Children's exploration as a window into their causal learning

Sophie Bridgers (sbridge@stanford.edu)
Department of Psychology, Stanford University

Yvonne Wang (yvonnewjy.wang@mail.utoronto.ca)
Daphna Buchsbaum (buchsbaum@psych.utoronto.ca)
Department of Psychology, University of Toronto

Abstract

How do children's beliefs about a causal system influence their exploration of that system? Children watched an experimenter try to make a machine play music by placing blocks on top; one block always activated the machine and the other block never did (Deterministic condition), or one block activated the machine a higher proportion of times than the other (Probabilistic condition). Subsequently, we measured children's exploratory behaviors without feedback (the machine never activated). We predicted that children in the two conditions would differ in their beliefs about how the system should work, leading to different hypotheses about why the machine was no longer working, and to differential exploration. Compared to the Probabilistic condition, children in the Deterministic condition intervened more often with the previously more effective block, experimented more with how to activate the machine, and explored for less time. Children's exploration provides a rich, nuanced view of their causal reasoning.

Keywords: cognitive development; causal learning; causal uncertainty; statistical learning; exploration

Introduction

Children continually learn about the causal relationships in their environment: Switches turn on lights; germs make people sick; unkind words make people sad. Learning such relationships is not trivial because causal links are not directly observable but rather inferred from statistical contingencies between events. Furthermore, causal relationships are graded in strength – you not only need to learn that a relationship exists but also the probability of the cause generating the effect, which in turn can lead to more or less certainty about the underlying causal structure (Griffiths & Tenenbaum, 2005).

From as young as 24 months, children can infer both the existence of causal links and the relative strength of different causes from deterministic and probabilistic data (Waismeyer, Meltzoff, & Gopnik, 2015; for older children see, Gopnik et al., 2004; Kushnir & Gopnik, 2007; Sobel, Tenenbaum, & Gopnik, 2004). Further, children are more likely to trust testimony that conflicts with probabilistic data (e.g., a block only sometimes makes a machine play music) than testimony conflicting with deterministic data (e.g., a block always makes a machine play music), suggesting sensitivity to the relative strength implied by these different patterns of evidence (Bridgers, Buchsbaum, Seiver, Griffiths, & Gopnik, 2016).

The majority of prior research on children's causal reasoning, however, has examined children's judgments of causal structure. That is, children are prompted to identify which objects are causes and which are not or to produce a single intervention on the system to bring about an effect (e.g., Bridgers et al., 2016; Buchanan & Sobel, 2011; Gopnik et al., 2004; Kushnir & Gopnik, 2005; Kushnir, Wellman, & Gelman, 2008; Sobel et al., 2004). These forced-choice dependent measures provide insight into children's inferences about what is and what is not a cause, but more graded measures could provide additional insight into children's sensitivity to causal strength, especially since causal strength itself is inherently graded. Such measures could also reveal children's certainty or confidence in their inferences. Here, we look to children's exploratory behavior as a window into their causal learning in the hope of gaining a more nuanced view.

Children are sophisticated active learners and acquire knowledge from their own direct exploration of the world (Schulz, 2012; Xu & Kushnir, 2013). Children's exploratory play is not just driven by their enjoyment but also by their inductive inferences and expectations about how the world works. For instance, if children learn that members of a category have an unobservable property, they expect other category members to also have that property and attempt to elicit the property from them (Baldwin, Markman, & Melartin, 1993; Butler & Markman, 2012). The stronger the cues provided to category membership are (e.g., the objects are similar vs. different in appearance; have the same vs. different labels), the longer children persist in their attempts to elicit the property, revealing sensitivity to gradations in the inductive strength of these cues (Baldwin et al., 1993; Schulz, Standing, & Bonawitz, 2008). Children also explore more when evidence is ambiguous or an event challenges their prior beliefs. They opt to play with causally confounded or belief-violating toys over novel toys (Schulz & Bonawitz, 2007; Stahl & Feigenson, 2015). In this play, they even spontaneously design novel interventions or experiments to test their beliefs, and generate sufficient evidence to disambiguate the causal system (E. B. Bonawitz, van Schijndel, Friel, & Schulz, 2012; Gweon & Schulz, 2008; Legare, 2011; Schulz & Bonawitz, 2007). Taken together, childrens certainty and surprise appear to influence how much and how long they explore, making exploration a good dependent measure of their underlying beliefs and confidence in those beliefs.

