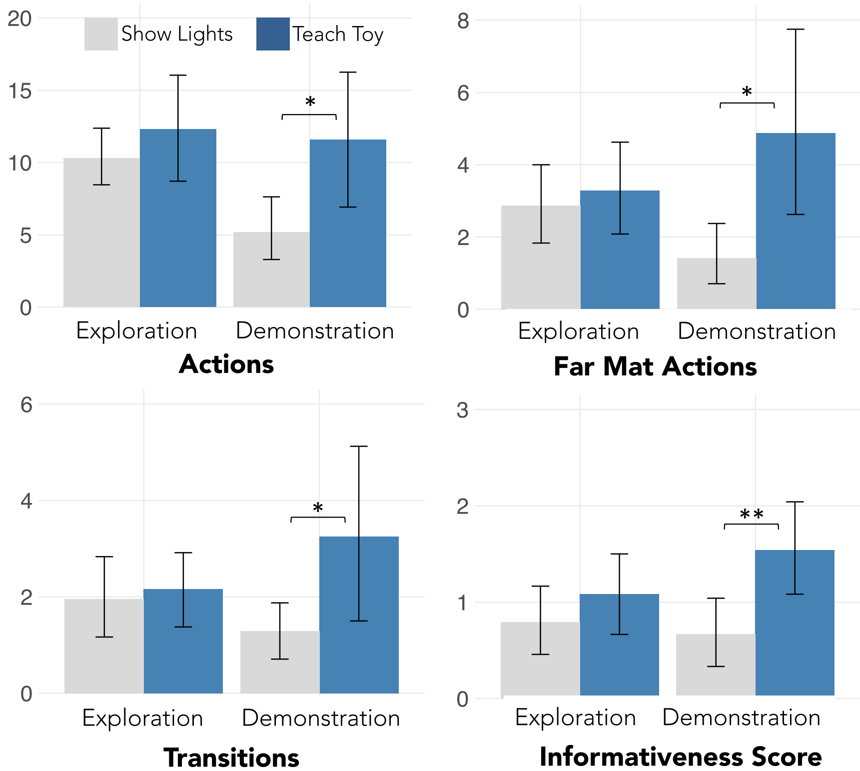

Figure 2. Results from Experiment 1. Error bars indicate 95% CI. *$p < .05$. **$p < .01$.

$p = .01$, Mann–Whitney *U* test). Furthermore, more children in the Teach Toy condition than the Show Lights condition eventually produced the set of demonstrations (at least 3 of the 4 unique actions) that fully disambiguated the toy (proportion of children who produced at least 3 of the 4 unique actions; Show Lights vs. Teach Toy: 16.6% vs. 54.2%, $p = .01$, Fisher's exact).

We then asked whether there were any age-related trends in children's demonstrations in each condition. In the Show Lights condition, none of the measures were correlated with age (Actions: $r(22) = .21$, $p = .32$; Far Mat Actions: $r(22) = .22$, $p = .30$; Transitions: $r(22) = .14$, $p = .50$; Informativeness Score: $r(22) = .29$, $p = .17$; Duration of Demonstration phase: $r(22) = -0.02$, $p = .93$). In the Teach Toy condition, only the Informativeness Score was positively correlated with age ($r(22) = .54$, $p = .006$, Actions: $r(22) = .26$, $p = .22$; Far Mat Actions: $r(22) = -.03$, $p = .9$; Transition: $r(22) = .23$, $p = .29$; Duration of Demonstration phase: $r(22) = .27$, $p = .20$). Further exploratory analyses showed that the correlation coefficients for Age × Informativeness Score did not differ across conditions (Fisher's $Z = 1.01$, $p = .31$), and the Informativeness Scores in the Exploration phase did not correlate with age).

Unlike the Exploration phase, children's behaviors in the Demonstration phase clearly differed across conditions. Children produced both more evidence overall and more costly evidence when the observer wanted to understand how the toy worked rather than merely see the toy activate. Similarly, children's demonstrations were more informative when the learner wanted to understand the causal structure of the toy than when he simply wanted to observe the effects, and children's tendency toward informativeness increased with age.

Comparisons Between Exploration and Demonstration Phases

Our primary hypotheses concerned differences between conditions during the Demonstration phase. However, we also explored whether children behaved differently during the Exploration and Demonstration phases, separately for each condition. Here, we report the paired-samples *t*-test results within each condition (see Supporting Information for full statistics from the mixed-effects

