Cognition 201 (2020) 104315

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Informative experimentation in intuitive science: Children select and learn from their own causal interventions

Elizabeth Lapidow*, Caren M. Walker

Department of Psychology, University of California, San Diego, United States of America

ARTICLE INFO

Keywords:
Cognitive development
Causal learning
Exploration
Scientific reasoning
Decision-making
Causal interventions

ABSTRACT

We investigated 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 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 two 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. These 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. Experiment 3 serves as a control demonstrating that children's success in the previous experiments is not due to their use of low-level strategies. We discuss these findings in terms of their significance for understanding the development of scientific reasoning and the role of self-directed actions in early causal learning.

1. 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 an analogy between formal scientific theory change and the processes underlying knowledge acquisition. Like scientific theories, children's intuitive theories are formulated, tested, and rationally revised in a way that combines their existing knowledge with 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 register and perceive statistical regularities in the data they observe (i.e., “intuitive statistics”, Saffran, Aslin, & Newport, 1996; Sobel & Kirkham, 2006; Xu & Garcia, 2008), and preschool-aged children use such regularities to infer abstract causal theories that allow for explanation and prediction of events in the world (e.g., Carey, 1985; Keil, 1989; Wellman & Gelman, 1992).

However, the scientific process is not limited to observation and interpretation of data available in the environment. Similarly, intuitive science requires designing, selecting, and executing actions to *evaluate* the accuracy of one's currently held beliefs and acquire new knowledge. This need for informative experimentation is especially apparent in the domain of causal inference, as observation alone is often insufficient to determine the dependencies that exist among variables in the world. 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 planter is dry. Multiple causal structures are consistent with this pattern of observation (see Fig. 1): It could be that the intense sunlight dried out the soil, and the plants wilted due to the lack of moisture (a causal chain, Fig. 1a). 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, Fig. 1b).

While passive observation of the world cannot disambiguate between these two possibilities on its own, an *intervention*, or action that fixes the value of a single variable, can: If variable X is the *cause* of variable Y, then, intervening to change X will also lead to a change in Y (i.e., the conditional intervention principle, Woodward, 2003). Returning to our houseplant example, you could intervene to change the dryness of the soil—perhaps by watering more often—and then check

* Corresponding author at: Department of Psychology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0533, United States of America.
E-mail address: elapidow@ucsd.edu (E. Lapidow).

https://doi.org/10.1016/j.cognition.2020.104315
Received 28 May 2019; Received in revised form 23 April 2020; Accepted 24 April 2020

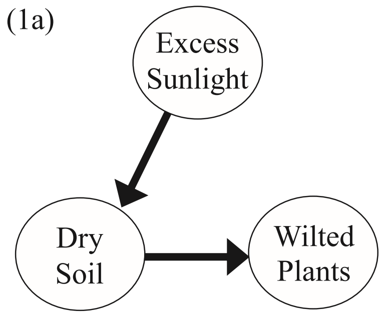

Fig. 1. Two possible causal structures for the same three variables: a causal chain (1a) and a common cause (1b).

to see if the plants in that location flourish (indicating a causal chain, Fig. 1a) or continue to wilt (indicating a common cause, Fig. 1b). Experimentation is therefore 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 planter) will always improve 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 around preschool-age, children have an intuitive tendency to produce informative actions that facilitate their learning. For example, 4- to- 6-year-olds 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; van Schijndel, Visser, van Bers, & Raijmakers, 2015).

On the other hand, this work contrasts with decades of research on the development of scientific reasoning, which overwhelmingly reports that even much older, school-aged children do not follow the principles of informative scientific experimentation in their spontaneous actions (Zimmerman, 2007; Zimmerman & Klahr, 2018). Specifically, 7- to-12-year-olds tend to 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 in these studies appear to select interventions based on their *tangible outcomes*, rather than their informativeness (e.g., Kuhn & Phelps, 1982; Schauble, 1990; Schauble, Glaser, Duschl, Schulze, & John, 1995; Siler & Klahr, 2012; Siler, Klahr, & Price, 2013; Tschirgi, 1980; Zimmerman & Glaser, 2001). That is, children often choose actions that will increase the likelihood of a desirable effect, but cannot disambiguate between the possible causal structures (e.g., choosing to shade the planter in Fig. 1).

