Don’t throw the (associative-learning) baby out with the bathwater just yet: Backwards-blocking reasoning with *multiple* potential causes in human children

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

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 code and network-modeling scripts are available upon reasonable request.

Conflicts of interests: none

Abstract

Causal reasoning is a fundamental cognitive ability that enables humans to learn about the complex interactions in the world around them. However, the available evidence suggests that the mechanism or set of mechanisms that underpin causal reasoning are not well understood. It is unclear, for example, whether causal reasoning is underpinned by a Bayesian mechanism, an associative mechanism, or both. Some theorists have argued that a Bayesian mechanism underpins causal reasoning because it can better account for backward-blocking (BB) and indirect screening-off (IS) findings in children and adults (e.g., Sobel, Tenenbaum, & Gopnik, 2004). However, the evidence is mixed about the extent to which learners engage in both kinds of reasoning. Here, we report three experiments that examine to what extent adults engage in BB and IS reasoning using the blicket-detector design (e.g., Gopnik et al., 2001), what mechanism best explains their behavior in this task, and under what conditions are adults’ causal ratings consistent with the predictions of the three competing computational and analytical models. The results of Experiment 1 revealed that adults’ causal ratings in the backwards-blocking condition (as well as in the indirect screening-off condition) were consistent with the predictions of the traditional and modified Rescorla-Wagner models when asked to reason about two objects. The results of the present study suggest that adults use associative processes to reason about two objects but a Bayesian-inference-like process to reason about three or more objects.

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

Few capacities are more important than the ability to reason and make inferences about cause-and-effect relations. This is a key cognitive ability that enables human learners to encode causal relations to inform prediction and inference (e.g., Leslie & Keeble, 1987; Oakes & Cohen, 1990), to intervene on those relations to generate new effects (e.g., Gopnik et al., 2001), and counterfactually to reason about causal events to determine what would have happened if alternative actions were chosen (e.g., Harris, German, & Mills, 1996; Sobel, 2004).

Most of the studies on causal reasoning in human children to date have used the blicket-detector design. In these studies, children are introduced to a machine called the "blicket detector" and told that the machine lights up and plays music when certain objects—namely, "blickets"—are placed on it. Children are then asked to determine which objects are blickets and to “make the machine go” by placing the blicket on the machine. One such commonly accepted finding is that causal reasoning generally emerges by the first half of the third year of life (e.g., Benton, Rakison, & Sobel, 2021; Gopnik & Sobel, 2000; Gopnik et al., 2001; Kimura & Gopnik, 2019; Meltzoff, Waismeyer, & Gopnik, 2012; Sobel & Kirkham, 2006; Sobel & Munro, 2006; Walker & Gopnik, 2014). However, the finding that has generated that the most discussion was that by Sobel, Tenenbaum, and Gopnik (2004). They showed that by 4 years of age children can engage in two forms of causal reasoning: “backwards-blocking” (henceforth, BB) reasoning and “indirect screening-off” (henceforth, ISO) reasoning. BB reasoning is the process by which learners discount or “block” causal cues shown to be redundant in producing some effect. ISO reasoning is the process by which learners discount or “screen off” a causal cue that is shown not to produce some effect and whose causal status is known unambiguously.

In their study, children were first shown that two novel objects, A and B, together caused the detector to activate when they were placed on the machine. Children were then shown that object A alone either failed to activate the detector (i.e., AB+ A-; ISO condition) or activated the detector when placed on it (i.e., AB+, A+; BB condition). Children in both conditions were then asked which of the two objects were blickets and to make the machine go by placing the blicket on the detector. Sobel et al. (2004) found that the 4-year-olds, but not the 3-year-olds, responded by placing object B on the machine in the ISO condition; these same children responded by placing object A on the machine in the BB condition. Subsequent research by Sobel and Munro (2009) found that 3-year-olds could also engage in BB and ISO reasoning if the machine possessed animate-like qualities.

These findings were interpreted to mean that human children can engage in BB reasoning and that Bayesian inference was likely the cognitive mechanism that underpinned their performance. Proponents of the Bayesian-inference perspective maintain that learners use a simple form of Bayes’ rule to reason about causal events and to choose the causal hypothesis—within a hypothesis space that can be exponentially large—that is most consistent with the observed data (e.g., Sobel et al., 2004; Gopnik & Wellman, 2012). These findings were also argued to rule out certain classes of associative learning models such as the traditional Rescorla-Wagner (RW) model (e.g., Rescorla & Wagner, 1972; Griffiths et al., 2011; Sobel et al., 2004). This it is because this model predicts that object B should be treated equivalently across the BB and ISO conditions. As we have just illustrated, this prediction is at odds with participants’ actual treatment of object B across conditions.