Here, we take up the hypothesis that children's free exploration of a causal system is supported by some of the same rational principles of inductive inference that inform their explicit causal judgments (e.g., Schulz, Standing, & Bonawitz, 2008). We examine children's exploration in the context of causal uncertainty, and in particular investigate how their exploration might differ after observing the deterministic vs. probabilistic patterns of evidence often used in classic experiments on causal judgments. We predicted that children would be sensitive to differences in causal strength implied by deterministic vs. probabilistic data, and that this would be reflected in their free play with a causal system. Using exploration as a dependent measure has the additional benefit that it is non-verbal and so could be particularly useful in measuring young children's certainty in their inferences; explicit meta-cognition is still developing in early childhood, making it difficult to elicit explicit certainty judgments from young children (Ghetti, Hembacher, & Coughlin, 2013; Hembacher & Ghetti, 2014).

A feature common to experiments using exploration as a dependent measure is that children are presented with objects that lack the previously observed property. This design decision prevents children from eliciting confirmatory evidence and eliminates the distraction of the interesting causal property, to more easily isolate how children's expectations affect their play (see Baldwin et al., 1993; Schulz, Standing, & Bonawitz, 2008). However, it also raises the question of what conclusions children are drawing about their failed attempts to elicit the property. This failure could be due to the causal system (i.e., it does not work as expected or has stopped working) or due to one's own actions (i.e., I am doing something wrong) (Karmiloff-Smith & Inhelder, 1974), a distinction to which even 16-month-old infants are sensitive when responding to their own failed actions (Gweon & Schulz, 2011). Thus, children's exploration may not only reflect uncertainty in their prior inferences about the causal system but also uncertainty about why the system is no longer working, making it a useful measure of both their initial inductive predictions and how such predictions inform later inferences about unexpected outcomes. Next, we present our specific experimental hypotheses about how children's sensitivity to these sources of uncertainty might manifest in their exploration of an inert causal system.

Overview

In the current experiment, four- and five-year-olds were introduced to a machine that could be activated by placing blocks on top. Children observed either deterministic or probabilistic evidence that one of two blocks was better than the other at making the machine go. Children were then given the opportunity to explore the blocks and the machine on their own, but during this time, the machine never activated no matter what interventions children performed.

We predicted that children's exploration would reflect both their initial inferences from the demonstrated data about how the system works and their subsequent inferences about why it was no longer working (Legare, 2011; Schulz, Hooppell, & Jenkins, 2008). In both conditions, we predicted children would first attempt to activate the machine with the block that was demonstrated to be more causally efficacious. However, when faced with evidence that this block was no longer working, children would explore differently across conditions.

Children in the Deterministic condition will likely develop a strong expectation that the previously more effective block should work (it always did before) and that the the previously less effective block should not (it never did before), while children in the Probabilistic condition should have less certainty about the causal strength of each block (both blocks previously succeeded and failed in activating the machine). We thus predicted that children in the Deterministic condition would persist in trying to activate the machine with the previously better block more than children in the Probabilistic condition, who would be more likely to explore the previously worse block.

The stronger belief in the previously better block's efficacy might also lead children in the Deterministic condition to infer that the block is no longer working because they are doing something wrong. If that is the case, children might not only persist in trying this block but also be more likely to experiment with different ways of activating the machine (e.g., placing the block in different locations on the machine) to try and find the right way to use the block (Legare, 2011). However, this stronger initial belief in the better block's causal strength might also lead children to give up more quickly, because of a belief that they are still doing something wrong (e.g., not placing the block in the right location) or that the system has somehow changed (e.g., it is out of batteries). In contrast, children in the Probabilistic condition might explore longer overall, but produce less variable interventions. Since the system is stochastic, if it is not activating it is not necessarily because the system has stopped working or that they are doing something wrong, so it is worth continuing to test out the blocks.