This apparent preoccupation with producing (or reproducing) positive outcomes, rather than testing causal hypotheses, has led some researchers to suggest that young children do not understand the goal of scientific experimentation (e.g., Carey, Evans, Honda, Jay, & Unger, 1989; Schauble, Klopfer, & Raghavan, 1991). Instead, Schauble et al. (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 determining the underlying causal structure of the world. According to this theory, young learners' interventions are often uninformative because *information is not their goal*. Indeed, this explanation has persisted as a common framework for understanding choice behavior in early scientific reasoning (e.g., Masnick, Klahr, & Knowles, 2017; Siler et al., 2013; Siler & Klahr, 2012; Zimmerman, 2007).

If true, this 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 determine 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 causal relations from observing the outcomes of interventions selected and performed by *others* (e.g. Meltzoff, Waismeyer, & Gopnik, 2012), there is conflicting evidence about whether the same is true for actions that children take themselves.

For example, while Schulz, Gopnik, and Glymour (2007) provide evidence that 3- to 6-year-olds understand the conditional relationship between an experimenter's actions and the causal structure of a novel system, recent findings indicate that even older children (5- to 8-year-olds) may struggle to apply this principle when planning and interpreting their *own* actions. Schulz et al. (2007) presented children with a causal system consisting of a toy with an on-off switch and two interlocking gears (A and B). Critically, observation of both gears spinning is insufficient for determining the underlying causal structure (e.g., a causal chain in which gear A turns gear B vs. a causal chain in which gear B turns gear A). As in the previous houseplant example, the only way to determine the causal structure is to carry out informative interventions (in this case, by removing and replacing individual gears to observe whether they spin in isolation). The authors found that young learners were able to identify the correct causal structure after observing the outcomes of an experimenter's interventions on the gears, and predict the outcomes of interventions in cases when the causal structure was already known.

In contrast, McCormack, Bramley, Frosch, Patrick, and Lagnado (2016) and Meng, Bramley, and Xu (2018) found that children have difficulty producing informative interventions and making causal inferences in exploratory and forced-choice contexts. Both studies presented children with a causal system composed of three variables (e.g., lightbulbs arranged on a circuit board), with two (Meng et al., 2018) or three (McCormack et al., 2016) competing hypotheses for how they are causally connected. Children were then given the opportunity to intervene in order to determine the correct hypothesis. While some of the actions children produced were informative, neither team found evidence of a strong preference for informative actions. According to McCormack et al. (2016), only 7- and 8-year-olds selected informative interventions significantly more often than chance, while 5- and 6-year-olds did not. Meng et al. (2018) report similar failure in 5- to 7-year-olds. Instead, both studies found evidence that children (across ages) select interventions in accordance with a positive testing strategy (PTS)—that is, they took actions that were expected to produce an effect if their current hypothesis were correct (Coenen, Rehder, & Gureckis, 2015; Klayman & Ha, 1987). According to McCormack et al. (2016), the most popular intervention was to intervene on the hypothesized '*root node*'—the variable that was expected to activate the remaining two, regardless of the true causal structure of the system. Meng et al. (2018) provide further evidence that although children's intervention choices rely on a combination of expected information gain and PTS, this mix is heavily skewed towards PTS (see Nussenbaum et al. (2019) for evidence of how the relative mix of strategies systematically changes across adolescence).

Importantly, evidence *for* PTS is not evidence *against* the 'engineering goal' account: While activating the putative root node of a system positively tests the largest number of causal links within 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). This limited success documented in previous studies cannot rule out the possibility that young children primarily select interventions according to an 'engineering,' rather than a 'scientific' goal. Thus, our first aim is to look directly at children's intervention preferences. Specifically, we ask whether young learners will privilege an informative action (i.e., one that has the potential to disambiguate between competing causal structures) over an uninformative one in a forced-choice design. Then, we 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 majority of children tested by Schulz et al. (2007), participants in Meng et al. (2018) failed to identify the correct causal structure more often than predicted by chance, and the 5- to 6-year-olds in McCormack et al. (2016) did so only for certain types of structures (i.e., for common cause, but not causal chain). It remains unclear whether children's previously reported failure to identify the correct causal structure was due to their inability to make inferences from self-generated evidence or due to the challenges associated with these more complex problems. Indeed, Frosch, McCormack, Lagnado, and Burns (2012) found that children struggle to make correct inferences about a similar 3-variable 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 research goes beyond this prior work to directly examine whether young children preferentially select and make inferences from their own actions in a way that is sensitive to the informative value of causal intervention. Three experiments examine how 4- to 6-year-olds respond 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 second task will therefore illuminate whether children's early interventions are primarily motivated by a '*science*' or '*engineering*' goal. Finally, Experiment 3 serves as a critical control to test whether children's interventions and inferences are truly indicative of their causal understanding. In addition to looking at which interventions young learners choose (and why), each of these experiments will also consider whether children are able to draw accurate inferences about a causal system from evidence they generate themselves.

