RESEARCH ARTICLE

Don't throw the associative baby out with the Bayesian bathwater: Children are more associative when reasoning retrospectively under information processing demands

Deon T. Benton[1] | David Kamper[2] | Rebecca M. Beaton[1] | David M. Sobel[3]

[1]Department of Psychology and Human Development, Vanderbilt University, Nashville, USA

[2]Department of Psychology, University of California, Los Angeles, USA

[3]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Los Angeles, USA

Correspondence
Deon T. Benton, Department of Psychology and Human Development, Vanderbilt University, Peabody College, 230 Appleton Place #552, Nashville, TN 37235, USA.
Email: deon.benton@vanderbilt.edu

This study was not preregistered. The study received approval by the study site's Institutional Review Board.

Abstract

Causal reasoning is a fundamental cognitive ability that enables individuals to learn about the complex interactions in the world around them. However, the mechanisms that underpin causal reasoning are not well understood. For example, it remains unresolved whether children's causal inferences are best explained by Bayesian inference or associative learning. The two experiments and computational models reported here were designed to examine whether 5- and 6-year-olds will retrospectively reevaluate objects—that is, adjust their beliefs about the causal status of some objects presented at an earlier point in time based on the observed causal status of other objects presented at a later point in time—when asked to reason about 3 and 4 objects and under varying degrees of information processing demands. Additionally, the experiments and models were designed to determine whether children's retrospective reevaluations were best explained by associative learning, Bayesian inference, or some combination of both. The results indicated that participants retrospectively reevaluated causal inferences under minimal information-processing demands (Experiment 1) but failed to do so under greater information processing demands (Experiment 2) and that their performance was better captured by an associative learning mechanism, with less support for descriptions that rely on Bayesian inference.

KEYWORDS
associative learning, Bayesian inference, causal reasoning, cognitive mechanisms, computational models

Research Highlights

- Five- and 6-year-old children engage in retrospective reevaluation under minimal information-processing demands (Experiment 1).
- Five- and 6-year-old children do not engage in retrospective reevaluation under more extensive information-processing demands (Experiment 2).
- Across both experiments, children's retrospective reevaluations were better explained by a simple associative learning model, with only minimal support for a simple Bayesian model.
- These data contribute to our understanding of the cognitive mechanisms by which children make causal judgements.

1 | INTRODUCTION

Few capacities are more important than the ability to reason and make inferences about causal relations. Causal reasoning enables human learners to make predictions and inferences (e.g., Bullock, et al., 1982; Shultz, 1982), to intervene on those relations to generate new effects (e.g., Butler et al., 2020; Schulz et al., 2007), and to reason about counterfactual claims—both about what might have been and how events could have turned out differently (e.g., Harris et al, 1996; Walker & Nyhout, 2020). These, and many other studies (e.g., Bonawitz & Lombrozo, 2012; Gopnik et al., 2001; Legare et al., 2010; Meltzoff et al., 2012; Walker & Gopnik, 2014), posit that young children have sophisticated causal reasoning capacities.

A fundamental question that underlies this research is *how* children make such inferences. One answer to this question is that children's causal inferences are best described by rational processes such as Bayesian inference. This process is thought to derive from more basic processes such as statistical learning that are present in early infancy (e.g., Gomez, 2002; Kirkham et al., 2002; Marcus et al., 1999; Saffran et al., 1996) and that with time enable infants to infer abstract patterns of coherent causal structure from probabilistic data (Gopnik & Wellman, 2012; Weisberg & Sobel, 2022). Although this view is often described as a computational level of analysis (cf. Marr, 1982), some advocates suggest that children use cognitive mechanisms that approximate or even represent Bayesian calculations (Bonawitz et al., 2014; Xu, 2019; see also Griffiths et al., 2015).

An alternative perspective is that associative learning alone is sufficient to describe children's causal inferences. On this view, children's causal knowledge reflects learned associations between causes and effects. Connectionist models—which learn largely via associative learning—have provided a proof of concept that causal learning can emerge from such associative processes (e.g., Benton et al., 2021; McClelland & Thompson, 2007). Additionally, comparative investigations between non-human animals and adults (e.g., Heyes, 2012) and studies of instrumental action and conditioning in human infants (e.g., Greco et al., 1990; Rovee-Collier, 1999) provide behavioral support for associative learning as a candidate mechanism for how children reason in the world.

One way to illustrate the tension between these hypotheses in development is through investigations of retrospective reasoning such as *backwards blocking* (Shanks, 1985). This is a form of reasoning that involves reevaluating the causal status of an ambiguous event based on learning more about the status of other unambiguous events (see also De Houwer et al, 2002; Kruschke & Blair, 2000; Larkin et al, 1998; Lovibond, 2003; Van Hamme & Wasserman, 1994, for other work on adults). One of the first studies to examine backwards blocking reasoning in children was carried out by Sobel et al. (2004). They introduced 3- and 4-year-olds to a machine called a "blicket detector" that lit up and played music when certain objects called "blickets" were placed on it (Gopnik & Sobel, 2000). Children were then shown that two novel objects, A and B, activated the machine when they were placed on it at the same time. Children were then shown that object A alone either

did or did not activate the machine. On both types of trials, children were then asked whether each object was a blicket. Children indicated that object A was a blicket when it activated the machine and that it was not a blicket when it did not activate the machine. Their judgments of object B also differed across these conditions. Children were more likely to conclude that object B was a blicket when object A failed to activate the machine than when A activated the machine. Using modified procedures, toddlers and even infants as young as 8 months of age showed a similar pattern of responses (Sobel & Kirkham, 2006).

These findings—and specifically the finding that children's causal inferences are sensitive to base rates (e.g., Sobel et al., 2004, Exp. 3)—have been interpreted as support for a Bayesian description of causal reasoning rather than as support for an associative learning mechanism. This is because some associative models (e.g., Rescorla & Wagner, 1972) predict that the strength between object B and the machine's activation is equivalent between the Backwards Blocking trial (where A is effective) and another trial in which A is not effective (labeled Indirect Screening-Off trials). Moreover, even a modified version of the Rescorla-Wagner model (e.g., Van Hamme & Wasserman, 1994) does not predict differences in such reasoning when the base rates of the causal effectiveness of an object is manipulated.

However, there are two facets of these data that warrant further consideration. First, McCormack et al. (2009) questioned what exactly was being reevaluated in a backwards blocking inference. They showed 4- and 5-year-olds that two objects (A and B) activated the machine together, and then that object A activated the machine alone. They compared children's causal status judgments for object B with a sequence in which a third object (C), unrelated to the compound set, activated the machine (i.e., AB+, C+). The 4-year-olds did not differ in their judgments (although 5-year-olds did—they were less likely to choose B than C). This control measure—which we adopt here—is a superior measure of assessing whether children reevaluate their causal judgments and specifically of examining whether children reevaluate the causal status of the object(s) shown independently, or the object only shown as part of the initial ambiguous data.

Second, although there are investigations suggesting Bayesian models are a better account for children's retrospective reasoning (e.g., Griffiths et al., 2011; Sobel et al., 2004), these investigations focus on a simplified case in which learners are asked to reason about exactly two candidate causes. Indeed, when three candidate causes are presented, some of children's inferences are better explained by Bayesian models, whereas other inferences are better explained by associative reasoning (Griffiths et al., 2011; Experiment 3). This suggests an intriguing possibility: As the number of candidate causes increases, children might fall back to simpler strategies such as associative learning from more rational reasoning strategies (akin to System 1/System 2 reasoning, Evans, 2003, 2011; Kahneman, 2011).

But how, exactly, is associative learning a simpler mechanism than Bayesian inference? The answer concerns the nature of the hypothesis spaces that underlie both models. Some associative models (including the one we instantiate here) posit a linear increase in the complexity

of the underlying hypothesis space based on the number of potential causes; that is, as the number of potential causes moves from 2 to n, the complexity of the hypotheses under consideration increases linearly from 2 to n, such that children must keep track of n associative values between each candidate cause and the effect. In contrast, in Bayesian models (as we will instantiate below) the underlying hypothesis space grows exponentially as the number of candidate causes increases. For example, if each object can either be a blicket or not and children are asked to reason about two potential blickets, then children would need to determine which of 2^2 or four causal hypotheses is correct. If, instead, children are asked to reason about just two more potential blickets for a total of four candidate blickets, then the underlying hypothesis space increases four-fold to 16 (i.e., 2^4) potential causal hypotheses. Thus, if children are sensitive to this increase in the size of the underlying hypothesis space and they possess limited information-processing abilities, then they might rely on simpler modes of processing such as associative learning than on more sophisticated forms of thinking that approximate normative Bayesian inference. The premise is that children have both associative and more rational causal reasoning mechanisms, but default to the former under more information processing demands.