Prior work has actually suggested that children engage in more variable exploration when presented with ambiguous or probabilistic evidence (see E. Bonawitz et al., 2011; Schulz, Hooppell, & Jenkins, 2008). Here, we predict that the probabilistic evidence will result in more even exploration of both blocks and longer exploration overall. In contrast to previous work, we anticipate that the deterministic evidence will render the inert system more surprising and so will lead to more novel interventions to try to figure out why this is the case.

Given these predictions, we not only look at children's overall persistence and exploration time, but also at which blocks children place on the machine and what else they try with the blocks and the machine to see if these behaviors also reflect different causal inferences and sources of uncertainty.

Methods

Participants

Seventy-seven children (41 4-year-olds and 36 5-year-olds; 38 females) were recruited from local museums in Toronto, Ontario. An additional 10 children were tested but excluded from analysis due to experimenter error ($n = 5$), missing date of birth ($n = 2$), or ending the experiment early ($n = 3$). Children were randomly assigned to the Deterministic condition ($n = 39$, $M(SD) = 59.43(6.30)$ months, 19 females) or the Probabilistic condition ($n = 38$, $M(SD) = 58.96(7.56)$ months, 19 females). The diversity of the sample was representative of the diversity of the local population.

Materials

The causal system was presented to children as a machine that could play music when blocks were placed on top. The "machine" was a decorated cardboard box as shown in Figure 2A. There were four wooden blocks, differing in shape and color, but similar in size: the red oval, yellow square, blue triangle, and purple star blocks. The machine appeared to play music when blocks were placed on top, but in reality, it contained a wireless door bell that could be activated surreptitiously by the experimenter via a hidden remote control. For some children, a bell they could ring to indicate that they were done exploring ($n = 36$) or a distractor toy ($n = 2$) were also placed on the table.[1]

Procedure

Participants were tested individually in a quiet off-exhibit location at the museum. Children's behavior was video recorded and coded offline by the second author.

Demonstration Phase The experimenter first introduced the child to the novel machine, explained that you could make the machine play music by putting blocks on top and that some blocks made it play music while others did not. The experimenter then brought out a pair of blocks, either the yellow square and red oval blocks, or the blue triangle and purple star blocks (counterbalanced across participants) and told children that she had never played with these blocks before.

Next, the experimenter demonstrated each block on the machine. In the Deterministic condition, each block was placed on the machine six times; one block deterministically activated the machine on all six trials (*better* block), while the other block failed to activate the machine on all six trials (*worse* block). In the Probabilistic condition, one block was placed on the machine three times and activated the machine twice, always on the first and third trial (*better* block); the other block was placed on the machine six times and also activated it two times, on the second and fifth trial (*worse* block). As in previous work, this pattern of data controls for the frequency with which each block activates the toy, providing stronger evidence that children are reasoning about the probability of each cause generating the effect and not just the number of times the effect was associated with the cause (Bridgers et al., 2016; Kushnir & Gopnik, 2007). We counterbalanced which block was the better block and whether the better block was demonstrated first.

The experimenter then asked the child which block was better at making the machine play music. If the child did not answer or chose both blocks, the experimenter asked them to select one. The block children pointed to, touched, or named was coded as their answer. If children answered incorrectly, the experimenter corrected them (e.g., "Oh remember, the red oval block was better at making the machine play music.")

Exploration Phase The experimenter then informed the child that she had to leave for a bit, and that the child could play freely with the blocks and the machine while she was gone. The experimenter left the table and pretended to be busy in another part of the room. During this time, the child could explore the blocks and machine but did not receive any feedback, i.e., neither block activated the machine. If the child asked the experimenter questions, she explained that she was still working but the child could keep playing with the blocks and machine, and let the experimenter know when they were done. The experimenter returned when the child indicated they were done or after two minutes.[2] Lastly, the child was given an opportunity to activate the machine with a new pair of blocks to ensure they left in good spirits.