2. 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 might be "working" (i.e., they spin when the toy is turned on) or "broken" (i.e., 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. Children 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 Fig. 2c). As in the previous houseplant example (Fig. 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*. Because the remaining gear behaves differently under the competing hypotheses, this observation uniquely provides information that disambiguates the causal structure. In contrast, removing the gear that varies between the two structures and observing the remaining (broken) gear in isolation would result in the same behavior under both causal structures and provide no new information.

Critically, children were given a choice between isolating and observing *only one* of the two gears prior to making an inference. If young learners indeed recognize and privilege actions that are most informative for causal learning, then they should prefer to observe the informative gear. 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, then those who select the informative action should also make an accurate inference.

2.1. Method

2.1.1. Participants

Forty-eight 4- to 6-year-old children (24 female, $M = 64.19$ months, $SD = 9.46$ months, range = 46–82 months) participated in Experiment 1. The target sample size was determined using Cohen's h with a medium effect size ($h = 0.5$), and an alpha of 0.05. The sample needed to achieve a power of 0.80 ($N = 32$) was increased to accommodate complete counterbalancing. Children were recruited and tested individually at a local science museum in a primarily urban area. Seventeen additional children were tested, but excluded due to experimenter error ($n = 4$), machine malfunction ($n = 7$), or failing to complete the entire testing session ($n = 6$). Parents provided written informed consent and participants gave verbal assent prior to beginning all procedures.

2.1.2. Materials

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 surreptitiously control the supply of power (which actually 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 Fig. 2) each depicted a cartoon illustration of either a single gear (Fig. 2a) or a gear pair (Fig. 2c). These were used 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, we to refer to them as 'multiple cause' and 'single cause' structures, respectively.

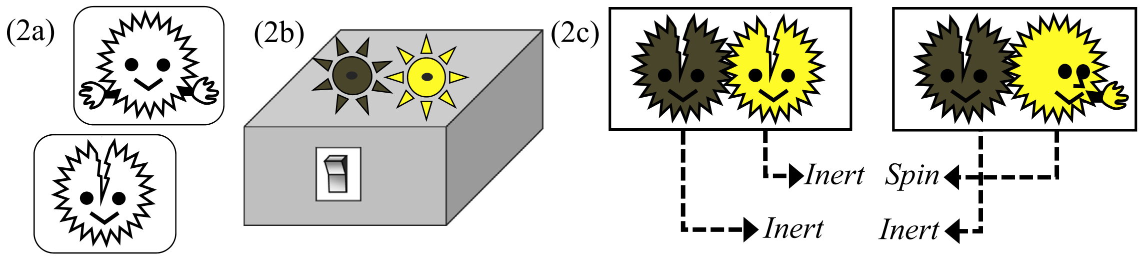

Fig. 2. Materials used in Experiment 1. (2a) Images used to illustrate 'working' and 'broken' gears. (2b) Schematic of gear toy. (2c) Images used to illustrate the multiple causes (left) and single cause (right) structures with the behavior of each gear indicated by the arrows. The informative gear (in this case, the yellow gear) is the one that behaves differently under each structure and can therefore distinguish between them. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

illustrate the possible causal status (working or broken) and causal structures (single or multiple causes) during the task. These illustrations were identical to those used in Schulz et al. (2007), in which working gears were represented with arms that allowed them to spin themselves and/or other gears, while (cracked) broken gears did not. The illustrated gears were color-matched to the physical gears used on the toy.

2.1.3. 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 causal status. Using the pictures, the experimenter explained, "Gears that are not 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] As noted above, only one of these two options is informative, since it will either spin or remain inert, depending on the true causal structure of the system. The uninformative gear, in contrast, is guaranteed to remain inert under both possible structures. Thus, children are presented with a forced choice between an informative and uninformative gear.