There is now considerable evidence demonstrating that children do default to simpler modes of thinking when their information-processing abilities are taxed (e.g., Doebel & Zelazo, 2015; Frye et al., 1995; Zelazo et al., 1996; Zelazo et al., 2003). For example, recently Kenderla and Kibbe (2023) demonstrated that 8- and 10-year-old children showed decreased reliance on working memory and greater dependence on manual exploration during a challenging virtual memory game. The goal of this game was to find three cards with shared and differing features. Given that children were not required to maintain information in memory when manually exploring, manual exploration ostensibly was a less cognitively effortful strategy than one that required an already resource-limited system such as working memory. Similarly, Richland et al. (2006) found that 3- and 4-year-old children made more featural and relational errors when asked to reason about multiple relations or when the task included a salient distractor than when asked to reason about a single relation without a distractor.

Even in infancy there is development from more associative to more rational inferences. Using an anticipatory eye-gaze measure, Sobel and Kirkham (2007) found that 8-month-olds exhibited backwards blocking inferences similar to preschoolers, but 5-month-olds' inferences were more associative in nature. When infants make judgments about the reliability of others' information, their decision-making seems to be best explained by associative processing (Sobel et al., 2020; Tummeltshammer et al., 2014). As children enter the preschool years, those judgments become more rational in nature (Sobel & Kushnir, 2013), although occasionally, they will default to associative forms of processing, particularly under information processing demands (e.g., Hermes et al., 2018; Luchkina et al., 2020). Further, on other kinds of retrospective causal reasoning tasks, as the information demands of the procedure increase, only older children between 3 and 7 years of age succeed (e.g., Erb & Sobel, 2014; Fernbach et al., 2012;

Sobel et al., 2017). Finally, beyond explicit causal reasoning tasks, preschoolers' performance on theory-of-mind and social-problem-solving tasks was adversely affected when they first completed tasks that taxed their information-processing abilities compared to when such capacities were not taxed (Caporaso & Marcovitch, 2021; Powell & Carey, 2017; Steinbeis, 2018). Considered together, these studies indicate that children use different reasoning processes under different information-processing demands; the higher those demands, the simpler the process (e.g., Cohen, 1988).

In the present study, we considered how children made retrospective inferences when first shown ambiguous data (i.e., three objects together produce an effect), followed by further evidence involving one of those objects (Experiment 1) or two of those objects (Experiment 2). In both cases, the logic of our design followed McCormack et al. (2009), in which we contrasted these retrospective inferences with control trials in which children saw the same initial ambiguous data, and then unrelated objects that had similar efficacy. The question across both experiments was whether children show qualitative evidence for Bayesian inference and associative learning with an edge towards associative learning. After presenting these behavioral data across two experiments, we present a pair of computational models to determine to what extent children's performance in Experiments 1 and 2 was better explained by Bayesian inference, associative learning, or both. The value of computational modeling here is that it can help to elucidate the cognitive mechanism or mechanisms by which children engage in retrospective reevaluation in ways that the experiments alone cannot. Specifically, by implementing as computer simulations theories about how children engage in retrospective reevaluation, it is possible to determine the theory—and by extension, the mechanism—that better accounts for the behavioral data.

2 | EXPERIMENT 1

Five- and 6-year-olds observed three objects (A, B, and C) together cause a machine to activate. Then they observed that object A either caused (Backwards Blocking trials) or failed to cause (Indirect Screening-Off trials) the machine to activate by itself. They were then asked whether each object individually caused the machine to activate. These experimental trials were compared to control trials in which children observed three different objects (A', B' and C') activate the machine together, followed by a fourth object (D), which either did (Backwards Blocking control) or did not (Indirect Screening-Off control) make the machine activate.

In these trials, a retrospective reevaluative causal inference is defined as participants treating the objects in the control trials that go on the machine together (A', B', and C') differently from the objects in the experimental trials that initially went on the machine together in the first demonstration, but whose individual efficacy was not revealed (i.e., B and C). In the Backwards Blocking trials, participants were said to engage in this form of reasoning if they were more likely to choose objects A, B, and C (i.e., the objects that were not shown on the machine

by themselves) in the control trials than objects B and C in the experimental trials (i.e., the objects that were not shown on the machine by themselves). The reason for this is straightforward: Given that A was shown initially in combination with B and C, observing subsequently that A causes the machine to activate by itself should affect participants' inferences about B and C. However, because object D was never shown in combination with A'–C', D's causal status should have no bearing on participants' treatment of objects A'–C'. In the indirect screening-off trials, participants were said to engage in this form of reasoning if they were more likely to choose objects B and C in the experimental trial than objects A, B, and C in the control trial. The rationale for why these ratings should differ is identical to that for the Backwards Blocking condition—because A was shown in combination with objects B and C, A's, but not D's, causal status should affect how participants rate the objects that never participated on the machine alone. Because McCormack et al. (2009) found that 5 and 6-year-olds made such retrospective inferences about two candidate causes, we have decided to test children of the same age.

2.1 | Method

2.1.1 | Participants

Participants were 32 five-year-olds (16 boys and 16 girls; $M = 64.81$ months, range = 60–71 months, standard deviation [SD] = 3.48) and 31 six-year-olds (17 boys and 15 girls; $M = 77.81$ months, range = 72–83 months, SD = 3.78). Sample size was determined based on previous studies on backwards blocking reasoning in children (e.g., Griffiths et al., 2011; Sobel et al., 2004). Two children were excluded from analysis for failing to participate ($N = 1$) or missing video (which made coding their responses impossible) ($N = 1$). We did not collect demographic information about the sample, but the demographic information about sample of children collected by the laboratory during this time was as follows: 82% White/Caucasian, 3% Black/African American, 4% Asian/Asian American, 0.5% Native American, and 11% of Mixed Descent. Sixteen percent identified as Hispanic/Latinx. Similarly, the overall household income level of families tested in the lab during this time was as follows: Less than 30K: 7%, 30–50K: 7%, 50–70K: 14%, 70–90K: 9%, 90–120K: 25%, Over 120K: 38%. The median income for the population as measured by the 2020 Census was ~$74K.

2.1.2 | Materials

The "device" used in the current study was a computer-animated version of the blicket detector (Gopnik & Sobel, 2000). The device was a white rectangle with a black border that measured 5.99 cm × 23.47 cm and that was presented on a computer screen. If the device was "on," the white region of the rectangle turned blue when objects touched it. If the device was "off," the white region remained white. A maximum of four differently colored circles were shown on the screen. Each cir-

cle measured 2.67 cm × 2.67 cm (see Figure 1 below). The machine was designed such that it activated immediately when the bottommost edge of a circle—predetermined to be a blicket—contacted it. At the start of any given trial, three or four equally spaced circles appeared above the machine. Finally, the videos contained a built-in script, which experimenters, but not the study participants, read. All video events were created in Microsoft PowerPoint.

2.1.3 | Procedure

All study procedures were reviewed and approved by the University's Institutional Review Board, and parental informed consent and child assent was obtained before each experimental session. Participants were tested in a quiet room in a local children's museum. At the beginning of the experiment, all participants were shown a pretraining video. The text, "We're going to play a game with my machine. This is a very special machine. It's my blicket machine. Blickets make the machine go. So, let's find all the blickets" appeared on the screen and was read to the participants by the experimenter. The video consisted of a rectangular base (i.e., the previously mentioned "blicket detector") and two shapes (i.e., a gray triangle and a gray pentagon). Crucially, these shapes were unrelated to the circles used during the experimental portion of the experiment. The pretraining phase began with the triangle (object A) and pentagon (object B) next to each other above the machine. Object A then descended until it contacted the machine, which immediately activated (i.e., the white region changed from white to blue). Object A then returned to its starting position above the machine. Object B then descended until it contacted and failed to activate the machine. Object B then returned to its starting position. Finally, both objects descended until they contacted and activated the machine. Participants were then asked whether each object was a blicket. This event ensured that participants understood the task and recognized that individual objects could activate the machine and that it activated if at least one effective object was placed on it.

Following this pretraining phase, participants were given four trials. Half the participants received two backwards blocking experimental trials and two backwards blocking control trials. The other half received two indirect screening-off experimental trials and two indirect screening-off control trials. The order of these trials within each condition was counterbalanced using a Latin square design. Different colored objects were used across all trials to prevent carryover effects. A schematic of this procedure is shown in Figure 1. Finally, all study responses were coded offline after each study session. Although study responses were coded offline, an experimenter was present throughout an entire study session.

