Does the intuitive scientist conduct informative experiments?: Children's early ability to select and learn from their own interventions

Elizabeth Lapidow (elapidow@ucsd.edu) & Caren M Walker (carenwalker@ucsd.edu)
Department of Psychology, University of California, San Diego, La Jolla, CA 92093

Abstract

We investigate whether children preferentially select informative actions and make accurate inferences from the outcome of their own interventions in a causal learning task. Four- to six-year-olds were presented with a novel system composed of two gears that could operate according to two possible causal structures (single or multiple cause). Given the choice between interventions (i.e., removing one of the gears to observe the remaining gear in isolation), children demonstrated a clear preference for the action that revealed the true causal structure, and made subsequent causal judgments that were consistent with the outcome observed. Experiment 2 addressed the possibility that performance was driven by children's tendency to select an intervention that would produce a desirable effect (i.e., spinning gears), rather than to disambiguate the causal structure. The results replicate our initial findings in a context in which the informative action was less likely to produce a positive outcome than the uninformative one. We discuss these results in terms of their significance for understanding both the development of scientific reasoning and the role of self-directed actions in early learning.

Keywords: cognitive development; causal learning; exploration; scientific reasoning; decision-making; experimentation

Introduction

The concept of the learner as an intuitive scientist—forming and evaluating hypotheses about the world—has provided an illuminating and productive model for understanding the mechanisms underlying cognitive development. In particular, 'Theory Theorists' have long advanced the analogy between the processes underlying knowledge acquisition and formal scientific theory change, in which children formulate, test, and rationally revise their intuitive theories in light of new evidence (Gopnik & Wellman, 2012). Indeed, much of what we know about self-directed learning in early childhood (and beyond) appears to resemble the basic inductive processes of science. From infancy, learners are sensitive to statistical information in the data they observe (e.g., Saffran, Aslin, & Newport, 1996; Xu & Garcia, 2008), and use these patterns to infer the abstract causal theories that allow for explanation, prediction, and action in the world (e.g., Carey, 1985; Keil, 1989; Wellman & Gelman, 1992).

However, the scientific process is not limited to passive observation and interpretation of statistical data. Instead, learning as an intuitive scientist also requires that children design, select, and execute informative interventions to evaluate the accuracy of their currently held beliefs and acquire new knowledge. The need for experimentation is especially apparent in the domain of causal learning, where observation alone is often insufficient. Instead, observations must typically be paired with appropriate and informative investigations in order to disambiguate between potential causes or causal structures (Pearl, 2000).

To illustrate, suppose that you notice that the houseplant sitting in a sunny spot on the windowsill has wilted, and the soil in the pot is dry. Multiple causal structures are consistent with this pattern of observation (see Figure 1): It could be that the intense sunlight dried out the soil, and the plants wilted due to this lack of moisture (a causal chain: Figure 1b). Or perhaps this is a variety of plant that requires shade, regardless of moisture. In this case, the sunlight is a direct cause of both wilting and dry soil, independently of one another (a common cause: Figure 1a).

While observation of the world alone cannot disambiguate between these two possibilities, taking specific actions on the world can. Due to the conditional relationship between patterns of intervention and causal structure, manipulating the variables in a system can reveal the causal relationships between them. That is, a learner who knows that variable X is the cause of variable Y *also* knows that intervening to change X will lead to a change in Y. Returning to our houseplant example, you could therefore discover the true casual structure by intervening to change the dryness of the soil—perhaps by watering more often—and then check to see if plants in that spot flourish (indicating a causal chain) or continue to wilt (indicating a common cause).

This makes intervention a powerful tool for determining causal structure, but its usefulness critically requires that the learner recognize and carry out *informative* interventions. For example, while intervening on the sunlight (e.g., by shading the flower pot) will always lead to improving the health of the plant, this desirable outcome would not provide information about the true underlying causal structure (i.e., whether wilting was caused by dry soil or by excess sunlight).