Exploration Phase Coding There were three main variables of interest. First, we measured the proportion of interventions children performed with the better block, out of the total interventions they made with just a single block (*single trials*). A single trial was when children placed a block that was off of the machine onto the machine, or lifted and then put back down part or all of a block that was already on the machine. Each single trial was coded according to whether the *better* or *worse* block was used. Only children who placed a block on the machine were included in this analysis (Deterministic condition: $n = 33$; Probabilistic condition: $n = 35$). A small number of trials (3.4%) in which children placed both blocks on the machine simultaneously were not included, but were considered a strategy as described below.

Second, we measured the total time in seconds children spent exploring the blocks and machine. Exploration began when children first touched a block or the machine and ended when they met any of the following criteria: They (1) explored for two minutes; (2) indicated that they were done; (3) did not interact with the blocks or machine for 15 seconds (end time was coded as the last second they touched the block and/or the machine); (4) only played with the blocks off of the machine for 15 seconds (end time was coded as the last second when they removed the block(s) from the machine). Children who never put the blocks on the machine and only played with the blocks off of the machine, were coded as having an exploration time of one second. Children who did not interact with the blocks or machine at all even after additional

[1] We initially experimented with how children could indicate that they were done. Children were either instructed to verbally alert the experimenter ($n = 39$), given a fun distractor toy they could switch to ($n = 2$), or told that the experimenter would bring out new blocks when she returned and given a bell to ring when they were done exploring ($n = 36$; we kept this latter version for our preregistered replication of this pilot experiment). These approaches were evenly distributed across conditions.

[2] We were initially concerned children might feel they had to explore until the experimenter returned and so would continue due to normative pressures rather than interest. Thus, for a subset of children (n = 23), evenly distributed across conditions, the experimenter checked-in prior to 2 minutes if they stopped exploring for 5-10 seconds, i.e., before their exploration had otherwise ended. Most children, however, did stop exploring before 2 minutes had passed, so we decided to remove these check-ins for the remaining children in this experiment and the children in our preregistered replication.

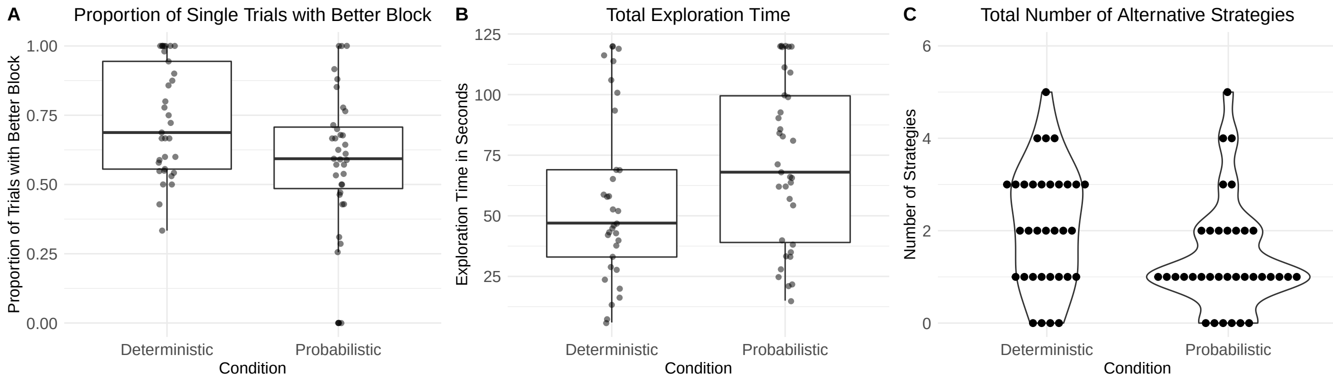

Figure 1: (A) Proportion of single trials children performed with the better block across conditions. (B) The amount of time children explored the blocks and/or the machine across conditions. Median and 1st and 3rd quartiles are displayed. (C) The total number of alternative strategies children performed across conditions. For all plots, dots are individual children.

encouragement from the experimenter were recorded as having an exploration time of zero seconds. These children did not indicate that they wanted to end the experiment; the experimenter also made clear through prompting that they could place blocks on the machine, so we were confident they understood the instructions and interpreted their lack of exploration as a choice not to explore, rather than confusion about the task. Nine children did not place a block on the machine, and this tendency did not differ by condition (Deterministic: $n = 6$; Probabilistic: $n = 3$; two-tailed Fisher's Exact test, $p = 0.481$). The end of exploration was determined offline according to the above criteria.

