Don’t throw the associative baby out with the Bayesian bathwater: Children process backwards blocking events *mostly* associatively when their information-processing abilities are stretched

Deon T. Benton1, David Kamper2, Rebecca M. Beaton1, David M. Sobel2

1Vanderbilt University  
2Brown University

Address correspondence to Deon T. Benton, Department of Psychology and Human Development, Vanderbilt University, Peabody College, 230 Appleton Place #552, Nashville, TN 37235

Data availability statement: The computational model scripts and data analysis code are available upon reasonable request.

Conflicts of interests: none

Don’t throw the associative baby out with the Bayesian bathwater: Children process backwards blocking events *mostly* associatively when their information-processing abilities are stretched

Submitted to *Developmental Science* on DATE

Abstract

Causal reasoning is a fundamental cognitive ability that enables children to learn about the complex interactions in the world. The mechanisms that underpin children’s causal reasoning, however, are not well understood. An open question is the extent to which children retrospectively reevaluate causal efficacy given ambiguous information, based on observing novel patterns of data. Here, we report two experiments that test children’s capacity to engage in such inferences. We also fit those data to different computational frameworks – one more associative and one more Bayesian – to consider the strengths and weaknesses of each approach, and the possibility that these approaches together better explain children’s causal reasoning than either approach individually.

Keywords: causal reasoning; cognitive mechanisms; computational models; associative learning; Bayesian inference

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; Gopnik & Sobel, 2000), to intervene on those relations to generate new effects (e.g., Butler et al., 2020; Gopnik et al., 2001; 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; Sobel, 2004; 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 algorithms that are underpinned by Bayesian inference. The crux of this idea is that learners use cognitive mechanisms that approximate Bayes inference to reason about causal events and to choose the causal hypothesis—among potentially infinitely many causal hypotheses—that is most consistent with the observed data (e.g., Bonawitz et al., 2014; Gopnik & Wellman, 2012; Griffiths & Tenenbaum, 2005, 2007; Xu, 2019). Crucially, causal reasoning starts with statistical learning capacities that are present in infancy (e.g., Gomez, 2002; Kirkham et al., 2002; Marcus et al., 1999; Saffran et al., 1996) but that develop into a system that infers abstract patterns of coherent causal structure from probabilistic data.

An alternative perspective is that associative learning alone is sufficient to describe children’s causal inferences. On this view, children build up a representation of causal structure from connecting and processing multiple associative relations and statistical regularities. 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 investigation between non-human animals and adults (e.g., Heyes, 2012) and studies of instrumental action and conditioning on 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 *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; Larkin et al, 1998; Kruschke & Blair, 2000; 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 judged that A was a blicket only when it activated the machine. Their judgments of object B also differed across these conditions. Children judged object B more likely to be a blicket when object A failed to activate the machine than when it did so. Using modified procedures, toddlers and even infants as young as 8 months 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 since been interpreted as support for Bayesian inference rather than associative learning. This is because some associative models such as the Rescorla & Wagner (1972) predict that the strength between object B and the machine’s activation is equivalent between the backwards blocking (where A is effective) and indirect screening-off (where A is not effective) trials. However, there are two facets of these data that warrant further consideration. First, what is not clear in these data is whether and how children reevaluate the causal status of object B. For instance, do children increase their belief that B is a cause when A fails to activate the machine but decrease their belief that B is a cause when A activates the machine, or are both occurring? when object A fails to activate the machine, do they increase their belief that B is efficacious, when object A activates the machine, do they decrease their belief about B, or are both occurring (Beckers et al., 2005; McCormack et al., 2009)?

McCormack et al. (2009) showed children a similar backwards blocking sequence (AB+, A+) to Sobel et al. (2004): Two objects (A and B) activated the machine together, and then object A activated it alone. They compared children’s causal status judgments for object B with a sequence in which a third object, 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). This control measure—which we adopt here—is a superior measure of assessing whether children reevaluate their causal judgments. Although these studies used different control trials, the critical point of agreement between these investigations is that at some point in development, children have the capacity to engage in backwards blocking reasoning (which is a form of ‘retrospective reevaluation'). A fundamental question remains, however: *How*—that is, by what cognitive mechanism—do children engage in this type of reasoning?