However, some caution should be exercised before accepting the conclusion that Bayesian inference rather than associative learning underpins how children process BB events. One reason concerns it is unclear whether, in fact, children in previous studies engaged in BB reasoning. This is because there are significant problems with how BB reasoning itself has typically been measured. For example, Sobel et al. (2004; see also Beckers et al., 2009; McCormack et al. 2009, Exp. 1; Sobel & Kirkham, 2006) operationally defined BB reasoning as greater B choices in the ISO condition than in the BB condition (although for alternative operationalizations, see De Houwer, Beckers, & Glautier, 2002; Larkin, Aitken, & Dickinson, 1998; Griffiths et al., 2011; Kruschke & Blair, 2000; Lovibond et al., 2003; Shanks, 1985; Van Hamme and Wasserman, 1994). This operationalization suffers from two key limitations. First, as Beckers et al. (2005) and McCormack et al. (2009) pointed out, it cannot be determined *why* participants treated object B differently between the BB and ISO condition. For example, participants could have treated B differently between the BB and ISO conditions because of a BB effect, an ISO effect, or both. Participants could have also treated B differently between these conditions because they observed a positive effect during the elemental (i.e., A+) phase in the BB condition but a negative effect during the elemental (i.e., A-) phase in the ISO condition. Crucially, neither of these alternative explanations would be evidence of a true retrospective reevaluation of object B by participants based on A’s *relation to and effect* *on* object B across both conditions (which is the intended inference).

The operationalization that we adopt here—which was independently discovered by McCormack et al. (2009, Exp. 2)—eschews this limitation. Specifically, we operationalize BB reasoning by comparing how participants treat object B following an AB+ A+ sequence of events (i.e., the BB experimental condition) to how they treat object B following an AB+ C+ sequences of events (i.e., the BB control condition). These two conditions differ in terms of the object that is shown during the elemental phase (i.e., A or C) and that object’s *relation* to B (and thereby the potential impact that this object has on how B is treated). For example, in the BB experimental condition, a dependency is presumably established between objects A and B because both objects appear together during the compound phase (i.e., the AB+ phase) of the condition. This means that A’s observed causal efficacy during the subsequent elemental phase (i.e., the A+ phase) *should* affect participants’ treatment of object B. In contrast, in the BB control condition, object C never appeared with object B. This means that C’s causal status should not affect participants’ evaluation of B. Crucially, the blicket effect itself is held constant such that the machine activates across both conditions and the compound and elemental phases.

Another reason to exercise caution before accepting the claim that human beings use Bayesian inference to engage in BB reasoning is that it remains unknown whether human children engage in BB reasoning for three (or more) objects. Consider a modified version of the standard BB event—which we use in the present study—in which children first see an ABC+ sequence followed by an A+ sequence. If BB reasoning is unaffected by the number of presented objects, then children should be less likely to label objects B *and* C as blickets compared to the same objects in a control event in which ABC+ is followed by D+. This question is worth addressing because if the goal is to elucidate and better understand the nature of the cognitive mechanisms that subserve causal reasoning *in the real world*, then it is crucial that we understand how causal reasoning unfolds in situations that mirror children’s natural environments.

One may question whether asking children to reason about three to four objects can really tell us more about the cognitive mechanisms that underpin causal reasoning than asking children to reason about two objects. This is because the two situations differ trivially by at most two potential causes. However, if Bayesian inference is the cognitive mechanism that underpins BB reasoning in human beings, then the difference between these two settings is far from trivial. This is because in the two-cause setting, participants need only to determine which of *four* candidate causal hypotheses generated the observed data. In contrast, in the three- or four-cause setting, participants need to determine which of *eight* (in the case of 3 objects) or *sixteen* (in the case of 4 objects) hypotheses is the one that generated the observed data. This means that participants must consider up to four times as many causal hypotheses across contexts.

Crucially, this difference may have important implications about whether an associative-learning mechanism or a Bayesian-inference mechanism underlies causal reasoning in children. For instance, it is possible that when children’s information-processing abilities are taxed—such as when they are asked to reason about three (or more) objects (see the General Discussion for a fuller discussion)—they may resort to simpler modes of causal inference that are better explained by associative processes. Thus, if participants’ performance in a multiple-object BB task adheres to the predictions of an associative-learning mechanism (see below), this would suggest associative learning is sufficient to account for causal learning in human children.