2.1.4 | Backwards blocking experimental and control trials

The two backwards blocking experimental trials began with three differently colored objects, which were located above the machine. The
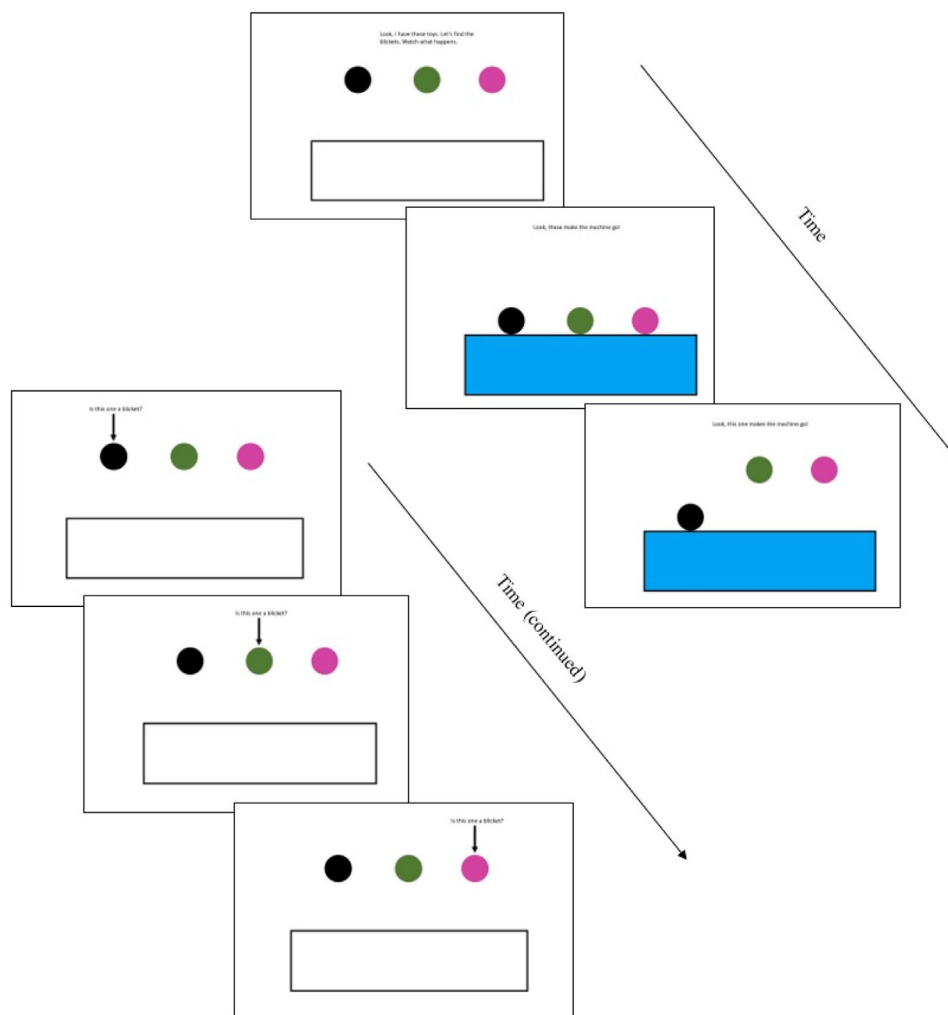
FIGURE 1 Schematic of a Backwards Blocking experimental trial. The upper-right portion of the figure shows the backwards blocking event as it unfolded across time. The lower-left portion of the figure shows the three objects and the text, "Is this one a blicket?" above each object across time.

text, "Look, I have these three toys. Let's find the blickets. Watch what happens" appeared above the objects. All three objects (i.e., objects A, B, and C) then descended until they contacted and activated the machine. At this point, the text, "Look, these also make the machine go!" appeared above the objects. The objects then returned to their starting positions.

The left- or right-most (counterbalanced) object (which here we will refer to as object A) then descended until it contacted and immediately activated the machine. The text, "Look, this one makes the machine go!" then appeared above the objects. This object then returned to its starting position. Children were then asked whether each object was a blicket. Specifically, the text, "Is this one a blicket?" with a downward-facing arrow then appeared above each object, and participants were asked to indicate whether each object was a blicket. Children received two of these trials, which were identical except for the color of the objects.

The two backwards blocking control trials began with four differently colored objects (i.e., objects A, B, C, and D), which were located above the machine. Objects A, B, and C then descended until they contacted and activated the machine; object D remained in place while objects A–C descended onto the machine. Object D then descended by itself until it contacted and activated the machine. The left-right position of object D was counterbalanced. Children were then asked whether each object was a blicket. Children once again received two trials, which were identical except for the color of the objects.

2.1.5 | Indirect screening-off experimental and control trials

The procedures for the indirect screening-off experimental and control conditions were identical to the backwards blocking trials except that object A (experimental trials) and D (control trials) failed to activate the machine. Table 1 below illustrates the key trial structures for the Backwards Blocking and Indirect Screening-Off conditions in Experiment 1.

TABLE 1 Schematic of the task structure for the backwards blocking and indirect screening-off experimental and control trials.

Condition	First learning phase	Second learning phase
Backwards blocking (experimental)	ABC+	A+
Backwards blocking (control)	ABC+	D+
Indirect screening-off (experimental)	ABC+	A-
Indirect screening-off (control)	ABC+	D-

2.2 | Results

Figure 2 shows participants' responses to "Is this a blicket?" for each object. Participants' yes/no responses were treated as the primary binary dependent measure. All analyses were conducted with the lme4 package in R (Bates et al., 2015). Deidentified data for all experiments, along with all analysis code, is available on OSF (https://osf.io/n6mvq/?view_only=a6b8231a6b9743c7bfe896ba1eab 58f3). Data were entered into a five-way mixed-effects logistic regression with Age as a continuous fixed effect, Condition (Backwards Blocking vs. Indirect Screening-Off) as the between-participants fixed effect, Trial Type (Experimental vs. Control), Objects (A vs. B vs. C vs. D), and Trial Number (Trial 1 vs. Trial 2) as the within-participants fixed effects, and participant as the random effect. This analysis yielded several experimental-effects and two-way interactions, which were qualified by a single three-way interaction among Condition, Trial Type, and Object, $\chi^2(2) = 64.85$, $p < 0.001$.

To unpack the nature of the interaction among Condition, Trial Type, and Object, we ran separate two-way mixed-effects logistic regressions separately for the Backwards Blocking and Indirect Screening-Off conditions, with Trial Type (Experimental vs. Control) and Objects (A vs. B vs. C vs. D) as the within-participants fixed effects and participant as the random effect. This analysis revealed a main effect of Trial Type, $\chi^2(1) = 9.62$, $p = 0.002$ and an interaction between Trial Type and Objects, $\chi^2(2) = 16.38$, $p < 0.001$. To explore this interaction, we constructed a set of one-way mixed-effects logistic regressions for the experimental and control trials within the Backwards Blocking condition. The Objects factor was treated as the sole within-participants fixed effect in these follow-up analyses. Participants were once again treated as a random effect to control for the within-participant variance from multiple responses. The one-way mixed-effects logistic regression for the control trials within the Backwards Blocking condition did not reveal a significant effect of Objects, $\chi^2(3) = 1.33$, $p = 0.72$. This means that participants treated the objects similarly in the control trials of the Backwards Blocking condition. In contrast, the second one-way mixed-effects logistic regression for the experimental trials within the Backwards Blocking condition revealed a significant experimental effect of Objects, $\chi^2(2) = 19.29$, $p < 0.001$. This experimental effect reflected the fact that participants judged object A as a blicket more often than object B, odds ratio = 204.79, 95% CI [33.96, 4609.11], $p < 0.001$, and object C, odds ratio = 129.67, 95% CI [18.75, 2824.63],

$p < 0.001$. However, participants treated objects B and C equivalently, odds ratio = 1.58, 95% CI [0.62, 4.19], $p = .34$.

The two-way mixed-effects logistic regressions for the Indirect Screening-Off condition also revealed a main effect of Trial Type, $\chi^2(1) = 26.91$, $p < 0.001$, a main effect of Objects, $\chi^2(3) = 67.32$, $p < 0.001$, and an interaction between Trial Type and Objects, $\chi^2(2) = 19.59$, $p < 0.001$. To explore this interaction, we constructed a set of one-way mixed-effects regressions for the experimental and control trials within the Indirect Screening-Off condition. The two one-way mixed-effects regressions for the experimental and control trials revealed a significant experimental effect of Objects, both χ^2-values > 36.78, both p-values < 0.001. In the experimental trials, participants judged object A as a blicket less often than any of the other objects, all odds ratios < 0.07, all p-values < 0.001. Likewise, in the control trial, participants considered object D to be less likely to be a blicket than any of the other objects, all odds ratios < 0.06, all p-values < 0.001. No other differences reached statistical significance.