With this possibility in mind, the present study reconsiders children’s backwards blocking capacities in the context of an observation of the Griffiths et al. (2011) data. In one of their experiments (Experiment 3), 4-year-olds were shown that two pairs of compound stimuli (A and B, and then A and C) were causally effective. Although children considered A to be more effective than B or C, they did not treat A any differently than objects that were presented as a single compound (X and Y that together activated the machine). At issue is whether having to reason about more than two objects produced information processing demands that caused children to rely on simpler modes of processing. There is a wealth of data that is consistent with this idea (Doebel & Zelazo, 2015; Frye, Zelazo, & Palfai, 1995; Zelazo, Frye, & Rapus, 1996; Zelazo et al., 2003). For example, a recent study by Kenderla and Kibbe (2023) demonstrated that when faced with a challenging virtual memory game, 8- and 10-year-old children showed a decreased reliance on working memory and a greater dependence on manual exploration. The goal of this game was to find three cards with shared and differing features. Given that children were not required actively to maintain information in memory when manually exploring, manual exploration was an ostensibly simpler and 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. Furthermore, Sobel and Kirkham (2006) found that 8-month-olds exhibited backwards blocking inferences similar to preschoolers, whereas 5-month-olds' inferences appeared more associative in nature (Sobel & Kirkham, 2007). Moreover, 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 normative, although occasionally they will default to associative forms of processing (Hermes et al., 2018; Luchkina et al., 2020). Additionally, there is evidence that preschool-age children's 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, this research indicates that although children can process information at higher levels, if the task that they are given requires information-processing abilities that extend beyond what they possess, then there will be a tendency for them to process information at lower levels and to rely on less sophisticated strategies and cognitive mechanisms (e.g., Cohen et al., 1998). It is worth noting that this idea is not novel and is related to Cohen’s (e.g., Cohen, 1988; Cohen, 1991; Cohen et al., 1999) information processing account. The crux of this perspective is that children will process information at the highest possible level until their information processing abilities are stretched.

Here we asked whether children could engage in backwards blocking reasoning for three and four objects as opposed to two. Our design was like that used by McCormack et al. (2009). Children observed as three objects (A, B, and C) together caused a machine called a blicket together to activate, and then as A either caused or failed to cause the machine to activate by itself. They were then asked whether each object caused the machine to activate. These trials were compared to control trials in which a fourth object D either caused or failed to cause the machine to activate (following the activation of the machine by A, B, and C together). Children were said to engage in retrospective reevaluation if they treated objects B and C differently across these two types of trials. When A activates the machine alone, judgments of the causal effectiveness of B and C should be lower than the judgments of B’ and C’ when children see that a fourth, unrelated object is effective. When A does not activate the machine alone, judgments of the causal effectiveness of B and C should be higher than B’ and C’. Because McCormack et al. (2009) found that 5 and 6-year-olds made such retrospective inferences, we considered the same age range here. After presenting these behavioral data across two experiments, we present a pair of computational models to illuminate possible cognitive mechanisms by which children arrived at their causal judgements.

**Method**

**Participants.** Participants were 32 5-year-olds (16 boys and 16 girls; *M* = 64.81 months, range = 60-71 months, SD = 3.48) and 31 6-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 human children (e.g., Gopnik & Sobel, 2000; 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 (compared with 83%), 3% Black/African American (9%), 4% Asian/Asian American (4%), 0.5% Native American (1%), and 11% of Mixed Descent (3%). Sixteen percent of the sample identified as Hispanic/Latinx (compared with 17% of the population). 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: 38K. The median income for the population as measured by the 2020 Census was ~$74K.

**Materials.** The “device” used in the current study was a computer-animated version of the blicket detector. The device was a white rectangle with a black border that measured 5.99 cm × 23.47 cm, presented on a computer screen. If the device was “on”, the white region of the rectangle turned blue when objects came into contact with it. If the device was “off”, the white region remained white. A maximum of 4 differently colored circles were shown on the screen. Each circle 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 read. All video events were created in Microsoft PowerPoint.

**Procedure.** Participants were tested in a quiet room in local children’s museum. At the beginning of the experiment, all participants were shown a pretraining video. 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) above the machine and next to one another. Object A then descended until it contacted and immediately activated the machine (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 the it activated if at least one effective object was placed on it.