There is one final reason to exercise caution before accepting the claim that Bayesian inference subserves how human children reason. This concerns the fact that other associative-learning processes beyond the traditional RW model may well explain how children process causal events that involve more than the number of objects standardly used in causal studies with children. For example, one class of models that *could* account for how children processed the BB events in the present study is connectionist artificial neural networks (e.g., Rogers & McClelland, 2014; Rumelhart et al., 1986). These models consist of “neuron-like” processing units, which are organized into layers. These layers typically include an input layer, a hidden layer, and an output layer. Layers within a connectionist model are connected to each other via modifiable weights. The input layer of a connectionist model is typically connected to the hidden layer immediately “above” it via adjustable connections. The hidden layer, in turn, is typically connected to the output layer via a different set of adjustable connection weights. Training in these models typically proceeds by presenting them with some pattern of activation along the input layer, comparing the model’s “observed” pattern of activation along the output layer to some “desired” pattern of activation along the same layer, and then using one or more learning algorithms or procedures to adjust the weights. The purpose of these weight adjustments is to reduce the difference between the observed and desired output.

An essential feature of these models is that they are fundamentally associative-learning devices—the weights encode associations among layers in the model, and the learning procedures operate either to strengthen or to weaken those associations. This means that these models can serve as a proof of concept that associative learning is sufficient to account for various aspects of cognitive development (e.g., Benton & Lapan, 2022; Benton et al., 2021; Cohen et al., 2002; Flusberg et al., 2010; Mareschal et al., 2000; Morton & Munakata, 2002; Munakata et al., 1997; Quinn & Johnson, 2000; Rakison & Lupyan, 2008; Westermann & Mareschal, 2004; for an extensive review see Yermolayeva & Rakison, 2013). Here we show that this modeling formalism is sufficient to explain how children processes the present BB events and provides a better account of the present data than a simple Bayesian model.

**Possible cognitive mechanisms underlying BB reasoning for multiple potential causes**

A key goal of the current experiments was to elucidate whether a Bayesian-inference or an associative-learning mechanism subserved children’s causal inferences. Thus, a critical first step was to derive the predictions of a simple Bayesian model and to instantiate the current experiments in a simple connectionist (computational) model to determine what predictions these models make for how children should process the present causal events. We restrict our discussion below to each model’s predictions but interested readers should consult the Appendix for the formal details of the Bayesian model. Details of the connectionist model are provided immediately below in the main text.

**Bayesian inference**

Proponents of the Bayesian-inference account maintain that human learners use a simple form of Bayes’ rule to determine which hypothesis is responsible for the observed data. Learners achieve this by combining their prior beliefs about each hypothesis (this is sometimes called the “prior”) with the likelihood that the observed data was produced by a particular hypothesis (this is sometimes called the “likelihood”). Given that learners were asked to reason about three potential causes (i.e., objects A-C) during the experimental trials in both the BB and ISO conditions and four potential causes during the control trials in both the BB and ISO conditions, the corresponding psychological hypothesis spaces consist, respectively of 8 and 16 hypotheses. Figure 1 below shows the hypothetical hypothesis space for three objects. Given that the predictions that a simple Bayesian model makes depends on the “prior probability” that any given object is a blicket, below we show the predictions of the model when the probability of a blicket is .5, .65, .8, .95, and 1. We also plotted the model’s predictions for various prior probabilities because it was unclear what participants’ baseline assumptions would be 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 (qualitatively and quantitatively) the model’s predictions for the different probabilities to children’s actual treatment of the objects. Figure 2A-E shows these predictions.

**Timeline

Description automatically generated with medium confidence**

Figure 1. 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.

|  |  |
| --- | --- |
| Chart, bar chart  Description automatically generated  A | Chart, bar chart  Description automatically generated  B |
| Chart, bar chart  Description automatically generated  C | Chart, bar chart  Description automatically generated  D |
| Chart  Description automatically generated  E |  |

Figure 2. This figure displays the of the Bayesian model for the BB and ISO conditions when *P*(Blickets) = .5 (3A), *P*(Blickets) = .65 (3B), *P*(Blickets) = .8 (3C), *P*(Blickets) = .95 (3D), *P*(Blickets) = 1 (3E).