Whether young learners are able to engage in this type of systematic experimentation is a subject of substantial debate. On the one hand, research on exploratory play suggests that even preschool-aged children have an intuitive tendency to produce informative actions that facilitate their learning: Children preferentially explore where they have incomplete or inconsistent knowledge (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012; Gweon & Schulz, 2008; Schulz & Bonawitz, 2007), and spontaneously select actions with the potential to improve their epistemic status (Cook, Goodman, & Schulz, 2011). On the other hand, this work contrasts with decades of research on the development of scientific reasoning, which overwhelmingly reports that even much older children *do not* follow the principles of informative scientific experimentation in their spontaneous actions (Zimmerman & Klahr, 2018): Children struggle with the control and isolation of variables, often designing confounded and confirmatory experiments rather than logically informative ones (e.g., Inhelder & Piaget, 1958; Klahr, Fay, & Dunbar, 1993; Siler & Klahr, 2012; Valanides, Papageorgiou, & Angeli, 2014). Critically, children also appear to select interventions based on their tangible outcomes, rather than their informativeness (e.g., Schauble, 1990; Tschirgi, 1980) (e.g., choosing to shade the plant in the above example).

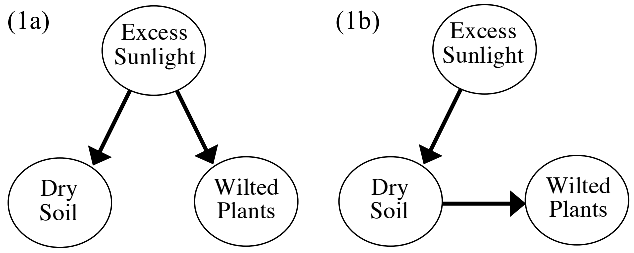

Figure 1: Common cause (a) and causal chain (b) structures.

This apparent preoccupation with producing (or reproducing) effects, rather than testing causal hypotheses, has led some researchers to suggest that children initially do not understand of the goal of scientific experimentation (e.g., Carey, Evans, Honda, Jay, & Unger, 1989; Schauble, Klopfer, & Raghavan, 1991). Instead, Schauble and colleagues (1991) proposed that early experimentation is motivated by an '*engineering'* goal, in which children engage in exploratory interventions in order to "make things happen," rather than the '*science'* goal of learning the underlying causal structure of the world. If true, this early inability or unwillingness to conduct informative experiments poses a major complication for the claim that children's self-directed learning intuitively follows a scientific process.

The current study, therefore, seeks to examine whether young children select and make inferences from their own actions in a way that supports their causal learning. While it is clear from past research that even infants successfully infer causality from observation of the outcomes of interventions that are chosen and performed by others (e.g. Meltzoff, Waismeyer, & Gopnik, 2012), it remains an open question whether the same is true for actions that children take themselves. Schulz, Gopnik, and Glymour (2007), for example, provide evidence that young learners understand and utilize the conditional relationship between causal structure and intervention. Specifically, 3- to 6-years-olds accurately identified the causal structure of a system after observing the outcomes of interventions on it *and* accurately predicted outcomes of interventions on a system when the causal structure was known.

In contrast, more recent findings indicate that even older children (5 to 8 years) may struggle to apply this principle to their *own* actions. Two studies—McCormack, Bramley, Frosch, and Lagnado (2016) and Meng, Bramey, and Xu (2018) – have examined children's causal interventions and inferences during exploration of a 3-node system. While some of the actions children produced in both studies were informative, neither team found evidence for a strong preference for informative actions. For example, according to McCormack and colleagues (2016), only 7- and 8-year-olds *consistently* selected informative interventions significantly more often than chance, while 5- and 6-year-old's did *not* select informative interventions above chance. Similarly, Meng et al. (2018) found that 5- to 7-year-olds average selection of informative interventions was not distinguishable from chance levels.

In fact, both studies found evidence that children select interventions in accordance with a positive testing strategy (PTS)—that is, taking actions that are expected to produce an effect if their current hypothesis is correct (Coenen, Rehder, & Gureckis, 2015; Klayman & Ha, 1987). In McCormack et al. (2015), the most popular intervention was turning on the hypothesized *root node*, which activated all other nodes in the system, regardless of the true causal structure. Meng et al. (2018) also provide evidence for children's use of PTS: Although the model that best captured children's intervention choices in their task relied on a combination of expected information gain and PTS, this mix was heavily skewed towards PTS.

