Don’t throw the associative baby out with the Bayesian bathwater: Children’s retrospective reasoning about multiple causes suggests multiple systems for causal inference

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

Conflicts of interests: none

Don’t throw the associative baby out with the Bayesian bathwater: Children’s retrospective reasoning about multiple causes suggests multiple systems for causal inference

Submitted to *Developmental Science* on DATE

Abstract

Causal reasoning is a fundamental cognitive ability that enables human beings to learn about the complex interactions in the world. The mechanisms that underpin 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; causal mechanisms; computational models; analytical models; associative learning; Bayesian inference

Don’t throw the associative baby out with the Bayesian bathwater: Children’s retrospective reasoning about multiple causes suggests multiple systems for causal inference

Causal reasoning – the capacity to make inferences about cause and effect relations – is central for learning about the world. Causal reasoning enables human learners to make predictions and inferences (e.g., Bullock, et al., 1982; Leslie & Keeble, 1987; Oakes & Cohen, 1990; Shultz, 1982), 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., 2014; 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 reasoning is governed by a complex, rational inferential system that is in place early in development. This view has been articulated as probabilistic causal models (e.g., Glymour, 2001; Pearl, 2000) describing how children represent causal knowledge (e.g., Gopnik et al., 2004), and algorithms based on Bayesian inference (e.g., Bonawitz et al., 2014; Griffiths & Tenenbaum, 2005) describing how children make causal inferences (e.g., Gopnik & Wellman, 2012; Griffiths & Tenenbaum, 2007; Xu, 2019). On this view, children’s causal reasoning starts with statistical learning capacities present in infancy (e.g., Gomez, 2002; Johnson et al., 2006; Kirkham et al., 2002; Marcus et al., 1999; Saffran et al., 1996), but develops into a system that infers abstract patterns of coherent causal structure from probabilistic data.

An alternative hypothesis is that the same starting points – statistical learning and associative reasoning – sufficiently explain children’s causal reasoning (an idea initially suggested by Piaget, 1930). Connectionist accounts of causal reasoning (e.g., Rogers & McClelland, 2004), 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) suggest that associative reasoning is 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), a retrospective inference in which learners reevaluate the causal status of an ambiguous event based on learning more about the status of other events (see also De Houwer et al, 2002; Larkin et al, 1998; Kruschke & Blair, 2000; Lovibond et al., 2003; Van Hamme & Wasserman, 1994, for other work on adults). Sobel et al. (2004) introduced 3- and 4-year-olds to a machine that lit up and played music when certain objects were placed on it (a blicket detector, Gopnik & Sobel, 2000). Children were then shown two objects (A and B). Together, A and B made the machine go, then object A was placed on the machine by itself, which also made the machine activate. Children were asked about the efficacy of both objects. Three- and 4-year-olds both stated that object B was less likely to be efficacious on these trials than on trials in which object A did not make the machine go by itself (a condition referred to as *indirect screening off*)[[1]](#footnote-1). Using an anticipatory eye-tracking procedure, Sobel and Kirkham (2006) found that 8-month-olds showed this same response pattern.

That children engage in backwards blocking inferences have been problematic for accounts that suggest children’s causal reasoning is described by classic associative mechanisms (e.g., Rescorla & Wagner, 1972) because the associative strength between object B and the machine is the same between the backwards blocking trials (where A is efficacious on its own) and the indirect screening off trials (where A is not). Both Beckers et al. (2005) and McCormack et al. (2009), however, argue that this difference cannot necessarily be explained by children retrospectively reevaluating the efficacy of object B based on object A’s efficacy. Instead, children might make a logical inference that only A does so when A activates the machine, given there is no baseline condition for simply observing subsequent positive efficacious events individually. McCormack et al. (2009) introduced children to a similar machine and showed them both a backwards blocking sequence (A and B together make the machine activate, then A makes the machine activate by itself with a third object, C, on the table, which is never placed on the detector) or a control sequence (A and B together make the machine activate, then C makes the machine activate by itself). The 4-year-olds they investigated did not discern between the efficacy of B in the backwards blocking sequence and the efficacy of B in the control sequence, although older children (5-year-olds) did so. The critical point of agreement between these investigations is that at some point in development, children have the capacity to make retrospective inferences. But the fundamental question remains as to how children engage in such reasoning.