Third, we noticed that during children's exploration, they indeed performed alternative actions that were not demonstrated by the experimenter. We considered these different actions experimentation or alternative strategies children were employing to try to activate the machine. We identified seven different types of strategies: (1) exploring the machine alone (e.g. knocking on or poking it; flipping it over), (2) exploring the blocks alone (e.g., tapping the blocks together off of the machine or on the table), (3) flipping a block over to try a different side, (4) placing the block in a different location on the machine, (5) placing both blocks on the machine, (6) dropping the blocks onto the machine from above, and (7) applying force when placing the blocks on the machine (See Figure 2A). Children received a score of 1 for each strategy type they produced and a 0 otherwise (i.e., children could be coded as exhibiting 0 to 7 different strategies). Note that this is not a measure of how many times children produced a strategy but rather a count of how many different kinds of strategies children exhibited. Only children who interacted with the machine at some point were additionally coded for strategies (Deterministic condition: $n = 33$; Probabilistic condition: $n = 35$; same children as those included in analysis of the proportion of single trials with the better block).

Results

Consistent with previous work, in response to the explicit question about which block was better at activating the machine, the majority of children correctly selected the *better* block (two-tailed Binomial test, $p < 0.001$; Deterministic: $37/39$; Probabilistic: $33/38$; two-tailed Fisher's Exact test comparing across conditions: $p = 0.263$). This was also true when looking only at children who later placed a block on the machine during exploration (two-tailed Binomial test, $p < 0.001$; Deterministic: $31/33$; Probabilistic: $32/35$; two-tailed Fisher's Exact test across conditions: $p = 1$). These children were similarly more likely to first intervene with the better block, rather than the worse block (two-tailed Binomial test, $p < 0.001$; Deterministic: $22/33$; Probabilistic: $24/35$; two-tailed Fisher's exact test across conditions: $p = 1$). Intriguingly, these children were more likely to identify the better block as the better cause ($63/68$) than to select it first to place on the machine ($46/68$; two-tailed Fisher's Exact test, $p < 0.001$), suggesting they may have had motivations other than maximizing the probability of activating the machine when they first intervened.

To compare the proportion of single trials on which children used the better block, the total time children explored, and the total number of strategies children exhibited across conditions, we conducted three one-way ANCOVAs with condition as a factor and age in months as a covariate.

Children in the Deterministic condition intervened with the better block on a higher proportion of single trials than children in the Probabilistic condition ($M\pm$SE = 0.73 ± 0.035 v. 0.59 ± 0.041, respectively; $F(1,65) = 6.17$, $p = .016$), and this tendency to intervene with the better block did not differ by age ($F(1,65) = 0.23$, $p = .635$). Children in the Deterministic condition, however, explored for a shorter amount of time than children in the Probabilistic condition ($M\pm$SE = 48.60 ± 6.14 seconds v. 66.40 ± 6.14 seconds, respectively). This difference was significant ($F(1,74) = 4.37, p = .040$), and the length of time children explored did not differ by age ($F(1,74) = 1.26, p = .264$). If we only consider the children who placed a block on the machine, the difference in exploration time is trending (Deterministic: $M\pm$SE = 57.33 ± 6.12 seconds; Probabilistic: 71.89 ± 5.76 seconds; $F(1,65) = 3.07$, $p = .084$). (See Figure 1A-B.)

Most children employed at least one alternative strategy in their attempts to make the machine play music (Deterministic: 29/33; Probabilistic: 29/35), and this tendency did not differ across conditions (two-tailed Fisher's exact test, $p = .735$). However, children in the Deterministic condition employed more strategies overall than children in the Probabilistic condition ($M \pm$SE = 2.09 ± 0.23 v. 1.43 ± 0.20, respectively; $F(1,65) = 4.74, p = .033$); children's overall tendency to perform these alternative actions did not differ by age ($F(1,65) = 0.18, p = .670$). (See Figure 1C.)