Following the pretraining phase, participants were given four trials. Half the participants received two backwards blocking trials and two backwards blocking control trials. The other half received two indirect screening off 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 Table 1.

Diagram

Description automatically generated

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.

**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 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 we will refer to here 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.

**Indirect Screening-Off Main 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.

**Results**

Figure 2 shows the number of times children responded “yes” to the question “Is this a blicket” for each object. Using this dependent measure, the data were entered into a five-way linear mixed-effects 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 yielded several experimental-effects and two-way interactions, which were qualified by 3 three-way interactions. These included a three-way interaction among Age, Condition, and Object, χ*2*(3) = 7.90, *p* = .05, a three-way interaction among Condition, Trial Number and Object, χ*2*(3) = 13.31, *p* = .006, and a three-way interaction among Condition, Trial Type, and Object, χ*2*(2) = 64.85, *p <* .001.

To explore the interaction among Age, Condition, and Object, we constructed separate two-way linear mixed-effects models between Age and Object for each condition. Age was included as a continuous fixed effect, Condition as a between-participants fixed effect, Object as a within-participants fixed effect, and participants as a random effect. Both linear models only yielded experimental effects of Objects, both χ2’s > 31.88, both *p*-values < .001, which indicated that participated treated the objects differently. Specifically, in the backwards blocking condition, participants considered object A (*M* = .89, *SD* = .31) to be more of a blicket than object B (M = .67, SD = .47), *t*(30) = 4.95, *p* < .001, and C (*M* = .71, *SD* = .46), *t*(30) = 3.89, *p* < .001*.* However, participants treated objects A and D (*M* = .85, *SD* = .36) equivalently, *t*(30) = .76, *p* = .45. In contrast, in the indirect screening-off condition, participants were less confident that object A (*M* = .54, *SD* = .50) was a blicket than object B (*M* = .79, *SD* = .41), and C (*M* = .84, *SD* = .37), both *t*’s > -5.03, both *p*-values < .001. However, participants were more confident that object A was a blicket than object D (*M* = .36, *SD* = .48), *t*(30) = 2.36, *p* = .02. This reflected the fact that the scores for object A were collapsed over Trial Type (in which A was seen by itself during the experimental trials but in combination with other objects during the control trials). Finally, participants were less confident that object D was a blicket than objects B and C, both *t*’s > 6.10, both *p*-values < .001.

To explore the second interaction among Trial Number and Object for each condition, Trial Number and Object were included as within-participants fixed effects and participants were included as a random effect. Although both linear models yielded experimental effects of Object, both χ2’s > 31.86, both *p*-values < .001, only the two-way linear mixed-effects model for the Indirect Screening Off condition yielded an interaction between Trial Number and Object, χ*2*(3) = 9.57. This interaction reflected the fact that participants treated the objects differently between the two trials. During trial 1 (when participants were asked to provide their first set of responses) participants treated object D (M = .55, SD = .50) as less of a blicket than object A (M = .72, SD = .45), *t*(114.97) = 2.31, *p* = .02, and object C (*M* = .78, *SD* = .42) as more of a blicket than either object B (*M* = .68, *SD* = .47), *t*(127) = -2.46, *p* = .02, or D, *t*(107.25) = 3.23, *p* < .01. In contrast, during trial 2, participants treated all the objects equivalently, all *t*’s < 1.78, all *p*-values > .08.

Chart, bar chart

Description automatically generated

Figure 2. Participants’ mean responses to whether each object was a blicket across the conditions and trial types. Bars show standard error.

To examine the final interaction among Condition, Trial Type, and Object, we constructed a set of one-way linear mixed-effects models for the experimental and control trials within the backwards blocking and indirect screening-off conditions. 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 linear model for the control trials within the backwards blocking condition did not reveal a significant effect of Objects,χ2(3) = 1.34, *p* = .72. This means that participants treated the objects similarly in the control trials of the backwards blocking condition. In contrast, the second one-way linear model for the experimental trials within the backwards blocking condition revealed a significant experimental effect of Objects, χ2(2) = 55.20, *p* < .001. This experimental effect reflected the fact that participants considered object A to be more of a blicket (*M* = .98, *SD* = 0.13) than object B (*M* = .55, *SD* = 0.50), *t*(30) = 6.45, *p<* .001, or object C (*M* = .61, *SD* = 0.49), *t*(30) = 5.62, *p* < .001. Participants treated objects B and C equivalently, *t*(30) = -1.07, *p* = .29.