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 cause 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.

2.2. Results and discussion

Children's responses to all questions were recorded during the experimental session and videotaped. We recorded 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. Below, we report all statistics using both frequentist tests and a Bayesian approach (expressed as Bayes factors [BFs], quantifying the likelihood of the data under the alternative hypothesis compared to the

[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 ($t(46) = -0.62$, $p = .538$ [*ns*]) or the number of correct causal inferences ($t(32) = 1.37$, $p = .18$ [*ns*]), so the two scripts were combined for all analyses.

null hypothesis).

A significant majority of children (70.83%) chose the informative intervention, isolating and observing the gear that could disambiguate between the possible causal structures, ($p = .005$, two-tailed binomial). We also conducted a Bayesian analysis to obtain a full posterior distribution over parameter estimates, using a uniform prior, $\theta \sim$ Beta (1, 1). The Bayes factor, $BF_{+0} = 23.775$, 95% credible intervals: [0.57, 0.82], provides strong evidence for the hypothesis that children have a preference for the informative gear. A logistic regression, treating age as a continuous factor and choice of informative gear as the dependent variable revealed no effects of age on performance (Wald, $z = 0.756$, $p < .450$).

Furthermore, of the 39 children who observed this disambiguating evidence,[3] *all but two* made the correct causal inference (94.12%, $p < .0001$, two-tailed binomial, $BF_{10} = 874{,}961.507$, 95% credible intervals: [0.81, 0.98]). 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 apply the outcomes of those interventions to accurately infer the causal structure of events in the world.

3. Experiment 2

The results reported above provide evidence that young children preferentially select and make accurate inferences 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).

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

The second experiment tests this alternative interpretation. 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 Fig. 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 find a tendency 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'.

3.1. Method

3.1.1. Participants

Forty-eight children (27 female, $M = 64.9$ months, $SD = 8.8$ months, range = 48–81 months) were included in Experiment 2. Recruitment procedures and demographics were identical to Experiment 1. Ten additional children were tested, but excluded due to experimenter error ($n = 3$), sibling or caregiver interference ($n = 2$), or failing to complete the entire testing session ($n = 5$).

3.1.2. Materials

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

3.1.3. Procedure

The task was similar to Experiment 1. However, the outcomes of each action 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 (Fig. 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."

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. All other procedures and coding criteria remained the same.

3.2. Results and discussion

There were no significant age differences between the samples of children tested in Experiments 1 and 2, $t(94) = 0.25$, $p = .802$ (*ns*).

Children again selected the informative intervention significantly more often than expected by chance (66.67%, $p = .029$, two-tailed binomial). Bayesian analysis, using a uniform prior, also supports the hypothesis that children have a preference for the informative gear, $BF_{+0} = 5.041$, 95% credible intervals: [0.54, 0.78]. Children's tendency to make this choice was not significantly different from their choice behavior in Experiment 1, ($p = .527$, two-tailed binomial), replicating our previous findings. In other words, children continued to privilege the informative action *even* when it was pit against an opportunity to produce a desirable outcome. Again, a logistic regression revealed no effects of age on choice of the informative intervention (Wald, $z = 1.005$, $p < 0.315$).

Performance on the final inference question also did not differ from Experiment 1. Of the 32 children who selected the informative gear, *all but one* used this information to infer the causal structure that was consistent with the observed outcomes of their interventions (96.88%, $p < .0001$, two-tailed binomial, $BF_{10} = 4{,}067{,}000$, 95% credible intervals: [0.84, 0.99]). These results provide evidence against the alternative, '*engineering* goal' explanation for children's success in Experiment 1.

[3] Although all children were asked to make a final inference, only those who generated an informative intervention observed evidence that would support further inferences that could be meaningfully classified as correct or incorrect.