Importantly, however, evidence *for* PTS is not evidence *against* the 'engineering goal' account: While turning on the putative root node of a system positively tests the largest number of causal links with in it (see Coenen et al., 2015), this is *also* the action that 'makes the most things happen'. Indeed, within the scientific reasoning literature, PTS behaviors are often treated as evidence that young learners are focused exclusively on the tangible outcomes of their interventions (Tschirgi, 1980; Zimmerman, 2007; Zimmerman & Glaser, 2001). These previous findings, therefore, cannot rule out the possibility that young children select primarily interventions according to 'engineering,' rather than 'scientific' goals. Thus, our first aim is to look directly at children's intervention preferences. We ask whether young learners will privilege an informative option (one that has the potential to disambiguate between competing causal structures) over an uninformative one in a forced choice design. We then examine whether children maintain their preference when this uninformative alternative is guaranteed to produce a desirable effect.

Our second aim is to examine whether children can utilize the outcomes of their own actions in later causal inference. Despite being older than the children tested by Schulz et al. (2007), participants in Meng et al. (2018) failed to identify the correct causal structure more often than chance, and the 5- to 6-year-olds in McCormack et al. (2015) did so only for certain types of structures. It is unclear whether children's failure to identify the correct causal structure was due to

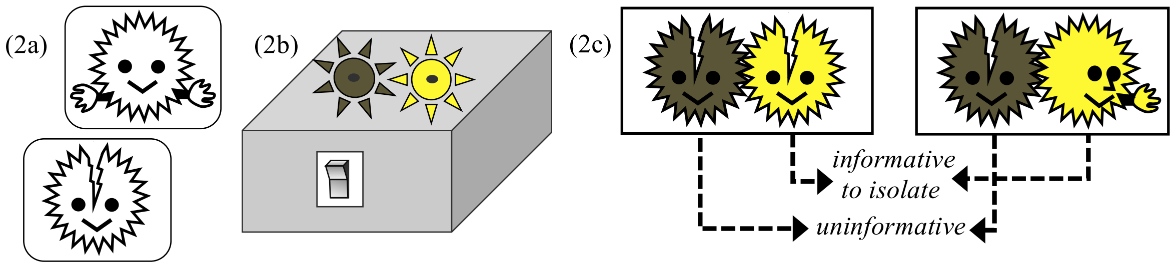

Figure 2: (a) Images used to illustrate 'working' and 'broken' gears. (b) Schematic of the gear toy. (c) Images used to illustrate the multiple causes (left) and single cause (right) structures with the informative option indicated.

their inability to make inferences from self-generated evidence, or due to the challenges associated with these more complex causal structures. In fact, Frosch and colleagues (2012) find that children struggle to make correct inferences about a similar 3-node causal system *even* when an experimenter generated the necessary evidence for them. We therefore designed the current task as a modified version of Schulz et al.'s (2007) paradigm. This is a context in which we know that young learners are able to reason about the conditional relationship between intervention and causal structure.

The current study aims to clarify whether young children preferentially select and successfully learn from their own actions in a way that is sensitive to the informative value of causal intervention. Two experiments examined how 4- to 6-year-olds responded to a forced choice between an informative and uninformative intervention in a causal learning task. Experiment 1 asks whether children will preferentially choose to take the informative intervention when selecting actions on a novel causal system. Then, in Experiment 2, the uninformative intervention is also guaranteed to produce a desirable effect. Choice behavior on this task will therefore distinguish whether children's early interventions are primarily motivated by a '*science*' or '*engineering*' goal. In addition to looking at which interventions young learners choose (and why), these experiments will also consider whether children are able to draw accurate inferences about a simple causal system from evidence they generate themselves.

Experiment 1

To investigate whether children preferentially choose interventions that support their causal learning, we used a task modeled on Schulz et al. (2007). Children were introduced to a gear toy featuring two interlocking gears and a switch. They learned that individual gears may be "working" (they spin when the toy is turned on) or "broken" (they are inert and prevent any interlocking gears from spinning). At test, children observed a pair of gears that failed to spin when the toy was turned on. They were told that this event could have resulted from two possible causal structures.[1] Either both gears are broken (a 'multiple causes' structure), or one gear is broken, preventing the other from spinning (a 'single cause' structure) (see Figure 2). As in the previous houseplant example (Figure 1), it is impossible to determine which of these represents the true causal structure from observation alone. Instead, a specific informative action must be performed: removing the gear that is broken in both structures and observing the behavior of the remaining gear *in isolation*. In contrast, removing the gear that varies between the two structures and observing the remaining (broken) gear would provide no information about the underlying causal structure. Children were given a choice between isolating and observing *only one* of the two gears prior to their inference. If young learners indeed recognize and privilege actions that are most informative for causal learning, then they should prefer to observe the gear that will disambiguate between the two structures. Afterwards, children were given the opportunity to observe the outcome of their chosen action, and were asked to judge which of the two structures was correct. If children are able to infer causal structure from their own actions, those who select the informative action should make the accurate inference.