As shown in Figure 2A-E, the model predicts that during the BB experimental and control trials participants should be maximally confident that objects A and D are blickets. In contrast, during the ISO experimental and control trials, participants should be maximally confident that objects A and D are not blickets. Importantly, the model makes these predictions regardless of the prior probability of blickets. In contrast, the model predicts that participants should categorize objects B and C at the same rate across the main trials in the BB and ISO conditions and objects A-C at the same rate across the corresponding control trials.

**Associative learning**

In contrast to the Bayesian model, we built a simple, two-layer connectionist computational model (Figure 3). This model was designed to simulate the current experiments. The model used to simulate the experiment reported here consisted of an input layer and an output layer—there were no hidden layers in these models. The input layer for the model consisted of four units, and the output layer consisted of a single unit. Each input unit corresponded to each of the four possible objects used in the experiment. Whenever an object was present, the activation value of its corresponding input unit was set to a value of “1”; the activation of these units was set to a value of “0” if the corresponding objects were not present. If a predetermined blicket was presented at the input layer, then the model was trained to turn on the single output unit (i.e., to produce an activation of 1). This process corresponded to an object activating the blicket machine. All simulations used a learning rate of .05 but no momentum. Model weights were initialized to 0, and the output units used sum-squared activation functions.

Diagram

Description automatically generated  
Figure 3. The neural network model used in the simulations reported here.

The models were trained on the same events as children. For example, networks, like children, were assigned randomly to the ISO condition or to the BB condition. To match the behavioral experiment, networks experienced two of each kind of event within a given condition. For example, during the two “experimental trials” for networks in the BB condition, the first three input units were turned on (i.e., the activation of each input node was set to a value of 1), and the network’s task was to learn to activate the single output unit (i.e., to set the activation of the single output unit to 1). Turning on the first three input units simulated placing objects A, B, and C on the blicket machine, and training the model to turn on the single output unit corresponded to networks learning that A-C caused the machine to activate. This segment of training corresponded to the ABC+ events. During the subsequent A+ elemental trials, only the first input unit was turned on, but again the network’s task was to activate the single output unit. The BB control trials were identical to the BB experimental trials except that the fourth input unit (corresponding to object D) rather than first input unit was turned on. The ISO experimental and control trials were identical to the BB experimental and control trials except that the network was trained to turn off the single output unit (i.e., set its value to 0) during the elemental phase of the ISO experimental and control trials. The compound (e.g., ABC+) and elemental (e.g., A+) phases—which were shown twice to be consistent with the behavioral study—lasted 200 epochs each. This mean that one complete simulation lasted 800 (i.e., 400 × 2). epoch. The predictions that this model makes for how participants should treat the BB and ISO events during the experiment are shown below in Figure 4.Although we report the results of a model that was trained for 800 total epochs in the main text, we ran additional simulations to ensure that the main results were not idiosyncratic to the precise number of training epochs.

Chart, bar chart

Description automatically generated

A

Figure 4. Connectionist model predictions for how participants should treat the objects between the main and control trials of the BB and ISO conditions.

B)

As can be seen in Figure 4 above, the model predicts that participants should treat objects A-C equivalently during the BB experimental trials. In contrast, the model predicts that participants should treat object A as more of a blicket than objects B and C during the same trials. For the ISO experimental trials, the model predicts that participants should treat object A as less of a blicket than objects B and C during the ISO experimental trials.

It should be noted that the Bayesian and connectionist model make identical qualitative predictions for all the conditions and trials except for the BB control condition: the simple connectionist model predicts that participants should treat objects A-D equivalently during this trial, whereas the simple Bayesian model predicts that participants should only treat objects A-C equivalently but should be maximally confident that object D is a blicket. Thus, it should be possible to determine which model participants relied on based on their performance during the BB control condition. Interestingly, both models predict that participants’ treatment of the redundant causes between the BB experimental and BB control trials should not differ. Likewise, both models predict that participants’ treatment of the redundant causes between the BB main and ISO main conditions should not differ. Thus, the simple connectionist model and Bayesian model do not predict BB reasoning according either to the new or old operationalization of BB reasoning. The present study was designed to test these predictions.

**The present investigation**

The present investigation had two goals. First, it was designed to determine whether 5- and 6-year-olds could engage in BB reasoning for three and four objects and when a more appropriate measure of such reasoning was used. Second, it was designed to gain greater insight into how—that is, by what underlying cognitive mechanism—children reasoned about the present causal events. We aimed specifically to determine which of two cognitive mechanisms—a Bayesian-inference mechanism or an associative-learning mechanism—best explained children’s causal inferences in the present context.