2.2.1 | Evidence of retrospective reasoning

To examine whether participants engaged in backwards blocking reasoning, data for the experimental and control trials within the Backwards Blocking condition were entered into a two-way mixed-effects logistic regression with Trial Type and Object as the within-participants fixed effects and participants as the random effect. This analysis revealed only a main effect of Trial Type, $\chi^2(1) = 17.72$, $p < 0.001$. This result indicated that participants did engage in backwards blocking reasoning. In particular, a follow-up, one-way mixed-effects logistic regression showed that participants were less likely to consider the objects whose efficacy was not shown individually in the experimental trial (i.e., objects B and C) to be blickets than the objects that were placed on the machine together in the control trial (i.e., objects A', B', and C'), odds ratio = 0.19, 95% CI [0.09, 0.78], $p < 0.001$.

We also ran the same analysis as above, but this time for the Indirect Screening-Off condition. Although this analysis also revealed a main effect of Trial Type, $\chi^2(1) = 4.39$, $p = 0.04$, a follow-up, one-way mixed-effects logistic regression indicated that participants' treated the objects that did not participate on the machine in the experimental trials (i.e., objects B and C) and the objects that did not participate on the machine in the control trials (i.e., objects A, B, and C) equivalently, odds ratio = 0.50, 95% CI [0.25, 1.01], $p = 0.052$.

2.3 | Discussion

In the experimental trials of Experiment 1, children were shown three objects that together activated a machine and then shown that one of those objects was or was not effective on its own. When that object was effective, children reevaluated the efficacy of the other two objects: They stated that they were less likely to be effective than objects in a control condition in which a fourth, unrelated object was effective. When that object was not effective, children did not retrospectively
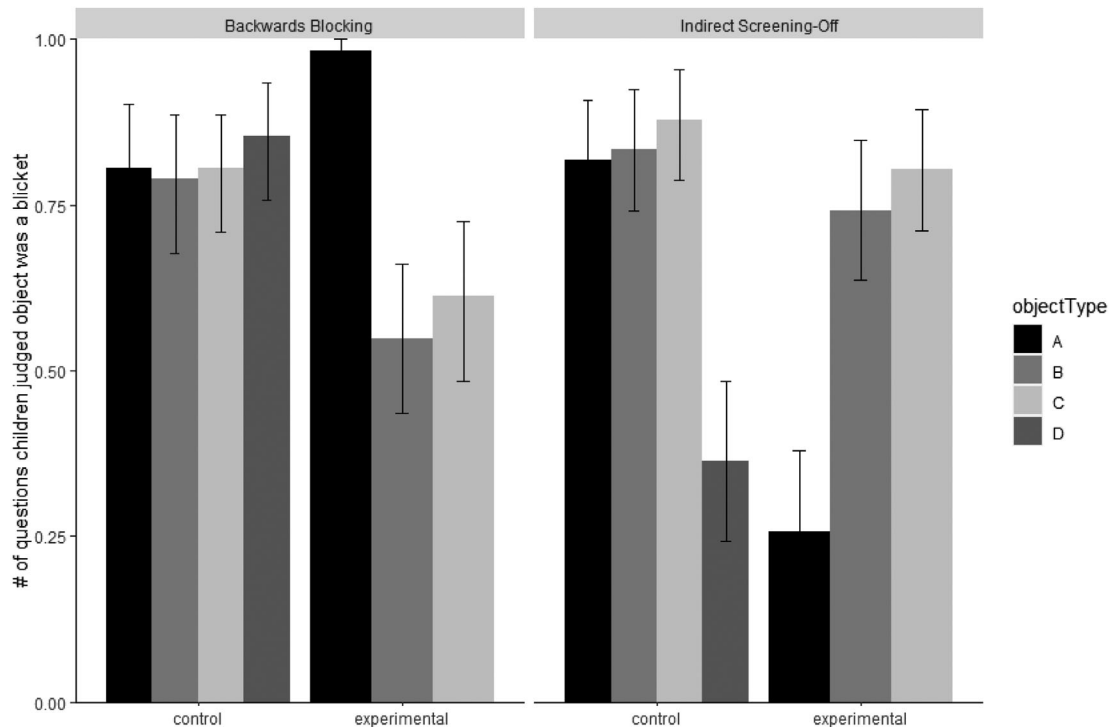
FIGURE 2 Participants' responses expressed as percentages to whether each object was a blicket across the conditions and trial types. Bars show standard error.

reevaluate the efficacy of the other objects and judged the objects equivalently across both conditions.

Before discussing possible cognitive mechanisms that might underlie these data, we wanted to consider a second, related type of retrospective inference. In Experiment 1, following the ABC+ event, participants were either shown an A+ event (in the Backwards Blocking condition) or an A- event (in the Indirect Screening-Off condition). Experiment 2 was similar to Experiment 1 except for what children observed following the ABC+ events (e.g., McCormack et al., 2009). In the experimental trial in the Backwards Blocking condition, they observed an AB+ event during the second learning phase; in the control trial in the same condition, children observed a DE+ event during the second learning phase. Children in the Indirect Screening-Off condition were shown the same series of events except that the machine did not activate. If children's ability to engage in various forms of retrospective reevaluation is related to their information processing, in Experiment 2 children should be less likely to engage in retrospective reevaluation than those in Experiment 1.

3 | EXPERIMENT 2

Experiment 2 was analogous to Experiment 1 except for the number of objects that were placed on the machine during the second part of the experimental and control trials. In the experimental trial, children were shown that three objects activated the machine together, and then that two of those three objects either did or did not activate the machine when they were placed on it together. These data were compared with

a control trial in which three different objects activated the machine, and then two additional novel objects either did or did not activate the machine in tandem.

3.1 | Method

3.1.1 | Participants

Participants were 32 five-year-olds (18 boys and 14 girls; $M = 65.31$ months, range = 60–75 months, SD = 3.65) and 32 six-year-olds (10 boys and 22 girls; $M = 76.56$ months, range = 65–83 months, SD = 4.33). Participants were recruited in the same manner as Experiment 1. Participants were 12% Asian/Asian American, 9% Black/African American, 10% Hispanic, and 69% White/Caucasian, but no other specific demographic data were collected (see Experiment 1 for overall demographic data from the laboratory).

3.1.2 | Materials and procedure

The materials and procedure for Experiment 2 were identical to that for Experiment 1 with the following exceptions: During the experimental trials in the Backwards Blocking condition following an event in which objects A, B, and C together activated the machine, two objects, A and B, descended onto and subsequently caused the machine to activate. Likewise, during the control trials in the same condition which consisted of five objects (i.e., objects A–E), objects D and E descended

TABLE 2 Schematic of the task structure for the backwards blocking and indirect screening-off experimental and control trials.

Condition	First learning phase	Second learning phase
Backwards blocking (experimental)	ABC+	AB+
Backwards blocking (control)	ABC+	DE+
Indirect screening-off (experimental)	ABC+	AB-
Indirect screening-off (control)	ABC+	DE-
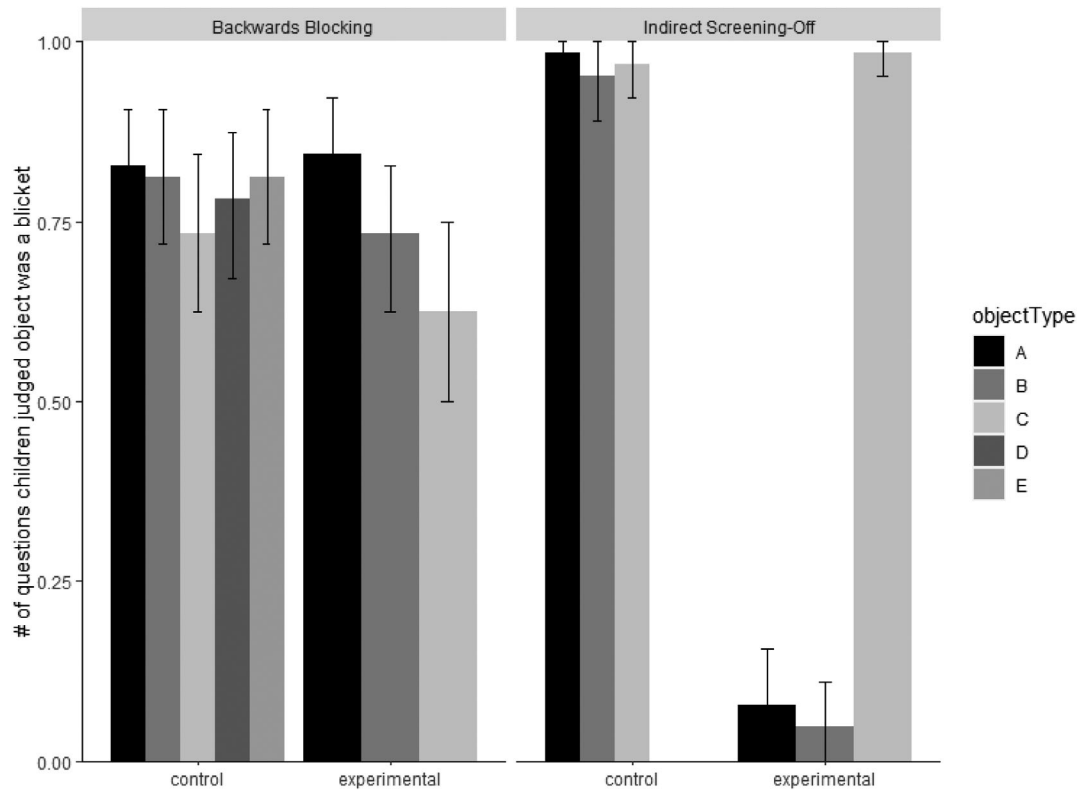
FIGURE 3 The participants' responses expressed as percentages to whether each object was a blicket across the conditions and trial types. Bars show standard error.