Methods

Participants Forty-eight children (M = 64.19 months, SD = 9.46 months, range = 46-82 months) participated in Experiment 1. Children were recruited and tested individually at a local science museum in a primarily urban area. Seventeen additional children were run, but excluded due to experimental error (n = 11) or failing to complete the testing session (n = 6).

Stimuli The task used a custom-built electronic gear-toy, colored plastic gears, and picture cards with colored illustrations representing the gears and causal structures.

The toy, previously used in Schulz et al. (2007), consisted of a 12"x12" cube with two metal pegs on top. Each peg was designed to hold one 3" diameter gear, such that two gears would interlock when positioned on top of the toy. Sensors inside the cube detected the presence of a gear on the pegs, causing them to spin when a switch attached to the front of the toy was flipped to the 'on' position. A hidden control on the back of the toy allowed the experimenter to

---

[1] These structures were also based on Schulz et al (2007) and were originally referred to as 'common cause' and 'causal chain.' However, in the current experiment, it is more appropriate to refer to them as 'multiple cause' and 'single cause' structures, respectively.

surreptitiously control the supply of power (which determined whether or not the switch caused the gears to spin).

A total of six uniquely colored gears (blue, yellow, pink, green, red, orange) were used: four during the training trials and two during the test trial. Gear colors used for each part of the procedure were counterbalanced across participants. Note that in our description of the procedure, we refer to the gears using letters (A-F) in place of the color names that were actually used to identify each gear during the experiment. The picture cards (see Figure 2) each depicted a cartoon illustration of either a single gear (Figure 2a) or a gear pair (Figure 2c). These were used to illustrate the possible causal status (working or broken) and causal structures (single or multiple causes) during the task. The illustrated gears were color-matched to the physical gears used on the toy.

Procedure Each testing session began with the toy on the table in its powered state, with the switch in the 'off' position, and two gears (A and B) in place on the pegs. The experimenter introduced the toy, indicating the switch on the front, and explained that it turned the toy on and off, allowing the child to try both actions. When the child turned the toy on, A and B would spin simultaneously, and when the child turned the toy off, both stopped spinning simultaneously. The experimenter then removed and replaced each gear in turn, explaining that, when turned off, gears can be taken on and off the toy.

The experimenter then put A and B away, saying, "You're going to get to see all the gears. But some of the gears are broken. When a gear is broken, it doesn't spin even when the toy is on, and it gets in the way of other gears spinning too." Children were then shown an example working gear (A) and a broken gear (C) in turn. The experimenter placed the gear on the right peg of the toy and the child observed it either spinning (A) or not spinning (C) when the toy was turned on. Each gear was paired with a matching picture card showing its casual status. Using the pictures, the experimenter explained, "Gears that aren't broken can use their arms to spin themselves," and, "Gears that are broken don't have any arms, they cannot spin, and keep other gears from spinning too." The experimenter then held up A and C in turn and asked the child to tell them, first, whether the gear was broken or working, and second, whether it would spin on the toy on its own. Children received feedback and, if necessary, correction on each response. As part of the feedback for the second question, the experimenter placed the gear on the left peg of the toy and flipped the switch. Thus, children observed that broken and working gears operate consistently regardless of which peg of the toy they are on.

Each child then received training on the two causal structures, presented as different combinations of gears: a multiple cause (C and D) and a single cause (D and B) structure. The order in which the two structures were presented was counterbalanced, as was whether the broken gear (D) in the single cause structure was on the left or right peg of the toy. For each structure, the experimenter placed both gears on the toy and turned it on. The toy was always depowered, and the gears always remained inert. The experimenter said, "The gears aren't spinning. Something is wrong." She then brought out a picture card depicting one of the possible causal structures and described it to the child. For example, for the single cause structure, she said, "The picture shows us that just one of the gears is broken. The D gear is broken and doesn't spin on the toy, and the B gear is not broken so it can spin on the toy. But when they're together, the D gear gets in the way of the B gear, and nothing moves." Each gear was placed on the toy individually, and children were asked to predict (with feedback and observation) whether it would spin when the toy was turned on. This procedure was then repeated for the other structure.