ANOVA). Children in the Show Lights condition spent less time acting on the stimuli during the Demonstration phase than the Exploration phases, (Exploration vs. Demonstration: 76 vs. 40 s, $t(23) = 4.94$, $p < .001$). They also produced fewer Actions (10.29 vs. 5.17, $t(23) = 3.87$, $p = .001$), which was reflected in both Far Mat Actions (2.88 vs. 1.42, $t(23) = 3.30$, $p = .003$), and Near Mat Actions (7.42 vs. 3.75, $t(23) = 2.79$, $p = .01$); the only behavioral measures that did not decrease between the Exploration and Demonstration phase were Transitions (1.96 vs. 1.29, $t(23) = 1.40$, $p = .175$). By contrast, no behavioral measure differed between the Exploration and Demonstration phase for the Teach Toy condition (Exploration vs. Demonstration: 68 vs. 57 s, $t(23) = 0.96$, $p = .35$; Actions: 12.29 vs. 11.58, $t(23) = .25$, $p = .80$; Far Mat Actions: 3.29 vs. 4.88, $t(23) = -1.36$, $p = .19$; Near Mat Actions: 9.0 vs. 6.71, $t(23) = 0.93$, $p = .37$; Transition: 2.17 vs. 3.25, $t(23) = -1.29$, $p = .21$). Thus, when the observer merely wanted to see the causal effects, children produced less evidence in demonstration than exploration; however, when the observer wanted to learn about the toy (as children themselves did during the exploration), they produced as much evidence in demonstration as in exploration.

In sum, although children produced comparable evidence when learning the causal structure of the toy themselves in both conditions, they provided different demonstrations for the observer depending on his goals. Collectively, results from Experiment 1 suggest that children do not indiscriminately reproduce the evidence they generate; they selectively reproduce evidence that is useful with respect to the learner's goals.

Experiment 2

Experiment 1 suggests that children can accommodate observers' different goals and adjust the evidence they provide accordingly. Can children also accommodate differences in learners' competence to help them achieve the same learning goals? For instance, an exceptionally bright learner might be able to infer the causal structure of the toy by observing only three of the four disambiguating demonstrations; however, a more ordinary learner might require all four unique demonstrations and benefit from seeing repeated demonstrations. Preschool-aged children are sensitive to others' epistemic competence (e.g., "smart," "not smart") as well as their access to relevant information, and use this information to decide from whom to learn (Lane et al., 2013). Given the results from Experiment 1 showing that children readily consider observer's goal to tailor their demonstrations, it is possible that children can also consider the observer's competence to tailor their selection of evidence.

To test children's ability to adjust their evidence according to the learners' abilities, in Experiment 2 children in both conditions were asked to "teach" the observer; the only difference was in whether the observer was introduced as an exceptionally competent learner (Exceptional condition) or an ordinary learner (Ordinary condition). We predicted that children would provide more evidence for an ordinary learner than for an exceptionally smart learner. Additionally, as in Experiment 1, we looked at whether there would be any differences between the kinds of evidence children produced in exploration and instruction. We predicted that in the Demonstration phase, children in the Ordinary condition would be more likely than children in the Exceptional condition to (a) generate more evidence overall, (b) generate the costly informative evidence that required changing the mats, and (c) generate evidence more comparable to the evidence children generated themselves during the Exploration phase.

Method

Subjects

Fifty children, $N = 25$/condition, M_{age}(range) = 5.4 (4.0–6.9), were recruited from a local children's museum (October 2014 to November 2014) and were randomly assigned to one of two conditions, mean age in months: (Exceptional vs. Ordinary: 65.5 vs. 65.0); no. of boys: (Exceptional vs. Ordinary: 17 vs. 11). The intended sample size was based on Experiment 1 ($N = 24$), but two additional children were tested on the last day of data collection; thus, we included them in the analysis. An additional 12 children were excluded from analyses due to failing the initial checks ($N = 1$), parental interference ($N = 2$), not completing the procedure ($N = 3$), or toy malfunction and experimenter error ($N = 6$). One additional child was excluded for not using the mat at all during teaching.

Materials

All materials and setup were identical to Experiment 1, except that the Elmo puppet was replaced by a human boy puppet; this ensured that children did not have strong a priori beliefs about the learner's competence.

Procedure

The general procedure was highly similar to Experiment 1. The only difference in the Exploration phase was that before children began playing, the experimenter mentioned that they would be teaching her friend about the toy after they are done playing. This led to an overall higher accuracy in answering post-play test questions about the toy (see Results) and no children received a second Exploration phase.

The key difference in the Demonstration phase was how the learner was introduced. After the same test questions as in Experiment 1, the experimenter brought out a puppet and introduced him as her friend "Paul." In the Exceptional learner condition, Paul was described as a smart friend who knew a lot and was quick to understand things. She asked, "Hey Paul, can you tell me what is 152 times 38? (Paul said: 5776!) Whoa! Paul, do you know how the light on the ceiling works?" Paul said: "Yes! If you flip the switch on the wall, it completes the circuit. Electricity flows into the light bulb, and the light turns on." In the Ordinary learner condition, Paul was described as a silly friend (as in Experiment 1) who did not know much and needed help to understand things. She then asked what is 5 plus 3 (Paul said: "Hmm, I don't know, 2?") and how the light on the ceiling works (Paul said: "Hmm, I'm not sure, it goes on and off but I'd like to know why. . ."). Thus, in many aspects, this Ordinary learner condition was similar to the Teach Toy condition in Experiment 1 except that children received more direct evidence that he was generally comparable, or slightly less competent, in his knowledge about the world to the young children themselves. In both conditions, Paul then looked at the toy and said, "Wow, that looks really cool! What is that?"). The experimenter then asked the child to teach Paul about the toy. All other aspects of the experiment were identical to the "Teach Toy" condition of Experiment 1.