**Current study**

Five- and 6-year-old children were introduced to a computer-animated machine called the “blicket detector” and were told that their task was to determine which objects activated the machine. They were told that objects that made the machine “go” were “blickets”; objects that did not make the machine go were not blickets. Participants then received either two BB main trials and two BB control trials or two ISO main trials and two ISO control trials. Participants in both conditions were then asked to indicate whether the objects in each trial were blickets. Participants were randomly assigned to the BB or ISO conditions.

**Method**

**Participants.** Participants were 32 5-year-olds (16 boys and 16 girls) and 31 6-year-olds (17 boys and 15 girls). Sample size was determined based on previous studies on BB 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). Although most children were from white, middle-class backgrounds, a range of ethnicities that resembled the diversity in the population were represented. All children were tested in a quiet room at a local children’s museum.

**Materials.** The “device” used in the current study was a computer-animated version of the canonical blicket detector (e.g., Gopnik & Sobel, 2000). The device was a white rectangle with a black border that measured 5.99 cm × 23.47 cm. If the device was “on”, the white region of the rectangle turned blue. If the device was “off”, the white region remained white. A maximum of 4 differently colored circles were used, and each circle measured 2.67 cm × 2.67 cm (see Figure 2 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 (for the BB or ISO experimental trials) or four (for the BB or ISO control trials) 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 science 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 main 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 was and was included to ensure that participants understood the task.

Diagram

Description automatically generated

Figure 5. Schematic of one of the two BB experimental events. The upper-right portion of the figure shows the BB 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.

Following the pretraining phase, participants were given four test trials—either the two BB experimental trials and 2 BB control trials or two ISO experimental trials and 2 ISO control trials—in counterbalanced order using a Latin square. Differently colored objects were used across all trials to prevent carryover effects.

The two BB main 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 (i.e., 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. The first and second BB experimental trials were identical except for the object colors (see Figure 5 for a schematic of the BB experimental event).

The two BB 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. The experimental and control trials used the same text.

Finally, the ISO experimental and control conditions were identical to the BB experimental and control conditions except that objects A (during the ISO main trials) and D (during the ISO control trials) failed to activate the machine (see Table 1 for a schematic).

|  |  |  |  |
| --- | --- | --- | --- |
| Schematic of Experiment 1 | | | |
|  | Compound | Elemental | Test |
| BB experimental trial | ABC+ | A+ | Is A/B/C a blicket? |
| BB control trial | ABC+ | D+ | Is A/B/C/D a blicket? |
| ISO experimental trial | ABC+ | A- | Is A/B/C a blicket? |
| ISO control trial | ABC+ | D- | Is A/B/C/D a blicket? |

Table 1. The +/- signs corresponds to whether the machine activates (+) or not (-)

**Results**

**Chart, bar chart

Description automatically generated**Figure 6. Participants’ mean responses to whether each object was a blicket across the conditions and trial types (i.e., “eventType”). Bars show standard error.

Figure 5 shows the results for this experiment. The dependent measure was the number of times that participants responded “Yes” to the “Is this a blicket” question. Thus, across two trials, the maximum number of times that a participant could respond “Yes” was 2; the minimum number of times that a participant could respond “Yes” was 0. Using this dependent measure, the data were entered into a four-way linear model with Age (5-year-olds vs. 6-year-olds) and Condition (BB vs. ISO) as the between-subjects factors and Trial Type (experimental vs. control) and Objects (A vs. B vs. C vs. D) as the within-subjects factors. This analysis revealed a main effect of Condition, *F*(1, 397) = 9.21, *p* < .005, a main effect of Objects, *F*(3, 397) = 8.78, *p* < .001, and a main effect of Trial Type, *F*(1, 397) = 13.29, *p* < .001.These significant main effects were qualified by a significant two-way interaction between Condition and Objects, *F*(3, 397) = 24.72, *p* < .001 and a significant two-way interaction between Condition and Trial Type, *F*(1, 397) = 5.47, *p* = .02. These two-way interactions were further qualified by a significant three-way interaction between Condition, Objects, and Trial Type, *F*(2, 397) = 21.05, *p* < .001. This three-way interaction is shown in Figure 3. Given that age did not factor significantly into any of the main effects or interactions, we collapsed across age for all subsequent analyses.