The third and fourth one-way linear models for the experimental and control trials within the indirect screening-off condition both revealed a significant experimental effect of Objects, both χ*2*-values > 76.81, both *p*-values < .001. Participants considered object A (*M* = 0.26, *SD* = 0.44) in the ISO experimental trials and object D (*M* = 0.36, *SD* = 0.48) in the ISO control trials to be less likely to be blickets than any of the other objects, all *t*-values > -7.45, all *p*-values < .001. Participants treated object B and C equivalently in the experimental trials, *t*(30) = -0.77, *p* = .29, and objects A-C equivalently in the control trials, all *t*-values < -1.07, all *p*-values > .29.

**Evidence of backwards blocking reasoning.** To examine whether participants engaged in backwards blocking reasoning—operationalized as higher combined ratings of objects A-C in the control trials than of objects B and C in the experimental trials—data were entered into a two-way linear mixed-effects model 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, *χ2*(1) = 21.97, *p* < .001. This result indicated that participants did engage in backwards blocking reasoning: they provided higher combined ratings of objects A, B, and C in the backwards blocking control trials (*M* = 0.80, *SD* = 0.40) than the combined ratings of objects B and C in the backwards blocking experimental trials (*M* = 0.58, *SD* = 0.49).

For completeness, we ran the same analysis as above, but this time for the indirect screening-off condition. This analysis also only revealed a main effect of Trial Type, *χ2*(1) = 4.42, *p* = .04. The results mirrored the results for the backwards blocking condition. Participants provided higher combined ratings of objects A, B, and C in the indirect screening-off control trials (*M* = 0.84, *SD* = 0.36) than the combined ratings of objects B and C in the backwards blocking experimental trials (*M* = 0.77, *SD* = 0.42). Similar to the results above for the backwards blocking condition, this result indicated that when the object that is shown in isolation was also shown in combination with other objects participants show stronger retrospective reevaluations.

**Discussion**

This purpose of this study was to determine whether children would engage in backwards blocking reasoning when three and four objects were used as opposed to two objects. The results indicated that participants did engage in backwards blocking reasoning. Specifically, we found that participants were less confident that objects that were only shown in groups in the backwards blocking experimental trial (i.e., objects B-C) were blickets compared to the objects that were only shown in groups in the backwards blocking control trial (i.e., objects A-C). Although participants did engage in backwards blocking reasoning, there were aspects of their inferences, such as their causal judgements during the backwards blocking control trials, that could be interpreted as less normative and more associative in nature. In that trial participants treated all four objects equivalently during this trial. This pattern of responses makes sense if participants homed in on the fact that all four objects were shown (and thus associated) with the machine’s activation an equivalent number of times. Nonetheless, one could argue that a more normative response would have been to treat the unrelated object D as more causally effective than objects A, B, and C. This is because children were given unequivocal evidence that D was causally effective because the machine activated when it was placed on the machine alone (compared to equivocal evidence of A's, B's, and C's causal effectiveness). This result is presumably at odds with previous backwards blocking research such as that by McCormack et al., 2009) and likely stemmed from the fact that children had to reason about four objects as opposed to the standard of two objects. Asking children to reason about four objects could have imposed a tax on children’s information processing that caused aspects of their performance to be less normative, even if other aspects of their inferences were left unaffected such as their backwards blocking performance. We conducted Experiment 2 to determine further to what extent children’s backwards blocking performance is sensitive to demands on their information-processing. Experiment 2 was identical to Experiment 1 except for what participants saw following the initial event in which A, B, and C activated the machine together. In the backwards blocking experimental trials, objects A and B were placed on the machine together, which again caused it to activate. In the backwards blocking control trials, objects D and E together were placed on and subsequently caused the machine to activate. The experimental and control trials for the indirect screening-off event were similar to those for the backwards blocking event except that objects A and B or objects D and E failed to make the machine go. We reasoned that asking children to make inferences about multiple objects during what was the single event portion of the backwards blocking and indirect screening-off events in Experiment 1 would continue to place extensive demands on children’s information-processing abilities. As a result, their causal inferences should remain aligned with associative learning, even if aspects of their inferences are consistent with normative processes captured by Bayesian inference.