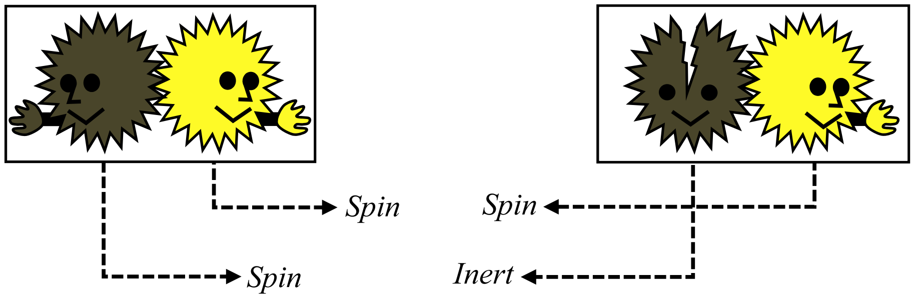

Fig. 3. Illustration of the causal structures presented in Experiment 2 with the behavior of gears under each structure indicated by the arrows. Unlike Experiment 1, the gear that is most likely to produce a desirable outcome (in this case, the yellow gear) is not informative. Children must be willing to forego this option in order to conduct an informative intervention (i.e., in this case, isolating the green gear). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Experiment 3

Experiments 1 and 2 provide evidence that children select informative interventions, and use the outcomes of their own actions to make accurate causal inferences. However, the study design leaves open an alternative explanation for children's success. Specifically, the picture cards used in both experiments to scaffold children's forced-choice depict the informative gear in different states in each image (spinning vs. inert), while the uninformative gear remains the same in both images (spinning vs. spinning or inert vs. inert). These images may have inadvertently provided children with a lower-level perceptual cue that highlighted the uncertainty (or novelty) associated with the informative gear. Furthermore, low-level 'matching' of the behavior of the observed gear to the card image may have facilitated children's selection of the correct causal structure. If so, children may have succeeded in Experiments 1 and 2 without any causal understanding.

Experiment 3 was therefore designed as a control task to remove all superficial cues (see Fig. 4). To do so, children were asked to determine the causal status of a single, unknown gear by choosing to pair it with one of two known gears. Images were only used to introduce and explain the difference between broken and working gears, and could not be used for low-level comparison or mapped to the outcome of interventions. Instead, selecting an informative action required understanding the causal properties of the system and potential outcomes (and implications) of interventions on that system. In addition, Experiment 3 introduced a pair of follow-up questions that required children to extend what they learned from their own intervention to determine the causal status of an entirely new, unknown gear. If children's performance on this revised task parallels their performance in Experiments 1 and 2, it would provide strong evidence of sophisticated causal reasoning.

4.1. Method

4.1.1. Participants

Forty-eight children (24 female, M = 64.2 months, SD = 10.3 months, range = 48–81 months) were included in Experiment 3. Children were tested in one of two conditions that matched the causal rules presented in Experiments 1 and 2: an Inhibitory condition (n = 24, M = 64.5 months, SD = 11.6 months, range = 49–81 months) and a Generative condition (n = 24, M = 63.8 months, SD = 9.2 months, range = 48–79 months). Recruitment procedures and demographics were identical to those of the first two experiments. Eleven additional children were tested, but excluded due to experimenter error (n = 4), sibling or caregiver interference (n = 2), machine malfunction (n = 2), or failing to complete the entire testing session (n = 3).

4.1.2. Materials

The stimuli consisted of the machine, gears, and training images (Fig. 2a) used in Experiments 1 and 2.

4.1.3. Procedure

Children were tested in one of two conditions. In the Inhibitory condition, as in Experiment 1, broken gears *prevent* working gears from spinning. In the Generative condition, as in Experiment 2, working gears *cause* broken gears to spin.

The beginning of the procedure was identical to the previous two experiments: The experimenter introduced the machine with two gears (A and B) in place, explained the role of the switch, and then demonstrated a working gear (C) and a broken gear (D). The operation and explanation of these gears was identical to either Experiment 1 or Experiment 2, depending on children's assigned condition. Again, these gears were paired with a matching picture card that depicted their causal status (broken or working, Fig. 4b) and children's understanding was checked. However, unlike in the previous experiments, these images and the two known gears remained visible throughout the rest of the experiment.