onto and subsequently caused the machine to activate. Objects D and E did not descend onto the machine during the initial event in which A, B, and C activated the machine and in this way were unrelated to objects A, B, and C. The experimental and control trials in the Indirect Screening-Off condition were identical to the backwards blocking trials except that the machine neither activated when objects A and B descended onto it in the experimental trial nor when objects D and E descended onto it during the control trial. The left- and right-most positions of objects A and B during the experimental trial and objects D and E during the control trial were counterbalanced. Table 2 shows the structure of the events used in Experiment 2.

3.2 | Results

Figure 3 shows participants' responses to "Is this a blicket?" for each object. Data were entered into a five-way mixed-effects logistic regression model with Age as a continuous fixed effect, Condition (Backwards

Blocking vs. Indirect Screening-Off) as the between-participants fixed effect, Trial Type (Experimental vs. Control), Objects (A vs. B vs. C vs. D), and Trial Number (Trial 1 vs. Trial 2) as the within-participants fixed effects, and participant as the random effect. This analysis only yielded a main effect of Trial Type, $\chi^2(1) = 14.33$, $p = 0.04$. This reflected that fact that across the Backwards Blocking and Indirect Screening-Off, participants were less likely to treat the objects in the experimental trials as blickets than objects in the control trials, odds ratio = 0.45, 95% CI [0.33, 0.62], $p < .001$.

3.2.1 | Evidence of retrospective reasoning

We next examined whether participants engaged in retrospective reasoning using the operationalization of it from Experiment 1. Data were entered into a two-way mixed-effects logistic regression with Trial Type and Object as the within-participants fixed effects and participants as the random effect. This analysis did not reveal any main effects
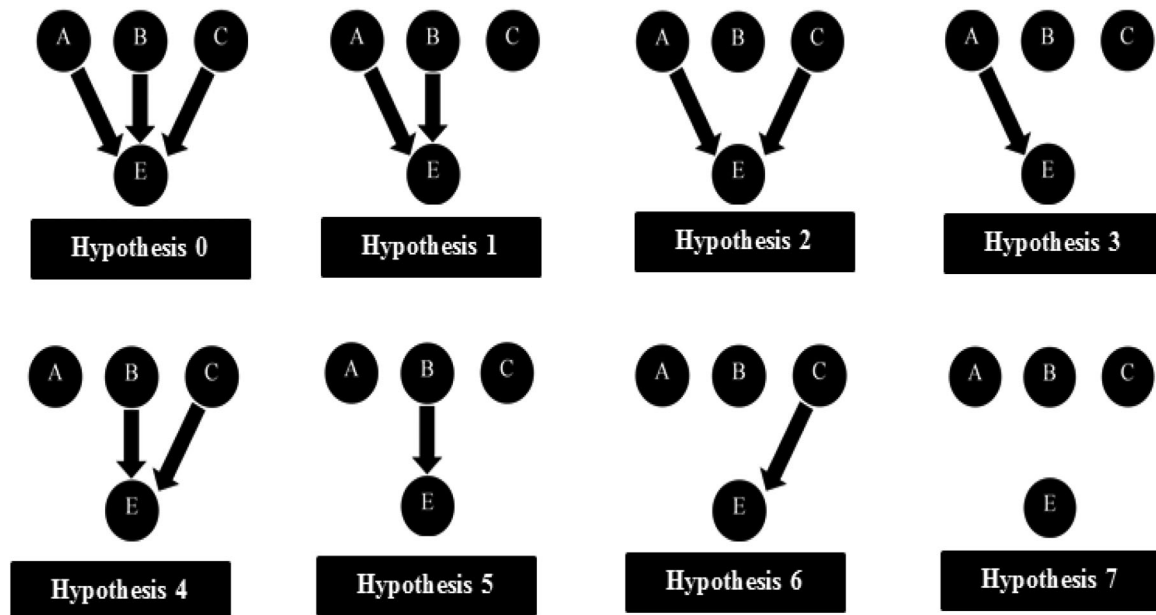
FIGURE 4 The eight different causal hypotheses indicating the possible causal relations for a causal event that involves three objects and one blicket detector. A, B, and C correspond to the three objects that were used on the machine and E indicates the activation of the machine.

or interactions, all χ^2-values < 1.91, all p-values > 0.18. The same picture emerged for the indirect screening-off condition—this analysis did not reveal any main effects or interactions, all χ^2-values < 1.79, all p-values > 0.41. Thus, unlike Experiment 1, there was no evidence that participants engaged in any form of retrospective reevaluation. This finding is likely the result of the increased demand on children's information processing abilities: Children were not only required to reason about 3 and 4 objects (as in Experiment 1), but they were also required to reason about 2 rather than 1 object during the second learning phase in the Backwards Blocking and Indirect Screening-Off conditions.

3.3 | Discussion

Unlike Experiment 1, in Experiment 2 there was no evidence that children engaged in retrospective reasoning. Specifically, children treated the objects equivalently between the experimental and control trials. In addition, across both experiments there was no evidence that retrospective reevaluation undergoes developmental change between 5 and 6 years of age. We return to this issue in the General Discussion. In the next section, we present fits from two computational models to determine whether an associative mechanism, a Bayesian mechanism, or some combination of both best captures children's judgements across Experiments 1 and 2.

4 | COMPUTATIONAL MODELS

We fit two computational models to the behavioral data. The first was a model based on Bayesian inference. This model was described initially by Sobel et al. (2004) and in more detail in Griffiths et al. (2011). The

second was a simple connectionist model, trained with the Delta Rule (Widrow & Hoff, 1960).

4.1 | Bayesian model

The Bayesian model we use here has been described previously (Griffiths & Tenenbaum, 2005; Griffiths et al., 2011; Tenenbaum & Griffiths, 2001). We refer the reader to these citations for more of a technical description. Here, we describe the basics of the model. Bayesian reasoning assumes the learner has a set of hypotheses H. Each hypothesis $h \in H$ is assigned a *prior probability*, $p(h)$, which indicates the initial belief that a learner has in a particular hypothesis prior to seeing data. After the learner observes data, d, the learner computes a posterior probability, $p(h|d)$—an updated belief about each hypothesis given the data. This is done using Bayes' rule, shown in Equation (1):

$$p(h|d) = \frac{p(d|h)\,p(h)}{\sum_{h' \in H} p(d|h')\,p(h')} \tag{1}$$

In this formula, $p(d|h)$ is the probability that the data d will be observed under a particular hypothesis h. This value is also known as the *likelihood* of the data.

Forming the initial hypothesis space for this model assumes that there is a set of objects O and a detector d, such that any object $o \in O$ can potentially cause d to activate. Given that participants are shown that the machine activates when objects with the label "blicket" are placed on its surface, a hypothesis h corresponds to a causal structure that posits whether individual objects have the causal effectiveness to activate the detector—that is, an arrow between a node representing an object and a node representing the machine's activation (see Griffiths & Tenenbaum, 2005, for more computational details; see Figure 4

TABLE 3 Model fit indices for the various models and instantiations for the data overall and the data for the backwards blocking, indirect screening-off, experimental, and control trials in Experiments 1 and 2 data.