**Experiment 2**

Experiment 2 was similar to Experiment 1 except for the number of objects that descended onto the machine during the second halves of the backwards blocking and indirect screening-off trials. During the second half of the backwards blocking and indirect screening-off experimental trials, two objects A and B descended onto the machine following the first half of this event in which objects A, B, and C descended onto the machine. These objects (as well as control objects) either did (backwards blocking) or did not (indirect screening-off) activate the machine. All other aspects of this experiment were identical to Experiment 1.

**Method**

**Participants.** Participants were 32 5-year-olds (18 boys and 14 girls; *M* = 65.31 months, range = 60-75 months, SD = 3.65) and 32 6-year-olds (10 boys and 22 girls; *M* = 76.56 months, range = 65-83 months, SD = 4.33). Participants were 12% Asian/Asian American, 9% Black/African American, 10% Hispanic, and 69% White/Caucasian.

**Materials & Procedure.** The materials and procedure for Experiment 2 was identical to that for Experiment 1 with the following exceptions: During the backwards blocking experimental events 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 (i.e., turn blue). Likewise, during the backwards blocking control events, two objects D and E descended onto and subsequently caused the machine to activate. 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 indirect screening-off experimental and control trials were identical to the backwards blocking trials except that the machine neither activated when objects A and B descended onto the machine during the indirect screening-off experimental trials nor when objects D and E descended onto the machine during the indirect screening-off control trials. The left- and right-most positions of objects A and B during the experimental trials and objects D and E during the control trials were counterbalanced.

**Results**

Figure 3 shows the number of times children responded “yes” to the question “Is this a blicket” for each object. The data for this experiment were entered into a five-way linear mixed-effects 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 vs. E), 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 between Condition, Trial Type, and Object, χ*2*(2) = 185.38, *p <* .001.

A graph of different sizes of bars

Description automatically generated  
Figure 3. Participants’ mean responses to whether each object was a blicket across the conditions and trial types. Bars show standard error.

To examine the three-way interaction between Condition, Trial Type, and Object, we constructed a set of one-way linear mixed-effects models for the experimental and control trials within the backwards blocking and indirect screening-off conditions separately. Object was treated as the single within-participants fixed effect in these follow-up analyses, and participants were again treated as a random effect. The one-way linear model for the control trials within the backwards blocking condition did not reveal a significant effect of Objects,χ2(3) = 4.55, *p* = .34. Thus, as in Experiment 1, participants treated the objects similarly in the control trials of the backwards blocking condition. Also consistent with Experiment 1, the second one-way linear model for the experimental trials within the backwards blocking condition revealed a significant effect of Objects, χ2(2) = 14.26, *p* < .001. This result reflected the fact that participants considered object A to be more of a blicket (*M* = .84, *SD* = 0.37) than object C (*M* = .63, *SD* = 0.49), *t*(31) = 3.38, *p<* .01. Participants treated the remaining objects equivalently.

As with Experiment 1, the third and fourth one-way linear models for the experimental and control trials within the indirect screening-off condition both revealed a significant experimental effect of Objects, both χ*2*-values > 1100.90, both *p*-values < .001. During the indirect screening-off experimental trials, participants considered objects A (*M* = 0.08, *SD* = 0.27) and B (*M* = 0.05, *SD* = 0.21) to be less likely to be blickets than object C (*M* = 0.98, *SD* = 0.13), both *t*-values > -21.10, both *p*-values < .001. Participants treated objects A and B equivalently, *t*(31) = 1.43, *p* = .16. During the indirect screening-off control trials, participants considered objects D (*M* = 0, *SD* = 0) and E (*M* = 0, *SD* = 0) to be less likely to be blickets than object A (*M* = 0.98, *SD* = 0.13), object B (*M* = 0.95, *SD* = 0.21), and object C (*M* = 0.97, *SD* = 0.18), all *t*-values > 35.79, all *p*-values < .001. Participants treated objects A-C equivalently.