Next, the experimenter proceeded to the test trial, presenting two new, unknown gears (E and F). The experimenter said, "I don't know how these gears work. I don't know if they're broken or not broken. Will you help me figure it out?" The experimenter then moved gear F out of sight, and placed gear E on one peg of the toy (Fig. 4a). Children were told they had to figure out whether or not E was broken, and that they would get a 'clue' to help them: They were allowed to choose *one* of the two gears from the introduction (*either* C or D) to observe on the toy

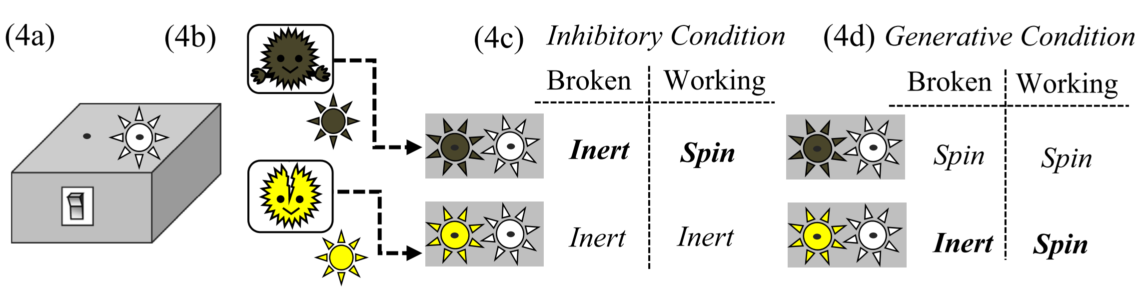

Fig. 4. Design of Experiment 3. (4a) The gear toy is presented with an unknown gear (white) on one peg. (4b) Children are given the choice of placing one of two known gears on the other peg: a working gear (top) or a broken gear (bottom). (4c-d) The pairing created by each choice and their possible outcomes (spin or inert), depending on the true causal status (broken or working) of the unknown gear in Inhibitory (4c) and Generative (4d) conditions. The informative option is the one that leads to different outcomes under each possibility and can therefore distinguish between them (i.e., the working gear in 4c and the broken gear in 4d).

along with gear E to see how the two gears would behave together when the toy was turned on. That is, they could choose to pair the unknown gear (E) with either a working gear (C) or a broken gear (D). As in the previous experiments, only *one* of these two possible interventions disambiguates between the two causal structures, which varied by condition (i.e., the working gear (C) is informative in the Inhibitory condition, and the broken gear (D) is informative in the Generative condition). In both cases, the gear that is informative is the one that will behave *differently* when paired with the unknown gear, depending upon the unknown gear's causal status. The other gear will behave the *same* regardless of the status of the gear it is paired with, and is therefore uninformative (see Fig. 4c and d). Thus, if children apply their causal understanding to determine those interventions that are most informative, they should select the opposite gears in each of the two conditions.

Following their choice, children made their selected intervention and observed its outcome. As in previous experiments, the outcome (spinning or inert) was alternated across participants who selected the informative gear. Afterwards, children were asked to infer whether gear E was broken or not, based on the evidence they generated.

Finally, to assess whether children could extend this newly acquired causal knowledge to inform additional inferences, those who answered correctly were also asked a follow-up question. The experimenter presented the *other* unknown gear (F), and used it to replace the informative gear on the toy. Children were asked one of two questions about this pair of gears, composed of the previously unknown gear (E) and the currently unknown gear (F). One type of follow-up question asked participants to make a *prediction* about whether E and F would spin when the toy was turned on. The other type of follow-up question asked participants to make an *inference* about the causal status of F (working or broken), after observing E and F spin or remain inert (alternated) when the toy was turned on. The type of follow-up question asked varied according to the causal status of gear E that was observed during the experiment (see Table 1).

4.2. Results and discussion

An analysis of variance showed no significant age differences among any of the three experiments, $F(2,141) = 0.205$, $p = .815$ (*ns*).

Even after removing low-level cues, a significant majority of children continued to select the informative gear (72.92%, $p = .002$, two-tailed binomial), which did not differ from Experiment 1 ($p = .87$) or Experiment 2 ($p = .44$). Bayesian analysis indicates strong evidence for the hypothesis that children prefer the informative option: $BF_{+0} = 59.508$, 95% credible intervals: [0.59, 0.83]. The numeric difference in frequency of informative choices between the Inhibition (79.16%) and Generative (66.7%) conditions matched the difference between Experiments 1 and 2, and was similarly non-significant $p = .278$. Logistic regression revealed no effects of age on choice of the informative intervention (Wald, $z = -0.639$, $p < .523$).