We followed up this three-way interaction with separate one-way linear models for the main and control trials within the BB and ISO conditions. The Objects factor was treated as the sole within-subjects factor in these follow-up analyses. The first one-way linear model for the control trials within the BB condition did not reveal a significant effect of Objects, *F*(3, 120) = 0.22, *p* = .88. This means that participants treated the objects similarly during the control trials of the BB condition. These results qualitatively match the predictions of the connectionist model but not the predictions of the Bayesian model. In contrast, the second one-way linear model for the main trials within the BB condition revealed a significant main effect of Objects, *F*(2, 84) = 11.04, *p* < .001. This main effect reflected the fact that participants considered object A to be more of a blicket (*M* = 1.97, *SD* = 0.18) than object B (*M* = 1.21, *SD* = 0.83), *t*(27) = 4.70, *p<* .001, or object C (*M* = 1.32, *SD* = 0.82), *t*(27) = 4.12, *p* < .001. Participants treated objects B and C equivalently, *t*(27) = -0.72, *p* = .48.

The third and fourth one-way linear models for the main and control trials within the ISO condition both revealed a significant main effect of Objects, both *F*’s > 22.76, both *p*’s < .001. This reflected the fact that participants considered object A (*M* = 0.52, *SD* = 0.87) to be significantly less of a blicket than objects B (*M* = 1.70, *SD* = 0.67) and C (*M* = 1.74, *SD* = 0.59) during the main condition, both *t*’s > -5.97, both *p*’s < .0001, and object D (*M* = 0.73, *SD* = 0.72) to be less of a blicket than objects A (*M* = 1.76, *SD* = 0.58), B (*M* = 1.69, *SD* = 0.59), and C (*M* = 1.76, *SD* = 0.50) during the control trials, all *t*’s > -6.01, all *p*’s < .001. Considered together, these data are qualitatively most consistent with the predictions of the connectionist but not the Bayesian model (see Figure 6A-C for a visualization of this fit).

**Assessing BB under the new operationalization of BB reasoning**

To examine whether there was evidence of BB reasoning according to the new operationalization of BB reasoning, data for the redundant causes within the BB experimental and control conditions were entered into a three-way linear model with Age (5-year-olds vs. 6-year-olds) as the between-subjects factor and Objects (A, B, and C) and Trial Type (main vs. control) as the within-subjects factors. If participants engaged in BB reasoning based on this operationalization, then there should a main effect of Trial Type. This is what we found. This result reflected the fact that participants were more likely to respond that a redundant object was a blicket during the control trials (*M* = 1.60, *SD* = 0.71) than during the main trials (*M* = 1.27, *SD* = 0.82), *F*(1, 139) = 5.28, *p* = .02 Follow-up planned comparisons revealed that participants were less likely to respond that object B was a blicket during the BB main trials (*M* = 1.21, *SD* = 0.83) compared to object A during the BB control trials (*M* = 1.61, *SD* = 0.72), *t*(27) = 3.06, *p* = .005. Moreover, participants were less likely to respond that object B was a blicket during the BB main trials compared to object B during the BB control trials (*M* = 1.58, *SD* = 0.72), *t*(27) = 2.17, *p* = .04. Finally, participants were less likely to consider object B to be a blicket during the BB main trials compared to object C during the BB control trials (*M* = 1.61, *SD* = 0.72), *t*(27) = 3.29, *p* = .003. No other differences reached statistical significance. Given the absence of evidence for BB reasoning for the remaining three comparisons (i.e., given that participants did not treat object C during the main trials differently than objects A, B, and C during the control trials), these data only provide minimal evidence of BB reasoning based on the new operationalization.

**Assessing BB under the old operationalization of BB reasoning**

To examine whether there was evidence of BB reasoning according to the old operationalization of BB reasoning, data for the redundant causes between the BB and ISO conditions were entered into a three-way linear model with Condition (BB vs. ISO) as the sole between-subjects factor, and Objects (A, B, and C) and Trial Type (main vs. control) as the within-subjects factors. Evidence for BB reasoning based on this operationalization would be a main effect of Condition (averaging over Objects and Trial Type). This is what we found. Specifically, participants were more likely to call a redundant object a blicket in the ISO condition (*M* = 1.73, *SD* = 0.58) compared to the BB condition (*M* = 1.48, *SD* = 0.77), *F*(1, 277) = 7.60, *p* < .01. Thus, there was stronger evidence for BB reasoning based on an old operationalization of it than based on the new operationalization of it.

|  |  |
| --- | --- |
| **Chart, bar chart  Description automatically generated**  A | |
| Chart, bar chart  Description automatically generated  B | Chart, bar chart  Description automatically generated  C |

Figure 7A-C. Participants data (6A) compared to the predictions of the best-fitting connectionist model trained for 800 epochs (6B) and the best fitting Bayesian model, where the *P*(Blicket) = .80 (6C).