Looking at the different strategies separately reveals that the modal strategy in the Deterministic condition was flipping the blocks over and in the Probabilistic condition, placing the blocks in different locations on the machine. Roughly the same number of children in both conditions changed the location of blocks (16 in Deterministic and 15 in Probabilistic) but about twice as many children flipped blocks over in the Deterministic than in the Probabilistic condition (21 v. 10 respectively), and this difference was significant (two-tailed Fisher's Exact test, $p = .007$). We are cautious to draw conclusions about the less common strategies since so few children exhibited them overall, but they do provide additional suggestive evidence that the conditions not only differed in the total number of strategies children exhibited but also in which strategies children employed. (See Figure 2B.)

Discussion

We provide evidence that children's exploratory behaviors can serve as a graded and detailed window into their causal reasoning. We find that presenting children with covariation information that deterministically or probabilistically supports a particular causal system leads them to explore this system differently, reflecting different inferences about causal strength and the uncertainty inherent in these inferences.

Children in both conditions drew rational inferences from the evidence they observed about the relative causal strength of the more effective (better) block compared to the less effective (worse) block. Children who previously observed deterministic evidence for the blocks' effectiveness attempted to activate the machine with the better block, rather than the worse block, more often than children who observed probabilistic evidence, suggesting a stronger inference that this block should work. Children's differential exploration across conditions suggests they did not simply draw a binary inference about which cause was better but rather were sensitive to the relative magnitude of causal strength.

Children's exploration also provided a richer picture of their causal inferences in this task than their causal judgments prior to the exploration phase. The causal judgments revealed that children had correctly inferred which block was more effective. If we had only considered this measure, however, we would not have seen that children were differentiating between the evidence presented in each condition and correctly retaining more uncertainty in the more ambiguous, probabilistic case. Similarly, if we had only looked at children's first intervention on the machine, we would have lacked the sensitivity to pick up on differences across conditions.

Interestingly, only about two-thirds (68%) of children across conditions first intervened with the better block, though over 90% explicitly identified this block as more effective. This is particularly surprising in the Deterministic condition; in prior work using similar or even weaker patterns of deterministic evidence, when explicitly asked to select a block to make the machine go, children overwhelmingly tended to intervene with the more effective block (e.g., Sobel et al., 2004; Sobel, Sommerville, Travers, Blumenthal, & Stoddard, 2009). Children's first intervention in our task suggests they were not simply motivated to generate the effect but rather to explore from the get go (e.g., perhaps some children wanted to understand why the worse block did not work or see if they could make it work). Direct questions likely place pressure on children to respond correctly, while self-directed exploratory play is more open-ended, removing such pressures and potentially revealing different behavior.

Prior work shows that children explore more when presented with events that violate their expectations. In many of these studies, including the present, children are not just faced with evidence that the causal system works differently than how they predicted but that it does not work at all. What inferences do children draw about the source of their own failed actions and how might their exploration reflect these inferences? Children in the Deterministic condition may have thought the problem lay with them since the evidence demonstrated by the experimenter strongly suggested that one block made the machine go and the other did not. These children tended to exhibit a wider variety of alternative actions, suggesting they may indeed have interpreted their failures to elicit music from the machine as the fault of their own actions rather than the causal system. Children in the Probabilistic condition, however, could explain away the lack of activation due to the system's stochasticity. Many did try at least one alternative strategy but overall did not experiment as much as children in the Deterministic condition, suggesting they did not necessarily think the problem lay with them but rather with the system itself. Indeed the one alternative strategy more children in the Probabilistic condition produced was exploring the machine on its own.

Though children in the Deterministic condition tried more strategies, they actually appeared to give up more quickly, exploring for a shorter amount of time overall than children in the Probabilistic condition. One possible explanation for this difference, consistent with why children in the Deterministic condition may have experimented more, is that the lack of machine activation led them to conclude more quickly that they were incapable of activating the machine either because they just could not figure out how or because it had stopped working. In contrast, children in the Probabilistic condition may have explored longer because they continued believing the machine might activate.