Results

As in Experiment 1, we did not predict differences in how children played with the toy during the Exploration phase; we predicted that these differences would emerge only in the Demonstration phase (see Supporting Information for mixed-effects ANOVA results). Again, before the main analyses, we first ensured that children were primarily using action-based demonstrations to teach rather than providing verbal instructions. Children's use of language during the Demonstration phase was comparable to Experiment 1 (Teach Toy condition) and similar across conditions. Thirty-two percent of children (8 of 25) in the Exceptional learner condition and 20% (5 of 25) in the Ordinary learner condition produced a verbal utterance during teaching (p = .52, Fisher's exact), and only 20% (5 of 25) and 12% (3 of 25) in each condition explained the causal structure by explicitly mentioning both the blocks and the mats. Thus, we focus on our main action-based measures as we did in Experiment 1; in the General Discussion, we provide additional analyses across both experiments to confirm that children's demonstrations were not influenced by whether or not they produced verbal instructions.

Evidence Generation During the Exploration Phase

As predicted, during the Exploration phase, children in the Exceptional learner and Ordinary learner conditions did not differ in any of the behavioral measures: Total playtime (Exceptional vs. Ordinary: 57 vs. 58 s, $t(40.11) = 0.20$, $p = .84$, see Figure 3 for results); Actions (14.32 vs. 12.40, $t(47.84) = 0.92$, $p = .36$); Far Mat Actions (4.16 vs. 3.52, $t(44.67) = 0.76$, $p = .45$); Near Mat Actions (10.16 vs. 8.88, $t(46.10) = 0.65$, $p = .52$); Transition (4.04 vs. 3.36, $t(46.05) = 0.59$, $p = .56$); or Informativeness Score (1.0 vs. 1.04, $Z = 0.25$, $p = .81$; Mann–Whitney U test).

All children used both blocks at least once to activate the lights; 96% of children (Exceptional learner) and 88% of children (Ordinary learner) used both mats at least once during the Exploration phase. Children also learned the toy's causal structure from their exploration and performed comparably on the test questions about the toy in both conditions: 96% of the children in the Exceptional condition and 100% in the Ordinary condition said that changing the block would activate a different light (p = 1.0, Fisher's exact), and 84% of children in both conditions understood that the same block on a different mat would activate the same light (p = 1.0). An exploratory analysis showed that children in Experiment 2 were generally more accurate than those in Experiment 1, 41 of 50 children in Experiment 2 answered both questions correctly (20 in Exceptional, 21 in Ordinary) compared with 27 of 48 in Experiment 1 (14 in Show Lights, 13 in Teach Toy; p = .008, Fisher's exact). This was presumably due to the fact that children were told (before the Exploration phase) that they would later have to teach a learner, and thus had a clearer reason to learn the causal structure of the toy during

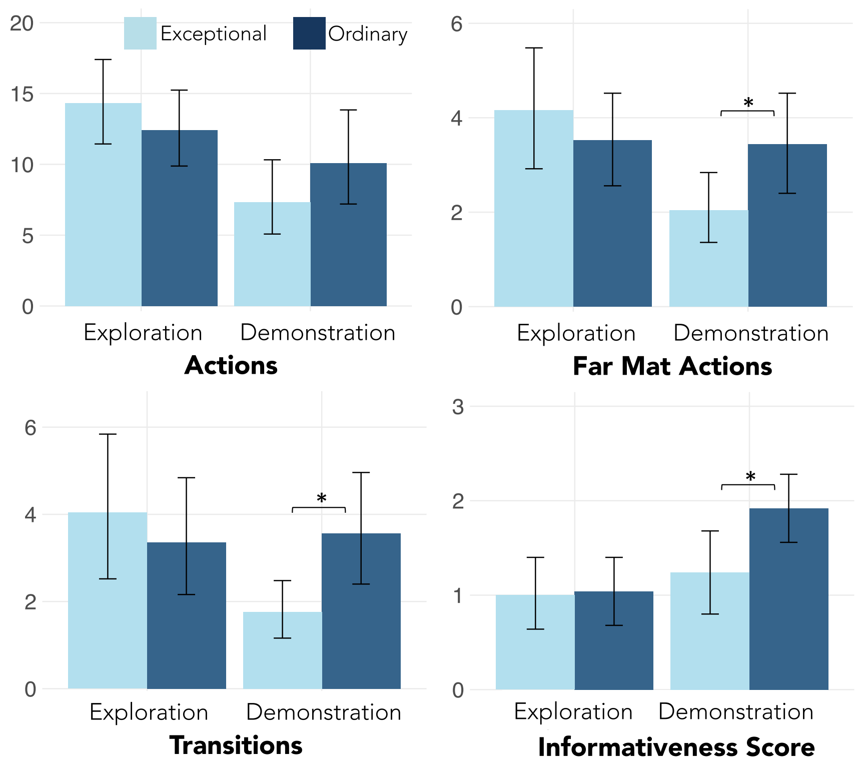

Figure 3. Results from Experiment 2. Error bars indicate 95% CI. $*p < .05$.

the Exploration phase. Overall, these analyses established that children's behaviors during exploration and their understanding of the causal structure of the toy were comparable across conditions.