During the test trial, the picture cards used during the training were left visible, one on either side of the toy. Gears E and F were placed on the toy and did not spin when the toy was turned on. This time, however, the experimenter said, "I don't know what's wrong here. I don't know why these gears aren't spinning. Will you help me figure it out?" The experimenter then produced two picture cards, identical to those seen during training, except that the depicted gears matched the colors of E and F. These cards were placed adjacent to the matching card from the training and each was described in the same terms. Children were told that they had to figure out which of the two pictures correctly showed why E and F weren't spinning together. Children were also told that they would get a 'clue' to help them: they could choose to see how *one* of the two gears (*either* E or F) would behave when the other gear was removed and the toy was turned on.[2]

After indicating their choice to the experimenter, children were allowed to remove the unselected gear, turn the toy on, and observe the outcome. If the informative gear was selected, the outcome (spin or inert) was counterbalanced, such that half of the children who selected the informative gear would observe evidence for the single cause structure, and the other half would observe evidence for the multiple causes structure. Regardless of choice or outcome, the experimenter would point to the gear when the toy was turned on and say, "Look!" before holding up the two picture cards depicting the possible structures, and asking children to pick the one that showed how the gears actually operated.

Results and Discussion

Children's responses to all questions were recorded during the experimental session and videotaped. We recorded

---

[2] As an attention and comprehension check, half of children (n = 24) were prompted to report the possible states of each gear before making their choice. This had no effect on either the number of informative interventions ($t(46) = -0.62$, $p = 0.538$ [*ns*]) or number of correct causal inferences ($t(32) = 1.37$, $p = 0.18$ [*ns*]), so the two scripts were combined.

whether each child chose to observe the informative or uninformative gear, as well as their final judgment about the true causal structure of the gears. For the subset of children who selected the informative gear, judgments were further coded for whether or not they were consistent with the outcome observed.

A significant majority (70.83%) chose the informative intervention, isolating and observing the gear that could disambiguate between the possible causal structures, ($p = 0.005$, two-tailed binomial). Of the 39 children who observed this disambiguating evidence, *all but two* made the correct causal inference (94.12%, $p < 0.0001$, two-tailed binomial). Together, these results suggest that young learners are not only sensitive to the informative potential of their own causal interventions, but they are also able to use the outcomes of those interventions to accurately infer the causal structure of events in the world.

Experiment 2

The results reported above provide evidence that young children preferentially select and learn from their own informative interventions in the course of causal learning. This is consistent with previous research on children's spontaneous exploration, while also extending this work to show that this preference for informative actions supports later inference. However, children's choice behavior on this task is also amenable to the opposite interpretation. As discussed above, the scientific reasoning literature often characterizes early experimenters as '*engineers*' (rather than '*scientists*') who incorrectly focus on generating effects (rather than information).

The informative gear in Experiment 1 was also the gear that had the potential to *spin* when isolated by intervention. It is possible, therefore, that children did not select the informative action because it would provide disambiguating evidence, but because it was more likely to produce this entertaining and desirable effect. If so, preference for informative action in Experiment 1 would actually be evidence *for* the claim that young children's interventions are motivated by producing effects, rather than learning about the world.

We conducted a second experiment to test this alternative. In Experiment 2, we changed the operation of the gears to include *generative* causes (i.e., working gears cause broken gears to spin), rather than inhibitory causes (i.e., broken gears prevent working gears from spinning): see Figure 3. At test, children observed a pair of spinning (rather than inert) gears that could be explained by appeal to either multiple (both gears spin) or a single cause (only one gear spins, causing the other to spin). Again, participants were given a forced choice between two interventions to determine the true causal structure.

Critically, however, this presents a choice between an uninformative action (isolating the gear that works under both structures), that is *guaranteed* to produce a desirable effect, and an informative action, (isolating the gear that works under one structure and is broken under the other), that has equivalent odds of producing or failing to produce the effect. This means that children must *forgo* the opportunity to produce a desirable effect in order to acquire information about how the causal system works.

If, as suggested by past work on exploratory play, children have an intuitive preference for informative actions, then we should continue to see a preference to isolate and observe the disambiguating gear. If, on the other hand, children show the opposite preference, choosing to select the uninformative gear, then this would suggest they are motivated by an '*engineering* goal'.