Taken together, these different aspects of children's exploration – their tendency to explore the better block, to experiment with different ways of activating the machine, and total

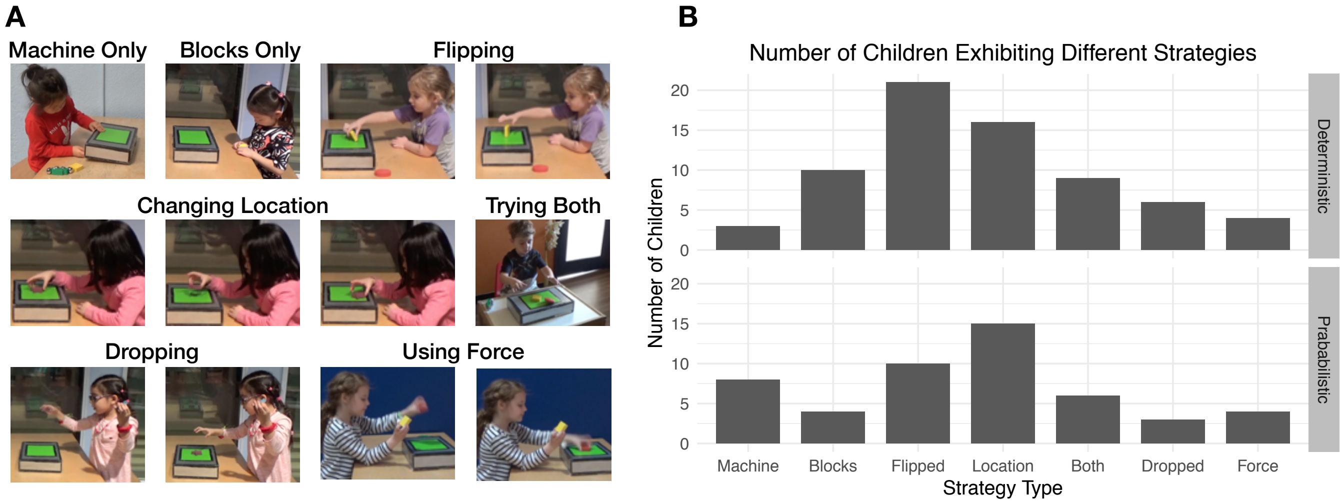

Figure 2: (A) Examples of alternative actions or strategies children performed to try to activate the machine. (B) The total number of children exhibiting each of the different strategies across conditions.

length of exploration – are suggestive of children integrating different sources of uncertainty. This includes uncertainty about the causal system itself (which block is better, and how much better) and uncertainty about the source of failure ("Is it me or the world?"; Gweon & Schulz, 2011).

In the deterministic case, children appear to persist in believing that the better block should work, so something must be wrong either with their own actions or with the blocks and the machine (e.g. it ran out of batteries). This behavior suggests continued certainty about how the causal system should work (even after receiving negative evidence about the previously better block, they still do not think it is likely that the previously worse block will work) but higher uncertainty about what has gone wrong (resulting in them trying more strategies to see if they can fix it).

In the probabilistic case, children appear less certain about the causal relationships. The better block continuing to not work is less surprising, because the system is stochastic. Children therefore demonstrate high uncertainty about the strength of the blocks and probability of activation (as evidenced by trying both blocks more evenly, rather than favoring the previously better one) but lower uncertainty about the source of failure (as evidenced by fewer strategies and overall more persistence).

Decades ago, Karmiloff-Smith and Inhelder (1974) argued that children's exploration is not just driven by their prior implicit theories but also by the evidence they generate as they explore, and that their failures to bring about an expected outcome would be interpreted as either relevant to their theory or to their action. Moving forward, looking at the time course of exploration (e.g., how early exploration differs from later exploration) could provide more compelling evidence for how children's inferences evolve as they accumulate evidence of the system's failure. For example, do the type of strategies children employ change over time? Prior work indicates children this age can design informative interventions to disambiguate causal evidence (Cook, Goodman, & Schulz, 2011). Are children's alternative strategies indeed targeting different hypotheses across conditions about why the machine is not working? In the present study, there are too few children who perform any particular strategy to probe these questions in more detail, but we are currently running a larger scale replication with over 100 children to address these questions (see `osf.io/sc54w` for preregistration).