Evidence Generation During the Demonstration Phase

As in Experiment 1, our primary question of interest was whether children's actions, Far Mat Actions, Transitions, and Informativeness Score would differ between conditions during the Demonstration phase. As in Experiment 1, the duration of the Demonstration phase did not differ across conditions (Exceptional vs. Ordinary: 50 vs. 57 s, $t(47.96) = 0.64$, $p = .52$). Additionally, and counter to our predictions, the difference between the two conditions in the number of Actions did not reach significance (Exceptional vs. Ordinary: 7.32 vs. 10.08, $t(46.65) = 1.31$, $p = .20$). However, consistent with our predictions, children in the Ordinary learner condition produced more actions on the Far Mat (Exceptional vs. Ordinary: 2.04 vs. 3.44, $t(42.81) = 2.03$, $p = .048$). As in Experiment 1, we did not have a directed hypothesis about the action on the Near Mat; exploratory analysis showed that it did not differ across conditions (5.28 vs. 6.64, $t(46.65) = 0.79$, $p = .44$). Furthermore, children in the ordinary condition produced more Transitions (Exceptional vs. Ordinary: 1.76 vs. 3.56, $t(33.82) = 2.24$, $p = .03$).

Relative to children in the Exceptional learner condition, children in the Ordinary learner condition were also quicker to provide the informative set of disambiguating evidence early in the Demonstration phase; the Informativeness Score of the first four actions was higher in the Ordinary condition than in the Exceptional condition (1.24 vs. 1.92, $Z = 2.22$, $p = .027$, Mann–Whitney U test). Additionally, 40% of children (10 of 25) in the Exceptional condition provided causally disambiguating evidence (at least 3 of 4 unique actions), while 68% of children (17 of 25) in the Ordinary condition did ($p = .087$, Fisher's exact). In the Ordinary condition, the Informativeness Score correlated positively with children's age ($r(23) = .44$, $p = .026$); this was not true in the Exceptional condition ($r(23) = -.203$, $p = .331$), and the difference between the two correlation coefficients was significant (Fisher's $Z = 2.25$, $p = .024$).

Thus, although children produced comparable amounts of evidence across conditions, they produced more costly and more causally informative evidence for the ordinary learner than for the exceptionally competent learner. As in Experiment 1, the tendency to front-load the most informative set of evidence increased with age in the Ordinary learner condition; however, this correlation was not observed in the Exceptional learner condition.

Comparisons Between Exploration and Demonstration Phases

As in Experiment 1, we explored whether children behaved differently during the exploration and Demonstration phases (see Supporting Information for a full report on the main effects and interactions). Compared with the Exploration phase, children in the Exceptional condition produced fewer Actions during the Demonstration phase (Exploration vs. Demonstration: 14.32 vs. 7.32, $t(24) = 4.092$, $p < .001$); fewer Far Mat Actions (4.16 vs. 2.04, $t(24) = 2.727$, $p = .011$); fewer Near Mat Actions (10.16 vs. 5.28, $t(24) = 3.138$, $p = .004$); fewer Transitions (4.04 vs. 1.76, $t(24) = 2.50$, $p = .020$). By contrast, none of these measures differed between phases in the Ordinary condition; children produced as many Actions (12.40 vs. 10.08, $t(24) = 1.188$, $p = .247$); as many Far Mat Actions (ordinary: 3.52 vs. 3.44, $t(24) = 0.122$, $p = .904$); as many Near Mat Actions (8.88 vs. 6.64, $t(24) = 1.322$, $p = .199$); and as many Transitions (3.36 vs. 3.56, $t(24) = -0.209$, $p = .836$) during the Demonstration as they did during Exploration. Thus, while children produced less evidence overall than they produced during exploration when the observer was an exceptionally competent learner, they produced as much evidence for the ordinary learner as they had produced for themselves in exploration, a pattern that is consistent with what we found in Experiment 1.

In summary, even when the observers in both conditions wanted to learn about the toy, children's behaviors differed depending on whether the observer was introduced as exceptional or ordinary. Consistent with our findings from Experiment 1, these results further suggest that children are sensitive to the competence of the learner and tailor their demonstrations accordingly.

General Discussion

Collectively, results across two experiments suggest that children can transition from exploration to instruction. As explorers, they learn from self-generated evidence; as teachers, they flexibly adjust the evidence they generate based on the observers' goals and competence. These results further suggest that children understand what information is useful for others and select the set of evidence that fulfills the learners' needs, while minimizing the costs involved by selectively generating costly evidence when such data are critical for learning.