(A) Model fit to the data overall

Experiment 1				Experiment 2			
Connectionist[‡]		Bayesian model		Connectionist[‡]		Bayesian model	
RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
0.15	0.11	0.17	0.17	0.13	0.11	0.16	0.13

(B) Model fit to the backwards blocking data only

Experiment 1				Experiment 2			
Connectionist[‡]		Bayesian model		Connectionist[‡]		Bayesian model	
RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
0.19	0.16	0.20	0.18	0.13	0.11	0.15	0.14

(C) Model fit to the indirect screening-off data only

Experiment 1				Experiment 2			
Connectionist[‡]		Bayesian model		Connectionist		Bayesian model[‡]	
RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
0.08	0.07	0.18	0.16	0.11	0.11	0.12	0.03

(D) Model fit to the experimental trials only

Experiment 1				Experiment 2			
Connectionist		Bayesian model		Connectionist		Bayesian model[‡]	
RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
0.19	0.16	0.19	0.16	0.16	0.14	0.14	0.12

(E) Model fit to the control trials only

Experiment 1				Experiment 2			
Connectionist[‡]		Bayesian model		Connectionist[‡]		Bayesian model	
RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
0.10	0.08	0.20	0.17	0.11	0.09	0.17	0.17

Abbreviations: MAE, mean absolute error; RMSE, root mean square error.
[‡]Corresponds to the better fitting overall model based on average RMSE and MAE.

for the hypothesis space). Griffiths et al. (2011) describe the formal parameterization of this hypothesis space and model that results in the hypothesis space shown in Figure 3, in which nodes A, B, and C represent objects A, B, and C each being placed on the machine respectively, and node E represents the "effect"—the machine activating.

To instantiate the model, each hypothesis is given a prior probability $p(h)$, which is a function of the child's belief about how likely any object is to be a blicket (i.e., the base rate of blickets), ρ. This prior corresponds to the number of blickets posited by the hypothesis. For example, in the figure, Hypothesis 0 posits 3 blickets, so its $p(h) = \rho^3$. Hypotheses 1, 2, and 4 posit exactly 2 blickets, so their $p(h) = \rho^2(1-\rho)$. Hypotheses 3, 5, and 6 each posit 1, which makes their $p(h) = \rho(1-\rho)^2$. Finally, Hypothesis 7 posits no blickets, which makes its $p(h) = (1-\rho)^3$.

For the purposes of this demonstration, we will assume that the model itself assumes that objects with causal efficacy will act deterministically on detectors.[1] As a result, the likelihood of each hypothesis is equal to 1 if that hypothesis could produce the data and 0 if not. This allows each model to be updated based on Bayes' rule, given the

data. The way the model then determines the probability that an object is a blicket is based on the posterior probability of the models in the hypothesis space. The probability that an object o is a blicket is the probability that it activates the machine, given the data d (i.e., $p(o \to E \mid d)$). This can be calculated by Equation (2)

$$p(o \to E \mid d) = \sum_{h \in H} p(o \to E \mid h)p(h|d) \qquad (2)$$

where $p(o \to E \mid h)$ is 1 if there is an edge between that object and the detector in that particular hypothesis, and 0 otherwise.

Crucially, because the predictions of this (or any) Bayesian model will depend on the prior probability that any given object is a blicket, we fit a Bayesian model with the following prior probabilities: 0.5, 0.65, 0.8, 0.95, and 1. We considered a range of prior probabilities because it was unclear what participants' baseline assumptions were about the prior probability of blickets in the absence of explicit manipulations to those probabilities. Thus, by deriving the model's predictions for various prior
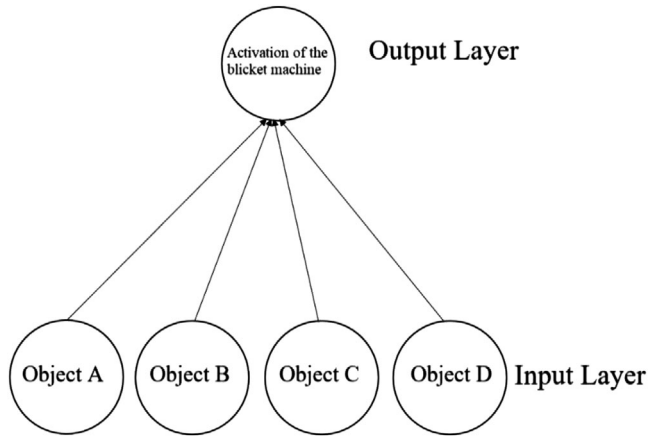
FIGURE 5 The connectionist model used to simulate Experiment 1.

probabilities, it was possible to compare the model's predictions for the different probabilities to children's actual treatment of the objects. The best quantitative fit of this model to the data in Experiments 1 and 2 is shown below in Table 3.

4.2 | Connectionist model

We also built a set of two-layer connectionist models. One of these models corresponded to Experiment 1 and the other corresponded to Experiment 2. The model architecture for the Experiment 1 simulations is shown in Figure 5. The rationale for building only a two-layer model was to explore whether a simple learning model trained with the Delta Rule (Kruschke, 1992; Widrow & Hoff, 1960)—which is formally equivalent to the traditional Rescorla-Wagner model (Danks, 2003; Gluck & Bower, 1988)—could be used to explain these data. Similar to children, we trained 16 models (i.e., "participants") per condition for both experiments (i.e., 32 total model runs for Experiment 1 and 32 total model runs for Experiment 2), and like the children, each model received two trials. Each new participant began with a fresh set of small random weights (sampled uniformly between ± 0.1). Finally, data were aggregated over the responses of each model to allow us to fit the model's responses to participants' count data (as shown in Figures 2 and 3).

The input layer for the model consisted of four units for Experiment 1 (corresponding to the four objects) and five units for Experiment 2 (corresponding to the five objects), and the output layer consisted of a single unit for the simulation of both experiments (corresponding to the activation of the machine). When object was placed on the machine, the activation value of its corresponding input unit was set to a value of 1 (and 0 otherwise). The input units could not take on any other values beside 0 or 1. If an object that was a blicket was placed on the machine, then the model was trained to turn on the single output unit (i.e., to produce an activation of 1).

All simulations used a learning rate of 0.05 but no momentum. Model weights were initialized to small random values (distribution range = ± 0.1), and the output units used sigmoidal or logistic activation

functions. The activation of the single output unit was interpreted as the model's confidence (or prediction) that a given object was a blicket and could range between 0 and 1 due to the sigmoid activation function (unlike the input units, whose input values were "hard clamped" or fixed).

Turning on the first three input units simulated placing objects A, B, and C on the machine, and training the model to turn on the single output unit corresponded to teaching the model that the machine activated when objects A-C were placed on it. During the subsequent A+ trials in Experiment 1 or the AB+ trials in Experiment 2, only the first input unit (for the simulation of Experiment 1) or the first and second input units (for the simulation of Experiment 2) were turned on, but again the model's task was to activate the single output unit. The control trials in the Backwards Blocking condition were identical to the experimental trials except that the fourth input unit (corresponding to object D in Experiment 1) or the fourth and fifth input units (corresponding to objects D and E in Experiment 2) were turned on following the ABC+ trial. The experimental and control trials in the Indirect Screening-Off condition were identical to the backwards blocking experimental and control trials except that the model was trained to turn off the single output unit (i.e., to produce an output activation of 0) during the experimental and controls for the simulations of Experiments 1 and 2. Each phase of the simulations—which were shown twice to be consistent with the behavioral study—lasted anywhere between 200 and 1000 epochs. This meant that one complete simulation lasted anywhere between 800 (i.e., 200 × 4) and 4000 (i.e., 1000 × 4) epochs. Networks were trained for different numbers of epochs to ensure that the model-fit results were not idiosyncratic to the precise number of training epochs. The best quantitative fit of this model to the data in Experiments 1 and 2 is also shown below in Table 3.

4.3 | Results

To assess the quantitative fit of the predictions of the connectionist and Bayesian models to the data, we computed the root mean square (RMSE) and mean absolute error (MAE) between each model's predictions (for the connectionist model these were the average activation of the single output unit in response to each object; for the Bayesian model these were point estimates) and participants' mean responses to the objects across Experiments 1 and 2. One or both metrics have been used in previous simulation studies to assess a model's quantitative fit to behavioral data (e.g., Bhat et al., 2022; Buss & Spencer, 2014; Spencer et al., 2022; Steyvers et al., 2003; Stojnic et al., 2023). Lower values on each metric indicate better model fit. Table 3 above shows the model fits for the different connectionist and Bayesian model instantiations across both experiments and for different subsets of the data (e.g., model fit to the data overall, to the backwards blocking data only, etc.).

The main finding from Table 3 is that, although the Bayesian model outperformed the connectionist model in 2 situations and exhibited comparable performance in 1 situation, the connectionist model generally performed better than the Bayesian model (achieving better fits

to the data in 7 of the 10 situations). These findings suggest that participants may simultaneously be relying on associative processing and Bayesian inference, even when there is a greater tendency to rely on associative learning to reason about multiple potential causes. Stated somewhat differently, these data neither clearly support the conclusion that children rely exclusively on Bayesian inference to reason about retrospective reasoning, nor do they permit the conclusion that children rely exclusively on associative learning about such inferences. Instead, these data support the conclusion that children weigh these two cognitive mechanisms differently depending on the number of potential causes about which they are asked to reason. Bayesian inference may be given more weight than associative learning when there are a small number of potential causes (such as in Sobel et al., 2004), but as the number of causes and the information processing demands of the task increase, participants may give more weight to associative learning (such as in the current study).