**Model fits**

To assess the quantitative fit of the predictions of the connectionist model to the data and the predictions of the Bayesian model to the data, we computed the root mean square (RMSE) and mean absolute error (MAE) between each model’s predictions and participants’ mean responses to objects A-C during the BB and ISO main trials and objects A-D during the BB and ISO control trials. These two metrics have been used in previous simulation studies to assess model’s quantitative fit to behavioral data (e.g., Bhat et al., 2022). Lower values on each metric indicate better model fit. Table X shows the fits for the different model instantiations.

|  |  |  |
| --- | --- | --- |
| Computational Models | root mean square (RMSE) | mean absolute error (MAE) |
| Connectionist Model (800 epochs)\* | .19 | .15 |
| Connectionist Model (1600 epochs) | .22 | .16 |
| Connectionist Model (2000 epochs) | .25 | .17 |
| Connectionist Model (3000 epochs) | .28 | .22 |
| Bayesian model (.5) | .58 | .54 |
| Bayesian model (.65) | .39 | .34 |
| Bayesian model (.80)\* | .29 | .21 |
| Bayesian model (.95) | .39 | .33 |
| Bayesian model (1) | .45 | .39 |

Table 2. Model fit indices for the various models and instantiations. \* indicates the best fitting connectionist and Bayesian models.

It should be clear from the table above that the connectionist model—regardless of the number of epochs for which it was trained—provided a better quantitative fit to the beahviroal data than any of the Bayesian models that was trained for 800 total epochs provided a better quantitative fit to the behavioral data than did any of the Bayesian models. Given that the connectionist model instantiates associative learning, these results suggest that participants may have used associative learning rather than Bayesian inference to process the present events. Taken together, the connectionist model provided a better qualitative and quantitative fit to the behavioral data than any of the Bayesian model instantiations. We discuss why this might be the case in the General Discussion.

General Discussion

This study had two aims. The first was to determine whether 5- and 6-year-olds would engage in BB reasoning for 3 and 4 objects. This departs from the typical convention of using two objects to study causal reasoning in human children. The second aim was to clarify how exactly (i.e., the cognitive mechanism by which) children reasoned about the present causal events. We were specifically interested in whether children’s causal inferences best conformed to the predictions of a simple Bayesian model or a connectionist (associative learning) model.

With respect to the first aim, we found evidence of BB reasoning under the old operationalization of BB reasoning but only minimal evidence of such reasoning under a new and more valid measure of BB reasoning. This finding extends previous research to show that when children are asked to reason about three objects and a more valid of operationalization of BB reasoning is used, children minimally engage in BB reasoning.

With respect to the second aim, the data were most consistent with the connectionist (associative learning) models. However, a question that we have not yet answered concerns the exact nature of the connectionist models’ associative learning. Specifically, how did these models—via associative learning—arrive at their causal judgements? To understand how the connectionist model’s judgements arose mechanistically, consider the BB (i.e., ABC+ D+) control trial. The rationale for focusing on this condition is that the Bayesian and associative models presented here make qualitatively identical predictions for the other conditions (i.e., the BB experimental and ISO experimental and controls trials). This means that the better qualitative and quantitative fit of the connectionist models to the data relative to the Bayesian models cannot be due to these trial conditions. Rather, the reason the connectionist models better fits the behavioral data than the Bayesian models is because this model uniquely predicts participants performance during the BB control 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. A key feature of these models is that they learn to reduce the difference between such observed and predicted output by adjusting connections between units in the model. In the present case, these connections or “weights” encode the association between each input unit (i.e., each object) and the single output unit (i.e., the machine’s activation). Thus, because the difference between the observed and predicted activation of the single output unit was equivalent for all four objects, the network made the same weight adjustments (both in sign and magnitude) to connections between each object and the single output unit. This explains why the network responded equivalently when “asked” whether each object (i.e., objects A-D) was a blicket. Given that participants’ causal responses mirrored the predictions of the connectionist model, this suggest that children arrived at their causal judgements via a similar associative-learning mechanism. This finding is significant because it has been suggested that causal reasoning in human children is best explained by Bayesian inference and rational processes rather than by associative processes. The present data add context to this debate by suggesting that whether children engage in Bayesian inference or associative learning may depend on how causal reasoning is measured and the number of objects about which children are asked to reason.