Experiment 3 also supports and extends evidence for the claim that children accurately utilize evidence generated through their own interventions in subsequent causal inferences. First, we replicated children's success on the initial question, with 32 of 35 children making the correct causal inference following their own informative intervention (91.43%, $p < .0001$, two-tailed binomial, $BF_{10} = 145{,}826.92$, 95% credible intervals: [0.77, 0.97]). Additionally, children succeeded on both types of follow-up questions, which required that they apply their knowledge to inform either a novel *prediction* or *inference*. Of the 32 children who received a follow-up question,[4] 93.75% answered correctly ($p < .0001$, two-tailed binomial, $BF_{10} = 262{,}400.25$, 95% credible intervals: [0.81, 0.98]). Looking at each type of follow-up question in turn, 16 of 18 children made a correct *prediction,* and all 14 children made a correct *inference*. Taken together, the results of Experiment 3 replicate and extend our previous findings.

Table 1
Type of follow-up questions asked in Experiment 3.

Condition	Children's Inference After Intervention	Question Type
Generative	Gear E is broken	*Inference*
	Gear E is working	*Prediction*
Inhibitory	Gear E is broken	*Prediction*
	Gear E is working	*Inference*

5. 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 prefer to take informative interventions (Experiment 1), but that these actions are not exclusively driven by their potential to produce desirable outcomes (Experiment 2). In Experiment 3, we replicate these results and also rule out the alternative that children's success was due to their reliance on a lower-level, non-causal strategy. These findings provide evidence against previous '*science* vs. *engineering*' account, which suggests children are initially concerned only with the practical (and not the informative) outcomes of their interventions (e.g., Schauble et al., 1991; Siler & Klahr, 2012). While it is clear from prior work that the desire to produce effects is one factor influencing children's choice of actions (see Lapidow & Walker, 2020 for an explanation of this behavior), the fact that young learners selected the informative action over the productive action in Experiment 2 (and in the Generative condition of Experiment 3) indicates that children's interventions are sensitive to the epistemic goal of experimentation. The current study therefore goes beyond past research on children's causal intervention (Meng et al., 2018; McCormack et al., 2016) by providing direct evidence against the claim that children select actions in order to '*engineer*' effects.

Second, children in the current experiments made ready and accurate use the outcomes of their own actions in novel causal inferences. This goes beyond prior work showing that children make appropriate inferences after observing the outcomes of experimenter-generated interventions (e.g., Schulz et al., 2007; Sobel & Kirkham, 2006), and contrasts with findings suggesting that children may be unable to draw causal inferences from their own explorations (McCormack et al., 2016; Meng et al., 2018). In addition, while previous research on exploratory learning (e.g., Cook et al., 2011; Schulz & Bonawitz, 2007) has shown young children's preference for informative actions, the bulk of this work has not required children to make subsequent causal inferences from the outcomes of those actions,[5] leaving it unclear whether and how children utilize self-directed exploration to support their learning.

[4] Only children who correctly answered the initial inference question about gear E were asked the follow-up question about gear F. This was done so that children's performance on these additional questions reflects their ability to extend their own causal inferences, rather than the experimenter's corrective feedback, to novel instances.

[5] Past work by Bonawitz et al., 2012 and van Schijndel et al., 2015 provides some evidence for this ability in older children (4- to 7-year-olds and 4- to 9-year-olds, respectively). However, both of these studies concerned improvement in children's predictions about causal phenomena (balance and shadow size) following their exploration of belief-inconsistent observations, not whether they make accurate inferences about causal structure from self-generated evidence.

Here, we provide strong evidence that children not only grasp the causal implications of the data they generate; they also apply this knowledge to inform their generalizations about novel causal variables.

Future work is needed to examine children's interventions and inferences in more complex and contextualized learning problems. As previously mentioned, the current task was designed to present a relatively simple causal system that does not rely on existing content knowledge (Schulz et al., 2007). This method has limited resemblance to real-world experimentation, in which the number of possible structures, interventions, and variables is often vast and uncertain. However, having established early success in children's performance on a small-scale, decontextualized problem, we can begin to investigate how these behaviors are expressed in more realistic contexts. In particular, future research should examine intervention and inference in causal systems with a greater number of variables (e.g., Meng et al., 2018; McCormack et al., 2016) and in domains in which learners hold prior beliefs (e.g., Schauble et al., 1995; Tschirgi, 1980).

To summarize, the current results demonstrate that young children both preferentially select informative interventions, and make accurate inferences from the outcomes they generate, during their own actions on a novel causal system. These experiments fill a critical gap in the well-worn proposal that early causal learning intuitively follows a process that is analogous to knowledge acquisition in science. Our findings further 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.