In Experiment 1, children in the Show Lights condition generated evidence that was easy to generate (i.e., on the near mat) and failed to fully disambiguate the causal structure of the toy (i.e., not showing that changing the mat does not affect the outcome); however, they were nevertheless helpful in fulfilling the learner's goal: seeing the red and green lights. Children in the Teach Toy condition generated evidence that was harder to generate (i.e., moving to the far mat) but disambiguated the causal structure of the toy, thus providing helpful information to fulfill the learner's goal. In Experiment 2, children who taught the ordinary learner produced more costly actions than those who taught the exceptional learner, and their first few actions were more causally informative. Moreover, when children were teaching, the informativeness of their first four demonstrations increased with age (Teach Toy condition in Experiment 1 and Ordinary learner condition in Experiment 2); this suggests that older children were more likely to front-load the most informative evidence at the onset of their teaching, although this tendency was not as strong when the learner was perceived as exceptionally competent. Thus, although children as young as four can adjust their actions to provide helpful information, their ability to plan the temporal sequence of evidence to provide maximally informative teaching may continue to develop.

Note that our cost manipulation was an important way to maximize the hypothesized differences across conditions; without the cost manipulation and their desire to minimize the costs for generating the evidence, children might have provided fully informative evidence regardless of the learner's goals or competence levels, even though they understood that some of the evidence is unnecessary for the learner. More generally, the degree to which children modify their teaching behaviors should depend on a number of factors; our experiment design provided contextual support to strengthen children's motivation to differentiate their demonstrations depending on some of these factors (e.g., learner's goal and competence).

In contrast to Experiment 1, in Experiment 2 we did not see significant differences in the total number of actions and the proportion of children who eventually provided fully disambiguating evidence.

A plausible explanation is that while the observer's goal differed across conditions in Experiment 1 (learn about the toy vs. seeing the toy's effects), such that causally disambiguating evidence was necessary in only one of the conditions, the observer's goal in Experiment 2 was always to *learn* about the toy; thus in both conditions, disambiguating evidence and additional demonstrations were useful for the learner. Overall, children in Experiment 2 were more likely to provide causally disambiguating evidence than children in the Show Lights condition in Experiment 1 (54% vs. 16.6%, $p = .003$, Fisher's exact), and also performed more Actions (8.7 vs. 5.17; $t(60.44) = 2.29$, $p = .025$), and Transitions (2.66 vs. 1.29; $t(71.61) = 2.63$, $p = .01$). While comparisons between experiments should be interpreted with caution, these results provide additional support for the idea that children provide evidence consistent with observers' goals, distinguishing showing someone an outcome from teaching someone how something works.

In addition to examining differences between conditions during the Demonstration phase, we were also able to observe within-subject differences in children's behaviors between the exploration and Demonstration phases. Relative to the Exploration phase, children performed fewer actions both when merely demonstrating the toy (Experiment 1, Show Lights condition) and when teaching the toy to an exceptionally smart learner (Experiment 2, Exceptional learner condition). Although these results might look superficially similar, the fact that children distinguish observers' goals (i.e., showing vs. teaching) suggests that children performed fewer actions in these two contexts for different reasons. When merely demonstrating the toy, children may have provided less information because additional evidence would be irrelevant for the observer's goal (i.e., the observer does not *want* any more); when teaching an exceptionally smart learner, children might have provided less information even though additional evidence is still relevant to his goal (i.e., the learner does not *need* any more). These results are consistent with the hypothesis that children distinguish others' goals and levels of competence, and selectively generate costly evidence only when it is useful for the observer.

Our work also raises a broader question about how much evidence children think is ideal for teaching, and how they might flexibly adjust the amount of evidence based on the leaner's characteristics such as their competence. For instance, given learners with varying levels of competence, children might provide *less* for a competent learner, *more* for an incompetent learner, or both. Note that the current study used a relatively coarse manipulation of competence that does not distinguish different factors that underlie children's representation of others' competence (e.g., raw intelligence, amount of knowledge, being confident, being admired by others, etc.); further work is needed to understand the precise nature of children's inferences about learners' competence and how this modulates their selection of evidence.

One limitation of the current work is that the effect sizes were relatively small across measures. Thus, although most measures showed the predicted differences across conditions with high consistency, some did not show a significant Condition by Phase interaction in the mixed-effects ANOVA (see Supporting Information). It is possible that with even higher costs to generating some evidence (e.g., placing the Far Mat even further or higher), we would have seen stronger effects. However, while the results with small effect sizes should be interpreted with appropriate caution, the systematic differences we observed in different dependent measures across both experiments provide support for the hypotheses that children can use the evidence they generate themselves to inform others, and adjust the evidence they provide according to the observers' goals and abilities.

The current work focused on analyzing children's actions as a way to directly compare how children generate evidence for their own learning (exploration) versus for others' learning (teaching). The low frequency of verbal instruction (27%, 20 of 74, collapsed across three teaching conditions) and the lack of causal content in their language strengthen our rationale for using action-based demonstrations as the primary mode of information transfer. In fact, children who produced language during the Demonstration phase (verbal: $N = 20$) or a subset of those who provided causally informative verbal instructions (informative verbal: $N = 11$) did not differ from children who did not verbalize at all (nonverbal: $N = 54$) in any of the action measures we used in the analyses (nonverbal vs. verbal: all ps $> .4$; nonverbal vs. informative verbal: all ps $> .5$; see Supporting Information for a full report). These results further suggest that language did not replace or reduce children's demonstrations, and that our main dependent measures appropriately reflect what, and how much, evidence children provided for the learner.