**Evidence of backwards blocking reasoning.** We next examined whether participants engaged in backwards blocking reasoning using the operationalization of it from Experiment 1. Data were entered into a two-way linear mixed-effects model 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, *χ2*(1) = 3.94, *p* = .05: Participants provided higher combined ratings of objects A, B, and C in the backwards blocking control trials (*M* = 0.79, *SD* = 0.41) than the combined ratings of object C in the backwards blocking experimental trials (*M* = 0.63, *SD* = 0.49). For the indirect screening-off condition, data were entered into a two-way linear mixed-effects model with Trial Type and Object as the within-participants fixed effects and participants as the random effect. This analysis revealed neither a main effect of Objects, *χ2*(2) = 1.40, *p* = .49, nor a main effect of Trial Type, *χ2*(1) = 0.35, *p* = .55.

Discussion

Data from Experiment 2 showed that participants continued to engage in backwards blocking reasoning, although there were aspects of the data that could be seen as less normative. Once again the relevant aspect appears to be participants’ equivalent responses to the five objects in the backwards blocking control trial. A more rational response perhaps based on Bayesian inference might be for participants—after seeing that objects D and E together activated the machine when placed on it (following the placement of objects A-C on the machine)—to consider objects D and E to be better blicket candidates than objects A-C (which they might treat equivalently). This suggests the possibility that participants relied on Bayesian inference and associative learning, although it remains unclear whether one form of processing was given more weight than the other. We address whether in fact participants relied on both processes (or one more than the other) below in the Computational Model section. Of particular import is whether participants’ causal inferences across Experiments 1 and 2 were best explained by an associative-learning mechanism, a Bayesian-inference mechanism, or some combination of both.

**Computational Models**

We fit two different 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).

**Bayesian Model.** The Bayesian model starts with a set of hypotheses *H*. Each hypothesis *h* ∈ *H* is assigned a *prior probability*, *p*(*h*), which indicates the initial belief in 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*), given an updated belief about each hypothesis given the data. This is done using Bayes’ rule, shown in Equation 1:

(1)

In this formula, *p*(*d | h*) is the probability of the data *d* given each a particular hypothesis *h* (also known as the *likelihood*).

Forming the initial hypothesis space relies on assuming that there is a set of objects *O* and a set of detectors *D*, such that any object *o* ∈ *O* can potentially cause any detector *d* ∈ *D* to activate. Given that participants are shown that the machine activates when blicket objects are placed on its surface, a hypothesis *h* corresponds to a structure that posits whether individual objects have the causal efficacy to activate the detector (see Griffiths & Tenenbaum, 2005, for more computational details). Griffiths et al. (2011) describe the formal parameterization of this hypothesis space and model that results in the hypothesis space shown in Figure 3.

To instantiate the model, each hypothesis is given a prior probability *p*(*h*), which is a function of the child’s belief about 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*) = *ρ*3. Hypotheses 1, 2, and 4 posit exactly 2 blickets, so their *p*(*h*) = *ρ*2(1−*ρ*). Hypotheses 3, 5, and 6 each posit 1, making their *p*(*h*) = *ρ*(1−*ρ*)2. Finally, Hypothesis 7 posits no blickets, making its *p*(*h*) = (1-*ρ*)3.

**Timeline

Description automatically generated with medium confidence**

Figure 3. 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.

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]](#footnote-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 determines the probability that an object is a blicket is based on the posterior probability of the models in the hypothesis space; that is, the probability that any object *o* is a blicket given the data *d* can be calculated by the equation in (2)

where *p*(*o*→*E* | *h*) is 1 if there is an edge between that object and the detector in h, 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: .5, .65, .8, .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 probabilities, it was possible to compare the model’s predictions for the different probabilities to children’s actual treatment of the objects. The quantitative fit of this model to the data in Experiments 1 and 2 are shown below in Table 1 in Results.

**Connectionist model**. We also built a set of two-layer connectionist models. One set of these models corresponded to Experiment 1 and the other set corresponded to Experiment 2. The model architecture for the Experiment 1 simulations is shown in Figure 4. 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, as was the case for the children.

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 .05 but no momentum. Model weights were initialized to small random values (distribution range = ± 0.1), and the output units used sum-squared activation functions (which enabled the weights to be modified with training). 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).