Methods

Participants Twenty-four children ($M = 65.4$ months, $SD = 9.59$ months, range = 46-82 months) were included in Experiment 2. Recruitment procedures and demographics were identical to Experiment 1. Four additional children were tested, but excluded due to experimental error ($n = 1$) or for failing to complete the testing session ($n = 3$).

Stimuli Materials were identical to those used in Experiment 1. However, new picture cards were created to depict the revised causal structures used in Experiment 2.

Procedure Procedures were similar to those used in Experiment 1. The script and outcomes of actions were modified in accordance with the revised definitions of 'broken' and 'working' gears. These changes are described below:

Children were initially told, "Some of the gears are broken. When a gear is broken, it can't spin on its own. It needs a gear that's not broken to make it spin." When shown the example gears and pictures (Figure 3), working gears were described as able to "use their little arms to spin themselves *and* to make other gears spin too!" Broken gears were described as unable to spin by themselves. Instead, broken gears "need a gear that's not broken on the toy with them to make them spin."

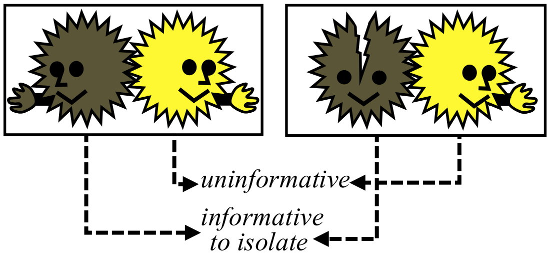

Figure 3: Illustration of the possible causal structures in Experiment 2.

In addition, the gear pairs were presented as operating according to one of two structures: Either both the gears (E and F) are working and can each spin on their own, or just one gear (E) is working, and "uses its little arms" to spin F, causing both to move. As in Experiment 1, whether the broken gear in the causal chain was the right or the left gear of the pair was counterbalanced across participants.

Results and Discussion

There were no age differences between the groups of children tested in Experiments 1 and 2, $t(70) = 0.03$, $p = 0.976$ (*ns*).

Children again selected the informative intervention significantly more often than expected by chance (79.17%, $p = 0.006$, two-tailed binomial). In fact, children's tendency to make this choice was not significantly different from their choice behavior in Experiment 1, $t(70) = 0.77$ $p = 0.442$ (*ns*). In other words, children continued to privilege the informative action *even* when it was pit against an opportunity to produce a desirable outcome.

Performance on the final inference question also did not differ from Experiment 1. Of the 19 children who selected the informative gear, *all but one* of them used this information to infer the causal structure that was consistent with the observed outcomes of their interventions (94.74%, $p < 0.0001$, two-tailed binomial). These results provide evidence against the alternative, '*engineering* goal' explanation for children's success in Experiment 1.

General Discussion

The current research sought to address two outstanding questions about children's intuitive experimentation: (1) Do children successfully identify and select informative interventions during exploration?, and (2) If so, can they draw appropriate causal inferences based on the outcomes they produce? These questions are critical, both for understanding the processes by which self-directed exploration contributes to early learning, and to address the disconnect between the claim that young learners are 'intuitive scientists,' and the claim that children are unsuccessful scientific experimenters.

First, our results demonstrate that 4- to 6-year-olds not only take informative interventions (Experiment 1), but that these actions are not driven by their potential to produce desirable outcomes (Experiment 2). These findings provide strong evidence against previous suggestions that children are initially concerned only with the practical (and not the informative) outcomes of their interventions. In particular, the '*science* vs. *engineering*' account, employed by Schauble and others (e.g., Schauble et al., 1991; Siler & Klahr, 2012) to explain children's choices in scientific reasoning tasks implies that the informative option should be less appealing than the uninformative, but productive one. The fact that the majority of children continued to select the informative action in Experiment 2 indicates instead that their choice of intervention was based on its potential to produce information and not positive outcomes. The apparent tendency to privilege producing effects seen in previous work may therefore be unrelated to children's understanding of the goals of experimentation, and an inaccurate reflection of early ability to identify and select interventions that improve their causal knowledge.