5 | GENERAL DISCUSSION

The purpose of this study was to examine whether and how children engage in retrospective reasoning under more strenuous information processing demands, in which they must track the efficacy of more than two objects. Experiment 1 indicated that when shown first that three objects activated a machine together, and then that one of those objects did so individually, the other two objects were judged as less likely to be efficacious than analogous objects in a control condition. When the individual object did not activate the machine on their own, judgments of the efficacy of the other objects were not different from the control condition. However, in Experiment 2 when two of the three objects were revealed to activate or not activate the machine together (following the ABC+ event), children did not show evidence of retrospective inference in either type of trial.

We subsequently fit a Bayesian model and a connectionist model to the data from both experiments. The Bayesian model did make some qualitative predictions about retrospective reevaluation that were seen in children's responses in Experiment 1 but not Experiment 2. However, overall, the connectionist model tended to provide better fits across the trials and experiments. In contrast to findings where children only must reason about two objects, increasing the demand characteristics of the experiment caused children to default to a more associative strategy. This was especially true in Experiment 2 where the information processing demands were even greater than in Experiment 1.

The value of the connectionist model is that it provided a plausible account of the nature of children's associative processing in the current study. This can be seen perhaps most clearly when one considers how the model arrived at its judgements for the objects in the control trials in the Backwards Blocking condition in the first study. For example, when the model saw three objects activate the machine together and then a fourth do so independently, it arrived at its causal judgements based on a relatively simple counting strategy. During the simulation of this trial, when all four objects were first presented to the model,

the resulting difference at the output layer between the activation of the single output unit and the predicted activation of that unit was equivalent for all four objects. Thus, because the difference between the observed and predicted activation of the output unit was equivalent for all four objects, the model made equivalent weight adjustments in sign and magnitude to the connections between each object and the output unit. Crucially, these connections instantiated each object's association with the machine's activation. As such, because objects A–D were shown with the "machine's activation" (i.e., the output unit = 1) an equal number of times, the strength of the association between each object and the machine's activation was equivalent. Given that the connectionist model provided a better fit overall (and in various specific places) than the Bayesian model, it seems likely that children might also be relying on a similar associative-based counting procedure.

In contrast, the Bayesian model predicted a clear difference between the causal effectiveness of the first three objects and the fourth objects in the control trials. Because the fourth object was placed on the machine by itself, its causal status as an effective object is unambiguous and should be high. In contrast, when all children know is that three objects activate the machine together, the only conclusion they can come to is that at least one of the other three objects has efficacy. A Bayesian model predicts that the probability that each is efficacious is greater than the base rate, but not necessarily at ceiling. Whereas the Bayesian model made qualitative predictions about retrospective reevaluation in the experimental trials that were mostly upheld (at least in Experiment 1), children made closer to ceiling-level responses in the control trials (particularly in Experiment 2).

But what accounted for why children engaged in retrospective reevaluation in Experiment 1 but not in Experiment 2? The current study suggests that when tasks exceed children's information-processing abilities, they will resort to less sophisticated strategies and cognitive mechanisms such as associative learning (e.g., Cohen et al., 2002), even though multiple processes (in this case, associative learning and Bayesian inference) may be simultaneously in operation but to different degrees.

Before closing, some potential criticisms are worth noting. First, in the present study, children's reasoning overall was more consistent with an associative model than one that is described by Bayesian inference. Yet, that does not mean that Bayesian models could not explain the data under some circumstances. For instance, one of the pieces of evidence for a Bayesian description of causal inference is that children are sensitive to and make different inferences about the base rates of causal properties (e.g., Griffiths et al., 2011; Sobel et al., 2004; Sobel & Munro, 2009). Here, we did not present children with base rates prior to them making an inference. If we were to have done so, and in the case where the base rate that any one object was a blicket was rare, children might have been cued not to use an associative counting strategy, even given multiple potential causes. In other words, their inferences about unambiguous data (i.e., individual objects that specifically do or do not activate the machine) should be unchanged, but other inferences about ambiguous data might be different. Although we can think of modifications to the associative model presented here, which could theoretically consider such base rate data, the simple connectionist

model that we used to simulate the data here would be less explanatory than the Bayesian model we present.

A second criticism concerns the artificial nature of the paradigm used here, which was necessitated by the COVID-19 pandemic. Testing remotely on a computer screen may have introduced a level of noise in the data that is fundamentally different than testing in person with real objects. Future studies should replicate our study using real objects and a real blicket machine. If such a study revealed that participants performed more normatively than associatively in person, this would suggest that children's normative inferences may not be as robust as originally thought—it is present when tested in person but nearly absent when tested on a computer. Such a finding would be interesting regardless because it would add nuance to the literature on children's causal inferences.

A third criticism concerns the logic behind our model fitting. Our model fits were based on aggregating a group of children's yes/no responses and fitting those averages to a model's stochastic predictions. Previous studies on children's causal inferences used such an approach. However, studies with adults asked them to make more graded inferences (e.g., rate on a scale of 1–10 how likely a particular object caused the machine to activate). Given that we investigated a slightly older sample than some other studies of retrospective reasoning in children, such a graded response measure could be used in a reproduction of these studies. This could further help distinguish between the qualitative predictions of each model and the quantitative model fits. Relatedly, the logic behind our decision for the sample size of the studies was based on prior studies that demonstrated children's reasoning that were better described by Bayesian models. The choice of aggregating children's yes/no responses might not have been sufficiently powerful here to demonstrate some of the more subtle inferences predicted by a Bayesian account.

A fourth potential criticism concerns the absence of developmental change in children's retrospective reevaluations: Children's backwards blocking and indirect screening-off inferences were unrelated to age in both experiments. Although we failed to observe an age effect, the current results do have developmental implications. If we are correct that children resort to more associative forms of processing when their information-processing capacities are stretched, then these results suggest that if younger children are tested in a replication of the current study their inferences should be even more associative than the 5- and 6-year-olds tested here. This is because younger children presumably possess less robust information-processing abilities than older children and thus should be more affected by the increase in the number of objects used (relative to past studies on retrospective reevaluation) than the 5- and 6-year-olds tested here. Conversely, if children older than that tested here or even adults are tested in a replication of the current study, then not only should they be less affected by the increase in the number of objects presumably because they possess more information-processing abilities than the children tested here, but their inferences should also better align with the predictions of the Bayesian model than the associative model.

Although it remains to be seen whether these predictions will hold in younger children, recent data by Benton and Rakison (2023) do support these predictions: In a study that was similar to the current one—including in the use of three and four objects—adults' backwards blocking inferences better aligned with Bayesian processes than associative ones. When one considers this finding given the current results, a clearer developmental picture emerges: They not only suggest that cognitive processing evolves from a more associative approach in younger children to a more Bayesian-oriented strategy in adults but that this developmental shift may be supported by increases in underlying information-processing. Nonetheless, future research will want to test younger children than that tested here to better assess the viability of the current information-processing account.

5.1 | Conclusion

This study constitutes one of the first systematic attempts to examine retrospective reasoning in human children in the context of multiple potential causes. A longstanding view has been that the cognitive mechanism by which people reason about causal events is Bayesian inference rather than associative processes. The experiments reported here support a different conclusion: Although children possess both associative-based and more rational mechanisms for reasoning about causes, they may rely more on associative learning than on Bayesian inference when their information processing is taxed.

ORCID
Deon T. Benton https://orcid.org/0009-0007-4922-5760

ENDNOTE
[1] The Griffiths et al. (2011) model assumes that this can be learned through a hierarchical process, and there have been other ways this model has been refined (see e.g., Goodman et al., 2011; Ullman & Tenenbaum, 2020); we are presenting a simpler model for the purposes of this investigation.

REFERENCES
Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Beckers, T., Vandorpe, S., Debeys, I., & De Houwer, J. (2009). Three-year-olds' retrospective revaluation in the blicket detector task: Backward blocking or recovery from overshadowing? *Experimental Psychology*, *56*(1), 27–32. https://doi.org/10.1027/1618-3169.56.1.27

Benton, D. T., & Rakison, D. H. (2023). Associative learning or Bayesian inference? Revisiting backwards blocking reasoning in adults. *Cognition, 241*, 1–11. https://doi.org/10.1016/j.cognition.2023.105626

Benton, D. T., Rakison, D. H., & Sobel, D. M. (2021). When correlation equals causation: A behavioral and computational account of second-order correlation learning in children. *Journal of Experimental Child Psychology, 202*, 105008. http://doi.org/10.1016/j.jecp.2020.105008

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*, *129*(4), 640. https://doi.org/10.1037/rev0000313

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*, 35–65. https://doi.org/10.1016/j.cogpsych.2014.06.003

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*(4), 1156. https://doi.org/10.1037/a0026471

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. *The Developmental Psychology of Time*, 209–254.