A diagram of a machine

Description automatically generated  
Figure 4. The connectionist model used to simulate Experiment 1.

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 backwards blocking control trials 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 indirect screening off experimental and control trials 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 A- and D- phases of the indirect screening-off experimental and control trials. Each phase of the simulations—which were shown twice to be consistent with the behavioral study—lasted anywhere between 200 and 1,000 epochs. This meant that one complete simulation lasted anywhere between 800 (i.e., 200 × 4) and 4,000 (i.e., 1,000 × 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. Below we show the quantitative fit of this model to the data in Experiments 1 and 2.

**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 1 below 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.).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| (A) Model fit to the human data overall | | | | | | | |
| Experiment 1 | | | | Experiment 2 | | | |
| Connectionist‡ | | Bayesian Model | | Connectionist‡ | | Bayesian Model | |
| RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| .15 | .11 | .17 | .17 | .13 | .11 | .16 | .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 |
| .19 | .16 | .20 | .18 | .13 | .11 | .15 | .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 |
| .08 | .07 | .18 | .16 | .11 | .11 | .12 | .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 |
| .19 | .16 | .19 | .16 | .16 | .14 | .14 | .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 |
| .10 | .08 | .20 | .17 | .11 | .09 | .17 | .17 |

Table 1. 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. ‡ Corresponds to the best fitting overall model based on average RMSE and MAE.

The main finding from Table 1 is that, although the connectionist model generally performed better than the Bayesian model (achieving higher performance in 7 of 10 total situations), the Bayesian model either outperformed the connectionist model in 2 situations or exhibited comparable performance in 1 situation. These model 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 causally nor do they permit the conclusion that children rely exclusively on associative learning to reason about causes. 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 give more weight to associative learning (such as in the current study).

General Discussion

The purpose of this study was to examine whether children can engage in backwards blocking reasoning for three or more objects. We found that children did engage in backwards blocking reasoning: In both experiments, children were less confident that objects that were only shown in groups (and never in isolation) in the backwards blocking experimental trials (i.e., objects B-C) were blickets compared to the objects that were only shown in groups in the backwards blocking control trials (i.e., objects A-C). We subsequently fit a Bayesian model and a connectionist model to the data in Experiment 1 and found that the connectionist model tended to provide the best fit across a number of contexts. For example, the connectionist model provided a better fit to the data overall across both experiments, to the backwards blocking data across both experiments, to the indirect screening-off data in Experiment 1, and to the control data across both experiments. Yet, the Bayesian model provided a better fit to the indirect screening-off data in Experiment 2 and to the experimental data in Experiment 2 and provided an equivalent fit to the experimental data in Experiment 1. Thus, these data seem to implicate the operation of Bayesian inference (or some mechanism that approximates this computational level description) and an associative learning mechanism, with a greater tendency for participants to process the events associatively. This finding extends previous research on this topic by showing that when children are asked to reason about three and four causes (as opposed to the standard two objects used in previous research), children reason normatively and associatively, with a clear tendency for associative processing.

Although children largely processed the present events associatively, a question that we have not yet answered concerns the exact nature of this associative learning. Specifically, how did the connectionist models (and perhaps by extension, the children)—through associative learning—arrive at their causal judgements? To understand how the connectionist model’s judgements arose mechanistically, consider the control trial in the backwards blocking (i.e., ABC+ D+) condition. 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. Stated plainly, the model’s responses were based on a simple associative “counting” strategy based on the number of times that a given object appeared with the blicket effect. As such, because objects A-D were shown with the “machine’s activation” (i.e., the output of the output unit) an equal number of times in the control trials of the backwards blocking condition, the strength of the association between each object and the machine’s activation was equivalent. Given that participants’ responses mostly matched the model’s predictions, this suggests that in those cases in which children’s responses were better explained by the connectionist model than by the Bayesian model children relied on a similar associative process.

More broadly, data from the current experiments align with the idea that children's reasoning about events is influenced by the level of strain on their information processing capacities. Specifically, the current set of findings is consistent with a large body of research (cited in the Introduction) that indicates that despite a general tendency for learners to process information at the highest possible level (e.g., Cohen et al., 2002), when tasks exceed children’s information-processing abilities, they resort to lower-level processing and less sophisticated strategies and cognitive mechanisms such as associative learning. A testable prediction of this account is that there should be a point at which children go from using a simple associative-based counting mechanisms in contexts like the present one to more rationale processes like Bayesian inference. Ongoing work in one of our labs is testing this possibility.