**Bayesian vs. Associative Reasoning**

Sobel et al. (2004) also demonstrated that changing the base rate of causally efficacious objects affected children’s backwards blocking inference. Four-year-olds observed that only two out of the first twelve objects they saw activated the machine, a backwards blocking inference about the 13th and 14th objects they observed on the machine was different than if they saw 10 out of the first twelve objects activate the machine. Children discounted the activation of object B with object A on the machine when few objects activated the machine, but did not when many objects did so. Three-year-olds did not show this inference, but they did if they understood that the objects held a deterministic relation to the machine’s activation because of a nonobvious property, regardless of the causal domain in which children were tested (Sobel & Munro, 2009).

These data were fit to a Bayesian model that these and other researchers have used to describe children’s causal inference (e.g., Gopnik & Wellman, 2012; Griffiths et al., 2011; Tenenbaum & Griffiths, 2001). For example, Griffiths et al. (2011) showed that both children and adults reasoned about backwards blocking with base rates according to a Bayesian algorithm, but that algorithm made predictions about other kinds of causal inferences. Particularly, both adults and 4-year-olds inferred that when three objects (A, B, and C) were placed on the machine in combinations (A and B activated it, and then A and C activated it), when efficacious objects were rare, object A was most likely to be efficacious, but not at ceiling levels, and B and C would have equal likelihood at the end of the sequence, greater than baseline, but lower than A.

However, one could also consider the evidence in favor of Bayesian inference as indicating that children make higher-older correlations among events. Benton et al. (2021) suggested that children can make causal inferences based on second-order correlations – associations of associations (see also Cuevas et al., 2006; Rakison & Benton, 2019; Yermolayeva & Rakison, 2016), providing a proof of concept that the same data in support of Bayesian models as a description of children’s causal inferences could be explained by higher-order statistical learning models, such as connectionist networks. McClelland and Thompson (2007) presented a connectionist account of the initial findings on backwards blocking on children; we suspect that deep learning architectures could be adopted to explain findings about such inferences that also took the base rates information of the candidate cause into account (although to our knowledge, no such model has been published).

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 – when shown that two pairs of compound stimuli (A and B, and then A and C) were efficacious, the 4-year-olds they investigated categorized A as efficacious more often than B or C, and less so than ceiling, but not differently from individual objects presented as a single compound (X and Y that together activated the machine). That is, children did not judge the likelihood that object A was efficacious as different from the efficacy of objects X and Y. At question is whether having to reason about more than two objects produced information processing demands that caused children to be less likely to use a rational Bayesian mechanism and more likely to look more associative in their inferences.

The presence of such information processing demands affecting children’s inferences might be surprising. SENTENCE ABOUT INFORMATION PROCESSING IN INFANT PERCEPTION OF CAUSALITY. Similarly, although Sobel and Kirkham (2006) found that 8-month-olds engaged in backwards blocking inferences like preschoolers, 5-month-olds’ inferences on the same measure looked more associative in nature (Sobel & Kirkham, 2007). When infants make inferences about the reliability of others’ information, their judgments appear more associative in nature (Tummeltshammer et al., 2014). As children enter the preschool years, those judgments become more based in rational inferences, although they still show default to certain kinds of associative inferences (Hermes et al., 2018; Luchkina et al., 2020). Relevant to the present investigation, Sobel et al. (2017; see also Erb & Sobel, 2014) showed that between the ages of 4-7, children develop the capacity to form larger hypothesis spaces of the potential causes they might need to hold to engage in more rational inferences. Because of this finding, we considered 5- and 6-year-olds in the present study.

Here, we asked children to make retrospective causal inference about three potential causes as opposed to only two, using a logic like that of McCormack et al. (2009). In Experiment 1, children observed three objects (A, B, and C) together have causal efficacy, and then A by itself either have or fail to have that same efficacy. They were asked whether each object was efficacious. These trials were compared with control trials in which they again observed three objects (A’, B’, and C’) have efficacy together, and then a fourth object (D) either have or fail to have that efficacy. A retrospective inference involves the judgments of objects B and C being different across these two types of trials. When A is efficacious, judgments of the efficacy of B and C should be lower than the judgments of B’ and C’ when children see that a fourth, unrelated object is efficacious. When A is not efficacious, judgments of the efficacy of B and C should be higher than B’ and C’ when children see that a fourth, unrelated object is not efficacious. Moreover, judgments of B and C should differ between these two trials; B and C are more likely to be judged as efficacious when A is not efficacious than when A is. 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, we present a pair of computational models that try to explain these results, followed by a second experiment that considers a prediction that emerges from the comparison of these models.