Second, these young children readily and accurately used the outcomes *of their own actions* when making judgments about the causal structure of a novel system. This goes beyond prior work showing that children make appropriate inferences after observing the outcomes of experimenter-generated interventions (Schulz et al., 2007), and contrasts with findings suggesting children may be unable to draw causal inferences from their own interventions (McCormack et al., 2016; Meng et al., 2018). In addition, while research on exploratory learning (e.g., Cook et al., 2011; Schulz & Bonawitz, 2007) has previously shown a preference for informative actions in young children, the bulk of this work has not required children to make subsequent causal inferences from the outcomes of those actions, leaving it uncertain whether and how children utilize self-directed exploration to support their learning.

Ongoing work aims to expand upon the current findings to investigate whether children are able to use the evidence generated by their own informative interventions to draw more sophisticated inferences. Specifically, we present children with cases in which the informative gear is paired with a novel gear after the intervention outcome is observed. Depending on the causal status (working or broken) of the informative gear, we can assess whether children will be able to *use* this information to update their existing causal representations, make predictions, and even draw inferences about the causal status of unknown gears.

This study also goes beyond past research on children's causal interventions (Meng et al., 2018; McCormack, et al., 2015) by directly examining intervention preference, and determining whether it is primarily driven by an action's informative potential or its tangible outcome. In contrast with previous work, the current results provide direct evidence *against* the claim that children select interventions in order to produce effects. Although our findings cannot explain children's previously reported tendency to engage in PTS, we show that this behavior is *not* due to their failure to appreciate the information-seeking goal of intervention and experimentation.

To summarize, the current results demonstrate that young children both preferentially select informative interventions, and make accurate inferences from the outcomes of those actions. These experiments fill a critical gap in the well-worn proposal that early causal learning intuitively follows a process that is analogous to belief revision in science. In sum, our findings suggest that young learners' causal interventions and inferences are sensitive to the principles of informative experimentation long before they are able execute and articulate those strategies in explicit scientific reasoning tasks.

References

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. E. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*.

Carey, S. (1985). *Conceptual change in childhood. The MIT series in learning development and conceptual change*. Cambridge, MA, US: MIT Press.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C.

(1989). ‘An experiment is when you try it and see if it works’: A study of grade 7 students’ understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*(5), 514–529.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, 102-133.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers’ exploratory play. *Cognition*, *120*(3), 341-349.

Frosch, C. A., McCormack, T., Lagnado, D. A., & Burns, P. (2012). Are Causal Structure and Intervention Judgments Inextricably Linked? A Developmental Study. *Cognitive Science*, *36*(2), 261–285.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psych Bulletin*.

Gweon, H., & Schulz, L. E. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers exploratory play. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 570–574.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence : an essay on the construction of formal operational structures*.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA, US: MIT Press.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for Scientific Experimentation: A Developmental Study. *Cognitive Psychology*, *25*(1), 111–146.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211.

McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children’s use of interventions to learn causal structure. *Journal of Experimental Child Psychology*.

Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology*, *48*(5), 1215–1228.

Meng, Y., Bramley, N., & Xu, F. (2018). Children’s causal interventions combine discrimination and confirmation. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, *49*(1), 31–57.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students’ transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, *28*(9), 859–882.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious Fun: Preschoolers Engage in More Exploratory Play When Evidence Is Confounded. *Developmental Psychology*.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, *10*(3), 322-332.

Siler, S. A., & Klahr, D. (2012). Detecting, Classifying, and Remediating: Children’s Explicit and Implicit Misconceptions about Experimental Design. In *Psychology of Science: Implicit and Explicit Processes*.

Tschirgi, J. E. (1980). Sensible Reasoning: A Hypothesis about Hypotheses. *Child Development*, *51*(1), 1–10.

Valanides, N., Papageorgiou, M., & Angeli, C. (2014). Scientific Investigations of Elementary School Children. *Journal of Science Education and Technology*, *23*(1), 26–36.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive Development: Foundational Theories of Core Domains. *Annual Review of Psychology*, *43*(1).

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223.

Zimmerman, C., & Glaser, R. (2001). *Testing Positive Versus Negative Claims: A Preliminary Investigation of the Role of Cover Story on the Assessment of Experimental Design Skills. CSE Technical Report*.

Zimmerman, C., & Klahr, D. (2018). Development of Scientific Thinking. In J. Wixted (Ed.), *Stevens’ Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed., pp. 1–25).