Before closing, some potential criticisms are worth noting. First, we cannot be sure that there are no contexts in which the balance of associative processing to Bayesian inference can be flipped. In the present study, children mostly relied on associative processing and minimally on Bayesian inference, but it is possible that in the right situation children would mostly rely on Bayesian inference and minimally on associative processing. For example, if participants assumed that blickets were rare in the present context, then participants should be *more* likely to block redundant causes; in other words, participants should be *less* likely to treat all potential blickets (expect for the ones that are explicitly shown to be causal) equally.

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. 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.

**Conclusion**

This study constitutes one of the first systematic attempts to examine backwards blocking and indirect screening-off reasoning in human children in the context of multiple three and four candidate 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: children rely on associative learning *and* Bayesian inference to reason about causal events.

References

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.

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.

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.

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.

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: children's use of simplicity and probability to constrain inference. Developmental psychology, 48(4), 1156.

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.

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.

Caporaso, J. S., & Marcovitch, S. (2021). The effect of taxing situations on preschool children’s responses to peer conflict. *Cognitive Development*, *57*, 100989.

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

Cohen, L. B. (1991). Infant attention: An information processing approach. In M.J. Weiss and P. R. Zelazo (Eds.), Newborn attention: Biological constraints and the influence of experience, (pp. 1-21). Norwood, N.J.: Ablex

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. Cognitive Development, 17(3-4), 1323-1343.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. Journal of Mathematical Psychology, 47(2), 109-121.

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.

Erb, C. D., & Sobel, D. M. (2014). The development of diagnostic reasoning about uncertain events between ages 4–7. PloS one, 9(3), e92285.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive development*, *10*(4), 483-527.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. Journal of Experimental Psychology: General, 117(3), 227.

Gomez, R. L. (2002). Variability and detection of invariant structure. Psychological Science, 13(5), 431-436.

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. Current Directions in Psychological Science, 24(2), 87-92.

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.

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.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138*(6), 1085.

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., 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.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. Cognitive psychology, 51(4), 334-384.

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. Cognition, 103(2), 180-226.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*(3), 233-259.

Heyes, C. (2012). Simple minds: a qualified defence of associative learning. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1603), 2695-2703.

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.

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.

Kimura, K., & Gopnik, A. (2019). Rational higher‐order belief revision in young children. *Child Development*, *90*(1), 91-97.

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.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. Psychological review, 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.

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.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. *Cognition*, *25*(3), 265-288.

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.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. Science, 283(5398), 77-80.

McClelland, J. L., & Thompson, R. M. (2007). Using domain‐general principles to explain children's causal reasoning abilities. Developmental Science, 10(3), 333-356.

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.

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.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193-207.

Powell, L. J., & Carey, S. (2017). Executive function depletion in children and its impact on theory of mind. *Cognition*, *164*, 150-162.

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.

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.

Rovee-Collier, C. (1999). The development of infant memory. Current directions in psychological science, 8(3), 80-85.

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

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

Shultz, T. R. (1982). Rules of causal attribution. Monographs of the society for research in child development, 1-51.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology Section B*, *37*(1b), 1-21.

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.

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: the development of causal reasoning in toddlers and infants. *Developmental psychology*, *42*(6), 1103.

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.

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.

Steinbeis, N. (2018). Taxing behavioral control diminishes sharing and costly punishment in childhood. *Developmental science*, *21*(1), e12492.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. Cognitive science, 27(3), 453-489.

Stojnić, G., Gandhi, K., Yasuda, S., Lake, B. M., & Dillon, M. R. (2023). Commonsense psychology in human infants and machines. Cognition, 235, 105406.

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.

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological science*, *25*(1), 161-169.

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.

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.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive development*, *11*(1), 37-63.

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... & Carlson, S. M. (2003). The development of executive function in early childhood. *Monographs of the society for research in child development*, i-151.

1. The Griffiths et al. (2011) model assumes that this can be learned through a hierarchical process; we are presenting a simpler model for the purposes of this investigation. [↑](#footnote-ref-1)