Nonetheless, potential criticisms are worth noting. One such criticism is that the results are inconsistent with the findings from previous studies on BB reasoning in human children. Such previous research showed that children unequivocally engage in BB reasoning when asked to reason about two objects; in contrast, the current study only provided equivocal evidence for BB reasoning when three and four objects were used. However, we believe that the present results extend rather than are at odds with such previous research. Specifically, the present study likely demonstrates that when children’s information-processing capacities are stretched such as when they are asked to reason about multiple potential causes, they may deploy and rely on simpler associative processes. Although the numerical difference between three and four objects is miniscule, by contrast the corresponding increase in the size of the underlying psychological hypothesis space is substantial. Such an increase in the size of the underlying psychological hypothesis space may have important ramifications on the cognitive mechanism that gets deployed by children, especially if children are sensitive to and affected by this increase. For example, children who are asked to reason about two candidate causes need only to represent and choose among *four* candidate causal hypotheses. Four candidate causal hypotheses may well be within the information-processing capacities of 5- and 6-year-olds. In contrast, children who are asked to reason about three candidate causes must now consider *eight* candidate causal hypotheses—this may exceed their restricted information-processing capacities.

It turns out that there is a wealth of data that is consistent with this general proposal (Doebel & Zelazo, 2015; Frye, Zelazo, & Palfai, 1995; Zelazo, Frye, & Rapus, 1996; Zelazo et al., 2003). One recent study by Kenderla and Kibbe (2023) showed that when 8- and 10-year-old children’s information-processing abilities were stretched in a virtual memory game—such as when children were asked to find three cards that shared one feature and differed on another feature—they relied less on working memory and more on manual exploration. 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 working memory. In a similar vein, Richland, Morrison, and Holyoak (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. Finally, 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).

Together, this research demonstrates 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. This may provide a developmental explanation for why children in the present study did not engage in BB reasoning or show evidence that they relied on Bayesian inference. 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. This issue should be explored more fully in future research.

A second potential criticism is that we cannot be sure that a simple Bayesian-inference mechanism underpinned participants’ performance in the present study. For example, if participants assumed that blickets were common in the present context—which is plausible given how frequently the detector activated in the present study—then participants should be *less* likely to block redundant causes; in other words, participants should be *more* likely to treat all potential blickets (expect for the ones that are explicitly shown not to be causal) equally. This could explain participants’ performance in the BB control condition—in that condition, participants treated all objects equally. However, this explanation cannot explain *all* the present data. This is because this explanation predicts that participants should have also treated objects A-C equivalently in the BB experimental condition as well, but this was not the case: Participants treated object A differently than either objects B or C in the BB experimental condition. This explanation is also unlikely given that overall the Bayesian models provided a poorer fit to the behavioral data than the connectionist models. This would not be expected if participants relied on Bayesian inference. Nonetheless, because we did not systematically manipulate base-rate information, this alternative explanation cannot be ruled out entirely. However, if we are correct that participants do not rely on Bayesian inference when asked to reason about multiple causes, we predict that their performance in this proposed future study would not differ from participants’ performance in the current study. However, if children’s causal judgements are affected by base-rate information, such that how they process BB event changes with changes to base-rate information, then this would suggest that participants may use Bayesian inference after all to reason about multiple candidate cause, at least when a Bayesian-inference mechanism is primed by explicitly and systematically manipulating base-rate information.

**Conclusion**

These potential criticisms notwithstanding, these experiments constitute one of the first systematic attempts to examine BB and ISO reasoning in human children in the context of three objects. A longstanding view has been that the cognitive mechanism by which human beings reason about causal events is Bayesian inference (e.g., Gopnik et al., 2004) rather than associative processes. The experiments reported here support a different conclusion: an associative-learning counting mechanism supports 5- to 6-year-old children’s reasoning about multiple potential causes.

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.

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

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.

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

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.

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.

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

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.

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.

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.

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart J. L. McClelland, & the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

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., & Munro, S. (2006, August). When Mr. Blicket wants it, children are Bayesian. In *Proceedings of the Cognitive Science Society* (pp. 810-816).

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.

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

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.

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.