Buss, A. T., & Spencer, J. P. (2014). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii. 1–103.

Butler, L. P., Gibbs, H. M., & Tavassolie, N. S. (2020). Children's developing understanding that even reliable sources need to verify their claims. *Cognitive Development*, *54*, 100871. https://doi.org/10.1016/j.cogdev.2020.100871

Caporaso, J. S., & Marcovitch, S. (2021). The effect of taxing situations on preschool children's responses to peer conflict. *Cognitive Development*, *57*, 100989. https://doi.org/10.1016/j.cogdev.2020.100989

Cohen, L. B. (1988). An information processing approach to infant cognitive development. In (L. Weiskrantz Ed.), *Thought without language* (pp. 211–228). Oxford University Press.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development*, *17*(3-4), 1323–1343. https://doi.org/10.1016/S0885-2014(02)00124-7

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, *47*(2), 109–121. https://doi.org/10.1016/S0022-2496(02)00016-0

Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the dimensional change card sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, *38*, 241–268. https://doi.org/10.1016/j.dr.2015.09.001

Erb, C. D., & Sobel, D. M. (2014). The development of diagnostic reasoning about uncertain events between ages 4–7. *PloS One*, *9*(3), e92285. https://doi.org/10.1371/journal.pone.0092285

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. S. B. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*(2–3), 86–102. https://doi.org/10.1016/j.dr.2011.07.007

Fernbach, P. M., Macris, D. M., & Sobel, D. M. (2012). Which one made it go? The emergence of diagnostic reasoning in preschoolers. *Cognitive Development*, *27*(1), 39–53. https://doi.org/10.1016/j.cogdev.2011.10.002

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive development*, *10*(4), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227. https://doi.org/10.1037/0096-3445.117.3.227

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review*, *118*(1), 110–119. https://doi.org/10.1037/a0021336

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222. https://doi.org/10.1111/1467-8624.00224

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620. https://doi.org/10.1037/0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085. https://doi.org/10.1037/a0028044

Greco, C., Hayne, H., & Rovee-Collier, C. (1990). Roles of function, reminding, and variability in categorization by 3-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 617.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*, 217–229. https://doi.org/10.1111/tops.12142

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*(8), 1407–1455. https://doi.org/10.1111/j.1551-6709.2011.01203.x

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384. https://doi.org/10.1016/j.cogpsych.2005.05.004

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition*, *103*(2), 180–226. https://doi.org/10.1016/j.cognition.2006.03.004

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*(3), 233–259. https://doi.org/10.1016/S0010-0277(96)00715-9

Hermes, J., Behne, T., Bich, A. E., Thielert, C., & Rakoczy, H. (2018). Children's selective trust decisions: Rational competence and limiting performance factors. *Developmental Science*, *21*(2), e12527. https://doi.org/10.1111/desc.12527

Heyes, C. (2012). Simple minds: A qualified defense of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Houwer, J. D., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(3), 965–985. https://doi.org/10.1080/02724980143000578

Kenderla, P., & Kibbe, M. M. (2023). Explore versus store: Children strategically trade off reliance on exploration versus working memory during a complex task. *Journal of Experimental Child Psychology*, *225*, 105535. https://doi.org/10.1016/j.jecp.2022.105535

Kimura, K., & Gopnik, A. (2019). Rational higher-order belief revision in young children. *Child Development*, *90*(1), 91–97. https://doi.org/10.1111/cdev.13143

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22. https://doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review*, *7*(4), 636–645. https://doi.org/10.3758/BF03213001

Larkin, M. J., Aitken, M. R., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1331.

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development*, *81*(3), 929–944. https://doi.org/10.1111/j.1467-8624.2010.01443.x

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 97.

Luchkina, E., Corriveau, K. H., & Sobel, D. M. (2020). I don't believe what you said before: Preschoolers retrospectively discount information from inaccurate speakers. *Journal of Experimental Child Psychology*, 189, 104701. https://doi.org/10.1016/j.jecp.2019.104701

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by 7-month-old infants. *Science*, 283(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Marr, D. (1982). *Vision: A Computational Investigation into the human representation and processing of visual information*. Henry Holt and Co., Inc. ISBN: 0716715678.

McClelland, J. L., & Thompson, R. M. (2007). Using domain-general principles to explain children's causal reasoning abilities. *Developmental Science*, 10(3), 333–356. https://doi.org/10.1111/j.1467-7687.2007.00586.x

McCormack, T., Butterfill, S., Hoerl, C., & Burns, P. (2009). Cue competition effects and young children's causal and counterfactual inferences. *Developmental Psychology*, 45(6), 1563. https://doi.org/10.1037/a0017408

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

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, 5(2), 193–207. https://doi.org/10.1016/0885-2014(90)90026-P

Powell, L. J., & Carey, S. (2017). Executive function depletion in children and its impact on theory of mind. *Cognition*, 164, 150–162. https://doi.org/10.1016/j.cognition.2017.03.022

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, 2, 64–99.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, 94(3), 249–273. https://doi.org/10.1016/j.jecp.2006.02.002

Rogers, T. T., & McClelland, J. L. (2014). Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, 38(6), 1024–1077. https://doi.org/10.1111/cogs.12148

Rovee-Collier, C. (1999). The development of infant memory. *Current Directions in Psychological Science*, 8(3), 80–85. https://doi.org/10.1111/1467-8721.00019

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

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

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology Section B*, 37(1b), 1–21. https://doi.org/10.1080/14640748508402082

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, 00, 1–51. https://doi.org/10.2307/1165893

Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology*, 22(1), 37–58. https://doi.org/10.1348/026151004772901104

Sobel, D. M., Erb, C. D., Tassin, T., & Weisberg, D. S. (2017). The development of diagnostic inference about uncertain causes. *Journal of Cognition and Development*, 18(5), 556–576. https://doi.org/10.1080/15248372.2017.1387117

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: the development of causal reasoning in toddlers and infants. *Developmental Psychology*, 42(6), 1103. https://doi.org/10.1037/0012-1649.42.6.1103

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science*, 10(3), 298–306. https://doi.org/10.1111/j.1467-7687.2007.00589

Sobel, D. M., Luchkina, E., & Tummeltshammer, K. (2020). Mechanisms for evaluating others' reliability when learning novel words. In J. Childers (Ed.), *Language and concept acquisition from infancy through childhood*. Springer, Cham. https://doi.org/10.1007/978-3-030-35594-4_9

Sobel, D. M., & Munro, S. E. (2009). Domain generality and specificity in children's causal inference about ambiguous data. *Developmental Psychology*, 45(2), 511524. https://doi.org/10.1037/a0014944

Sobel, D. M., 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.

Tenenbaum, J., & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24(4), 629–640. https://doi.org/10.1017/S0140525X01000061

Spencer, J. P., Ross-Sheehy, S., & Eschman, B. (2022). Testing predictions of a neural process model of visual attention in infancy across competitive and non-competitive contexts. *Infancy*, 27(2), 389–411. https://doi.org/10.1111/infa.12457

Steinbeis, N. (2018). Taxing behavioral control diminishes sharing and costly punishment in childhood. *Developmental Science*, 21(1), e12492. https://doi.org/10.1111/desc.12492

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27(3), 453–489. https://doi.org/10.1207/s15516709cog2703_6

Stojnić, G., Gandhi, K., Yasuda, S., Lake, B. M., & Dillon, M. R. (2023). Commonsense psychology in human infants and machines. *Cognition*, 235, 105406. https://doi.org/10.1016/j.cognition.2023.105406

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science*, 25(9), 1730–1738. https://doi.org/10.1177/0956797614540178

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, 2, 533–558. https://doi.org/10.1146/annurev-devpsych-121318-084833

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, 25(2), 127–151. https://doi.org/10.1006/lmot.1994.1008

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, 25(1), 161–169. https://doi.org/10.1177/0956797613502983

Walker, C. M., & Nyhout, A. (2020). Asking "why?" and "what if?": The influence of questions on children's inferences. *The Questioning Child: Insights from Psychology and Education*, 252–280. https://doi.org/10.1017/9781108553803.013

Weisberg, D. S., & Sobel, D. M. (2022). *Constructing science: Connecting causal reasoning to scientific thinking in young children*. MIT Press.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Stanford Univ Ca Stanford Electronics Labs.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, 126(6), 841. https://doi.org/10.1037/rev0000153

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., Sutherland, A., & Carlson, S. M. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, *68*(3), i-151.