Children's experimentation also raises interesting questions about conditions that lead to innovation. Along certain dimensions, and consistent with prior work (e.g., E. Bonawitz et al., 2011; Schulz, Hooppell, & Jenkins, 2008), deterministic evidence seemed to constrain children's exploration: they were less likely to explore the block another person had demonstrated to be less effective and explored for a shorter total amount of time than children in the Probabilistic condition. Along other dimensions however, they appeared to be more exploratory and innovative. They were more likely to experiment with how they placed blocks onto the machine. Children's ability to generate alternative means for achieving a goal and flexible problem solving in the face of failed actions is an interesting avenue for future work.

Exploration is a powerful, ecologically valid dependent measure that is more sensitive than binary questions and does not rely on children's language skills or explicit introspection. It does come with limitations; it is indirect and influenced by other factors besides children's beliefs. The use of open-ended, dynamic measures such as exploration, however, in conjunction with direct questions will allows us to paint a richer, more graded picture of children's inferences, as well as offer the potential of investigating how these inferences might change across time and affect behavior. Just as children harness the power of their exploratory play to learn about the world, we, as scientists, can harness this same play to learn more about what and how children learn.

Acknowledgments

We would like to thank Katrina Palad, Maureen Huang, Sandy Chao, and Aaron Philipp-Muller for their help with data collection and coding. We also thank the meta-reviewer and three reviewers of this paper for their thoughtful and constructive feedback. This research was supported by the Social Sciences and Humanities Research Council of Canada [435-2018-0890], and the Canada Foundation for Innovation.

References

Baldwin, D., Markman, E., & Melartin, R. (1993). Infants' ability to draw inferences about nonobvious object properties: evidence from exploratory play. *Child Development*, *64*(3), 711–728.

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

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.

Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology*, *52*(1), 9–18.

Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development*, *82*(6), 2053-2066.

Butler, L. P., & Markman, E. M. (2012). Preschoolers Use Intentional and Pedagogical Cues to Guide Inductive Inferences and Exploration. *Child Development*, *83*(4), 1416–1428.

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

Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives*, *7*(3), 160–165.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, *111*(1), 3–32.

Griffiths, T., & Tenenbaum, J. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

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

Gweon, H., & Schulz, L. (2011). 16-Month-Olds Rationally Infer Causes of Failed Actions. *Science*, *332*(6037), 1524–1524.

Hembacher, E., & Ghetti, S. (2014). Dont look at my answer: Subjective uncertainty underlies preschoolers exclusion of their least accurate memories. *Psychological Science*, *25*(9), 1768–1776.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*(3), 195–212.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*.

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*(1), 186–196.

Kushnir, T., Wellman, H., & Gelman, S. (2008). The role of preschoolers' social understanding in evaluating the informativeness of causal interventions. *Cognition*, *107*(3), 1084–1092.

Legare, C. H. (2011). Exploring Explanation: Explaining Inconsistent Evidence Informs Exploratory, Hypothesis-Testing Behavior in Young Children. *Child Development*, *83*(1), 173–185.

Schulz, L. (2012). The origins of inquiry: inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389.

Schulz, L., & Bonawitz, E. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050.

Schulz, L., Hooppell, C., & Jenkins, A. (2008). Judicious imitation: children differentially imitate deterministically and probabilistically effective actions. *Child Development*, *79*(2), 395–410.

Schulz, L., Standing, H., & Bonawitz, E. (2008). Word, thought, and deed: the role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology*, *44*(5), 1266–1276.

Sobel, D., Sommerville, J. A., Travers, L. V., Blumenthal, E. J., & Stoddard, E. (2009). The role of probability and intentionality in preschoolers' causal generalizations. *Journal of Cognition and Development*, *10*(4), 262-284.

Sobel, D., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303–333.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: an action measure. *Developmental Science*, *18*(1), 175–182.

Xu, F., & Kushnir, T. (2013). Infants Are Rational Constructivist Learners. *Current Directions in Psychological Science*, *22*(1), 28–32.