CRediT authorship contribution statement

Elizabeth Lapidow: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Caren M. Walker**: Writing - original draft, Writing - review & editing, Supervision.

Acknowledgements

This work was supported by the Hellman Fellowship to C. M. Walker, the San Diego Fellowship, and the National Defense Science and Engineering Graduate Fellowship to E. Lapidow. We are grateful to the parents and children who participated in the study, as well as the Reubin H. Fleet Science Museum for their help with recruitment. Finally, we thank Carolyn Collora, Sally Tang, and Nicky Sullivan for their efforts towards data collection.

References

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002.

Carey, S. (1985). Conceptual change in childhood. *In The MIT series in learning development and conceptual change*MIT Presshttps://doi.org/10.1016/S0016-6995(85)80176-5.

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. https://doi.org/10.1080/0950069890110504.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology, 79*, 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. https://doi.org/10.1016/j.cognition.2011.03.003.

Frosch, C., 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. https://doi.org/10.1111/j.1551-6709.2011.01208.x.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin.*. https://doi.org/10.1037/a0028044.

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* (pp. 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.* Basic Books.

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

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*(1), 111–146. https://doi.org/10.1006/cogp.1993.1003.

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

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. *Advances in Child Development and Behavior, 17*(C), 1–44. https://doi.org/10.1016/S0065-2407(08)60356-0.

Lapidow, E., & Walker, C. M. (2020). *The search for invariance: Repeated positive testing serves the goals of causal learning. In Language and Concept Acquisition from Infancy Through Childhood. (pp. 197–219)*Cham: Springerhttps://doi.org/10.1007/978-3-030-35594-4_10.

Masnick, A. M., Klahr, D., & Knowles, E. R. (2017). Data-driven belief revision in children and adults. *Journal of Cognition and Development.*. https://doi.org/10.1080/15248372.2016.1168824.

McCormack, T., Bramley, N. R., Frosch, C., Patrick, F., & Lagnado, D. A. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology.*. https://doi.org/10.1016/j.jecp.2015.06.017.

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. https://doi.org/10.1037/a0027440.

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

Nussenbaum, K., Cohen, A. O., Davis, Z., Halpern, D., Gureckis, T. M., & Hartley, C. A. (2019). Causal information-seeking strategies change across childhood and adolescence. *PsyArXiv.*. https://doi.org/10.31234/osf.io/qukac.

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. https://doi.org/10.1126/science.274.5294.1926.

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. https://doi.org/10.1016/0022-0965(90)90048-D.

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences, 4*(2), 131–166.

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. https://doi.org/10.1002/tea.3660280910.

Schulz, L. E., & Bonawitz, E. B. (2007). *Serious.* Fun: Preschoolers Engage in More Exploratory Play When Evidence Is Confounded. Developmental Psychologyhttps://doi.org/10.1037/0012-1649.43.4.1045.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*(3), 322–332. https://doi.org/10.1111/j.1467-7687.2007.00587.x.

Siler, S. A., & Klahr, D. (2012). Detecting, classifying, and remediating: Children's explicit and implicit misconceptions about experimental design. *Psychology of science: Implicit and explicit processes*https://doi.org/10.1093/acprof:oso/9780199753628.003.0007.

Siler, S. A., Klahr, D., & Price, N. (2013). Investigating the mechanisms of learning from a constrained preparation for future learning activity. *Instructional Science, 41*(1), 191–216.

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology, 42*, 1103–1115. https://doi.org/10.1037/0012-1649.42.6.1103.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*(1), 1–10. https://doi.org/10.2307/1129583.

Valanides, N., Papageorgiou, M., & Angeli, C. (2014). Scientific investigations of elementary school children. *Journal of Science Education and Technology, 23*(1), 26–36. https://doi.org/10.1007/s10956-013-9448-6.

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.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of Core domains. *Annual Review of Psychology, 43*(1), 337–375. https://doi.org/10.1146/annurev.ps.43.020192.002005.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* Oxford University Press.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, 105*(13), 5012–5015. https://doi.org/10.1073/pnas.0704450105.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001.

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.* (In 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)*John Wiley & Sons, Inc.https://doi.org/10.1002/9781119170174.epcn407.