However, given that language is a powerful tool for communicating abstract knowledge in everyday pedagogical contexts, one might wonder why children rarely used language in our experiment. A few reasons might have contributed to the low rate of verbal teaching. First, compared to prior work where children were verbally taught by the experimenter before teaching, children in our study never received verbal instruction about the toy; their knowledge was acquired via self-generated evidence during their exploration. Second, the toy had a complex causal structure that might have been rather difficult to explain verbally for young children. Indeed, a majority of children's comments referred to the superficial features of the toy (e.g., colors of the blocks or mats, description of the effect) to support their demonstrations, and less than 15% of children provided causal information that could have helped the learner disambiguate the causal structure of the toy. Finally, although prior work has successfully used puppets as learners to study children's teaching (e.g., Ronfard & Corriveau, 2016), children might be less likely to engage in verbal communication with a puppet than with a human. While using puppets as learners allowed us to control for the potential effects of social feedback (or the lack thereof) from the learner and manipulate the competence of the learner without changing the learner's identity, whether children would also modulate teaching based on the learner's feedback (e.g., facial expressions, questions) is an important open question for future work.

Given the focus on these action-based measures in our study, one might also wonder whether children might have produced more informative evidence in some conditions simply by virtue of being more active. For instance, children in the Teach Toy condition (Experiment 1) might have had a nascent idea of what it means to help someone learn, and simply did everything they could think of; similarly, children in the Ordinary learner condition (Experiment 2) might have noticed that the learner "needed help to understand things" and thus performed more actions overall. However, we think this account is unlikely for a number of reasons. First, there were a number of causally irrelevant actions the children could have performed (e.g., they could have repeatedly performed more actions on the closer mat, slid both blocks on the same mat back and forth, and stacked the blocks). Nonetheless, children in the Teach Toy condition (compared to the Show Lights condition in Experiment 1) and the Ordinary learner condition (compared to the Exceptional learner condition in Experiment 2) not only generated more causally relevant actions overall than children in the other conditions but were also likely to provide these informative demonstrations immediately after the instruction to teach (as suggested by the difference in Informativeness Score). Second, we observed differences between conditions in children's actions on the Far Mat (costly actions). This suggests that children in the Teach Toy condition (Experiment 1) and the Ordinary learner condition (Experiment 2) did not simply perform more actions overall, but that their actions were targeted (at cost) to produce more causally informative evidence. Finally, in Experiment 2, the overall number of actions was actually well matched across conditions; the finding was not that children did more in general but specifically that children were more likely to transition between the mats to produce causally informative evidence when the learner was less competent.

Note that in our first experiment, we manipulated whether or not the observer wanted to learn. In real life, however, good informants not only conform to the learners' explicit requests about what they want; they consider what is genuinely beneficial for the learner. Recent work shows that children are capable of overriding others' explicit requests when they know the request would not fulfill the agents' goal (Martin & Olson, 2013). Our study similarly shows that children provide informative, tailored evidence for a learner even in the absence of explicit requests for specific kinds of information; children consider the learner's goal and competence, and adjust the information they provide to achieve those ends. Children's ability to select appropriate causal evidence for learners is especially noteworthy in light of prior work that suggests even school-aged children experience difficulty with controlling variables to generate evidence that confirms or disconfirms competing hypotheses (Chen & Klahr, 1999). One possible reason for such competence is that children were particularly motivated to provide informative evidence as teachers, while prior work placed children in the contexts of learners who are being questioned. However, in real-world pedagogical contexts, effective teachers proactively consider many things in determining what evidence to provide: the learner's individual traits, prior knowledge, epistemic and physical competence, motivation to explore, time available, etc. Our study provides a first step in studying what properties of the learner children take into account in adjusting their informative behaviors. Future work might examine not only how children *select* evidence as teachers (see also Bass et al., 2017; Rhodes et al., 2010) but also how they *structure* or *sequentially order* informative evidence to ensure accurate learning.

One interesting possibility is that the decision about what to teach and what information to provide involves not only an understanding of others' goals, knowledge, and desires (ToM) but also a consideration of their expected utility. That is, teachers may act to increase learners' perceived rewards and decrease their perceived costs. Recent work has formalized this intuition to suggest that people reason about their own and others' goal-directed actions via a *naive utility calculus* (Jara-Ettinger et al., 2016), and a number of recent empirical studies shows that even young children reason about others' actions in ways that are consistent with predictions of utility-based models (e.g., Bridgers, et al., 2016; Jara-Ettinger et al., 2015). The current work provides indirect support for the idea that children reason about expected rewards and costs. Although children were more likely to provide costly information when it was valuable to the learner, this might be either because they considered their own costs in generating the evidence, considered the learner's costs in acquiring or processing the information, or both. Future work might more directly test the idea that children's teaching is informed by a joint consideration of their own and others' utilities by systematically varying the costs and rewards to both the teacher and the learner.

Reasoning about the expected costs and benefits of teaching and learning might require relatively sophisticated inferences. The same information may have higher expected value when it is generalizable or provides the means to a higher-end goal than when it is only transiently useful. Similarly, information can be costly not only in terms of the time and effort for receiving and processing information but also the learner's situational constraints and opportunity costs (e.g., the learner might be in a rush or have insufficient time to process additional information). Thus, it is possible that we might find age-related differences in contexts where their understanding of costs and rewards depends on such additional knowledge. Further exploring these ideas will shed light on the representations and inferential processes that underlie our ability to decide what to teach and communicate, and how to help others more generally.

In this study, the measure of effective teaching was children's ability to select an appropriate set of demonstrations about a toy. Clearly, however, there is far more to teaching than selection of evidence. Caregivers and teachers in both formal and informal educational settings are not only sensitive to what the learners want to learn but also *predict* what they might like to learn, and even draw normative decisions about what the learner ought to learn. They use multiple modalities of information transfer, such as demonstrations, verbal communication, and graphical representations to instruct learneres, and even scaffold learners' own exploratory behaviors to help them generate their own evidence. Furthermore, teachers can order and structure the sequence of evidence to make learning more effective. In the current work, we used an experimental paradigm that allowed children to freely demonstrate a causal mechanism without the experience of being taught. However, the scope of children's informative behaviors was nevertheless constrained in a number of ways: the learner's learning goals were explicitly provided, and the range of possible actions was well-specified by the causal structure of the toy. While this paradigm enabled us to assess children's sensitivity to the others' goals, competence, and the cost of information, our experimental setup might have limited our ability to examine how children flexibly modulate their teaching in more naturalistic contexts. The fact that only a small number of children accompanied their demonstration with verbal instruction is suggestive of the possibility that children's teaching behaviors, while impressive, still differ from how an adult would teach about the toy. Future work might look at how children's understanding of teaching extends beyond simple demonstrations.

The ability to transition from learners to informants is critical for successful accumulation of knowledge. Furthermore, being an effective informant involves more than having knowledge about the world; it requires the ability to flexibly weigh the costs (e.g., effort) against the benefits of pedagogy and select the right kinds of evidence depending on others' wants and needs. The current study provides an important first step toward understanding the cognitive roots of this ability, suggesting that even young children generate informative evidence from what they learned from exploration, making sophisticated decisions about how to teach others.

References

Adolph, K. E., Eppler, M. A., & Gibson, E. J. (1993). Crawling versus walking infants' perception of affordances for locomotion over sloping surfaces. *Child Development*, *64*, 1158–1174. https://doi.org/10.2307/1131332

Ashley, J., & Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Social Development*, *7*, 143–163. https://doi.org/10.1111/1467-9507.00059

Bass, I., Bonawitz, E., Shafto, P., Ramarajan, D., Gopnik, A., & Wellman, H. (2017). I know what you need to know: Children’s developing theory of mind and pedagogical evidence selection. *Proceedings of the 39th annual conference of the Cognitive Science Society*, London, UK.

Bensalah, L., Olivier, M., & Stefaniak, N. (2012). Acquisition of the concept of teaching and its relationship with theory of mind in French 3-to 6-year olds. *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2011.10.008

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. https://doi.org/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. https://doi.org/10.1016/j.cognition.2010.10.001

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2016). Children consider others’ expected costs and rewards when deciding what to teach. In A. Papafragou, D. grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the Cognitive Science Society* (pp. 710–715). Austin, TX: Cognitive Science Society.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children’s imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*, 331–340. https://doi.org/10.1016/j.cognition.2010.12.001

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*, 1098–1120. https://doi.org/10.1111/1467-8624.00081

Clegg, J. M., & Legare, C. H. (2016). Instrumental and conventional interpretations of behavior are associated with distinct outcomes in early childhood. *Child Development*, *87*, 527–542. https://doi.org/10.1111/cdev.12472

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers’ exploratory play. *Cognition*, *120*, 341–349. https://doi.org/10.1016/j.cognition.2011.03.003

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. https://doi.org/10.1016/j.cognition.2014.04.013

Gweon, H., & Schulz, L. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers’ exploratory play. *Proceedings of the 30th annual meeting of the Cognitive Science Society* (pp. 570–574), Austin, TX.

Gweon, H., Shafto, P., & Schulz, L. E. (2014). Children consider prior knowledge and the cost of information both in learning from and teaching others. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 565–570). Austin, TX: Cognitive Science Society.

Gweon, H., Shafto, P., & Schulz, L. E. (2018). Development of children's sensitivity to over-informativeness in learning and teaching. Retrieved from psyarxiv.com/qtxay. https://doi.org/10.17605/osf.io/qtxay

Hagedorn, J., Hailpern, J., & Karahalios, K. G. (2008). *VCode and VData: Illustrating a new framework for supporting the video annotation workflow*. Extended Abstracts of AVI 2008.

Heyman, G., & Legare, C. H. (2013). Social cognitive development: Learning from others. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 749–766). New York, NY: Oxford University Press.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*, 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children’s understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14–23. https://doi.org/10.1016/j.cognition.2015.03.006

Kaush, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*, 288–294. https://doi.org/10.3758/BF03194066

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, *38*, e31. https://doi.org/10.1017/S0140525X14000090

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

Kretch, K. S., & Adolph, K. E. (2016). The organization of exploratory behaviors in infant locomotor planning. *Developmental Science*, 1–17. https://doi.org/10.1111/desc.12421

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants’ traits weigh heavily in young children’s trust in testimony and in their epistemic inferences. *Child Development*, *84*, 1253–1268. https://doi.org/10.1111/cdev.12029

Legare, C. H. (2011). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, *83*, 173–185. https://doi.org/10.1111/j.1467-8624.2011.01691.x

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*, 732–739. https://doi.org/10.1016/j.cognition.2008.06.013

Magid, R. W., Yan, P., Siegel, M. H., Tenenbaum, J. B., & Schulz, L. E. (2017). Changing minds: Children's inferences about third party belief revision. *Developmental Science*. https://doi.org/10.1111/desc.12553

Martin, A., & Olson, K. R. (2013). When kids know better: Paternalistic helping in 3-year-old children. *Developmental Psychology*. https://doi.org/10.1037/a0031715

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, *1*, 131–156. https://doi.org/10.1207/S15327647JCD010201

Rhodes, M., Bonawitz, E., Shafto, P., Chen, A., & Caglar, L. (2015). Controlling the message: Preschoolers' use of information to teach and deceive others. *Frontiers in Psychology*, *6*, 867. https://doi.org/10.3389/fpsyg.2015.00867

Rhodes, M., Gelman, S., & Brickman, D. (2010). Children's attention to sample composition in learning, teaching and discovery. *Developmental Science*, *13*, 421–429. https://doi.org/10.1111/j.1467-7687.2009.00896.x

Ronfard, S., & Corriveau, K. H. (2016). Teaching and preschoolers' ability to infer knowledge from mistakes. *Journal of Experimental Child Psychology*, *150*, 87–98. https://doi.org/10.1016/j.jecp.2016.05.006

Ronfard, S., Was, A. M., & Harris, P. L. (2016). Children teach methods they could not discover for themselves. *Journal of Experimental Child Psychology*, *142*, 107–117. https://doi.org/10.1016/j.jecp.2015.09.032

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070. https://doi.org/10.1111/1467-8624.00334

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*, 382–389. https://doi.org/10.1016/j.tics.2012.06.004

Schulz, L., Goodman, N., Tenenbaum, J., & Jenkins, A. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, *109*, 211–223. https://doi.org/10.1016/j.cognition.2008.07.017

Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology*, *44*, 1266–1276. https://doi.org/10.1037/0012-1649.44.5.1266

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. https://doi.org/10.1111/cdev.12502

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2- and 3-year-old children. *Developmental Psychology*, *53*, 642–651. https://doi.org/10.1037/dev0000278

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*, 779–797. https://doi.org/10.1037/a0034191

Sobel, D. M., & Letourneau, S. M. (2016). Children's developing knowledge of and reflection about teaching. *Journal of Experimental Child Psychology*, *143*, 111–122. https://doi.org/10.1016/j.jecp.2015.10.009

Sobel, D. M., & Sommerville, J. A. (2010). The importance of discovery in children's causal learning from interventions. *Frontiers in Psychology*, *1*, 176. https://doi.org/10.3389/fpsyg.2010.00176

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*, 91–94. https://doi.org/10.1126/science.aaa3799

Strauss, S., Calero, C. I., & Sigman, M. (2014). Trends in neuroscience and education. *Trends in Neuroscience and Education*, *3*, 38–43. https://doi.org/10.1016/j.tine.2014.05.001

Strauss, S., Ziv, M., & Stein, A. (2002). Teaching as a natural cognition and its relations to preschoolers' developing theory of mind. *Cognitive Development*, *17*, 1473–1487. https://doi.org/10.1016/S0885-2014(02)00128-4

Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552. https://doi.org/10.1017/S0140525X0003123X

van Schijndel, T. J. P., Visser, I., van Bers, B. M. C. W., & Raijmakers, M. E. J. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, *131*, 104–119. https://doi.org/10.1016/j.jecp.2014.11.008

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*, 343–357. https://doi.org/10.1016/j.cognition.2014.07.008

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Lagattuta, K. H. (2004). Theory of mind for learning and teaching: The nature and role of explanation. *Cognitive Development*, *19*, 479–497. https://doi.org/10.1016/j.cogdev.2004.09.003

Williamson, R., Meltzoff, A., & Markman, E. (2008). Prior experiences and perceived efficacy influence 3-year-olds' imitation. *Developmental Psychology*, *44*, 275. https://doi.org/10.1037/0012-1649.44.1.275

Wood, D., Wood, H., Ainsworth, S., & O'Malley, C. (1995). On becoming a tutor: Toward an ontogenetic model. *Cognition and Instruction*, *13*, 565–581. https://doi.org/10.1207/s1532690xci1304_7

Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development*, *19*, 457–477. https://doi.org/10.1016/j.cogdev.2004.09.002

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Mixed-effects ANOVA results