**Experiment 1**

**Method**

***Participants***

Participants were 32 5-year-olds (16 boys and 16 girls, *Mage* = 64.81 months, SD = 3.48) and 32 6-year-olds (17 boys and 15 girls, *Mage* = 77.81 months, SD = 3.78). 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 additional children were tested, but not included in the 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***

We presented children with a computer-animated version of the blicket detector (e.g., Gopnik & Sobel, 2000). On a computer screen, the machine 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 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 (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 ensured that participants understood the task and recognize that individual objects could activate the machine and that the machine activated if at least one efficacious 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 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.

**Backwards Blocking Trials and Control.** The two BB main trials began with three differently colored objects, which were located above the machine. The experimenter read the text on the screen, “Look, I have these three toys. Let’s find the blickets. Watch what happens.” All three objects (i.e., objects A, B, and C) then descended until they rested on top of the machine, which activated by lighting up blue and making a sound. At this point, text appeared on the screen, which the experimenter read: “Look, these also make the machine go!” The objects then returned to their starting positions.

Then the left- or right-most (counterbalanced) object (which we will refer to here as object A) descended until it contacted and immediately activated the machine. The experimenter read the text on the screen, “Look, this one makes the machine go!” 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 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. Again, children received two trials, which were identical except for the color of the objects.

**Indirect Screening Off Trials and Control.** The procedure for children who received the indirect screening off trials and controls was identical to the backwards blocking trials, except that when the A object (Experimental trial) or the D object (Control trial) was placed on the machine by itself, the machine did not activate.

Table 1. Schematic of Experiment. Children receive two trials of BB Experimental and Control or ISO Experimental and Control

|  |  |  |  |
| --- | --- | --- | --- |
| Schematic of Experiment | | | |
|  | 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**

Figure 2 shows the mean percentages of responses to the “Is it a blicket?” question for each object at the end of each type of trial. To account for the within-subject nature of this question, we analyzed the frequency with which children stated each object was a blicket across the trial types and conditions with a Generalized Estimating Equation assuming a robust correlation matrix. Condition (Backwards Blocking vs. Indirect Screening Off), Trial Type (Experimental vs. Control), Object (A, B, C, and D), and age (in months) were fixed factors.

Figure 2. Participants’ mean responses to whether each object was categorized as a blicket across the conditions and trial types. Bars show standard error.**Chart, bar chart

Description automatically generated**

This model revealed a no significant main effects, but two significant interactions: a significant two-way interaction between object and condition, Wald 2(2) = 8.05, *p =* .05, and a significant three-way interaction between object, condition, and age, Wald 2(3) = 9.43, *p =* .02. In particular, parameter estimates of this model suggest that as children got older, they treated objects B and C differently than object A between the two conditions, B = 0.04, SE = 0.005, 95% CI [0.003, 0.02], Wald2(1) = 6.83, *p =* .009, for object B, and B = 0.02, SE = 0.01, 95% CI [-0.003, 0.04], Wald2(1) = 3.03, *p =* .08 for object C.

To analyze these interactions, we divided children by condition and age group. In the Backwards Blocking condition, both the 5-year-olds (60% of the time vs. 85%) and 6-year-olds (56% vs. 75%) categorized the B and C objects as less likely to be blickets in the experimental trials than the control trials, *t*(14) = 2.42 and *t*(15) = 2.42 respectfully, both *p*-values = .03. In contrast, in the Indirect Screening Off condition, the 5-year-olds did not differ in their average judgment of the efficacy of objects B and C in the experimental trials (80%) and the control trials (85%), *t*(15) = 1.57, *p* = .14, while the 6-year-olds showed a larger difference (77% vs. 86%), but only at a non-significant trend, *t*(16) = 1.86, *p* = .08.

Finally, we considered performance between the trials. Children treated object A differently in the experimental trials of the Backwards Blocking and Indirect Screening Off conditions, judging A as a blicket on 98% of the trials in the former and 26% of the trials on the latter, *t*(62) = 9.11, *p* < .001. Their judgments of objects B and C also differed between these conditions, with them stating these objects were blickets on 58% of the backwards blocking trials and 77% of the indirect screening off trials, *t*(62) = -2.23, *p* = .03.

**Discussion**

Experiment 1 showed evidence that children retrospectively reevaluated the causal efficacy of ambiguous data, although the evidence suggesting that they do so was not as clear as in previous investigations. When shown ambiguous evidence that three objects together have efficacy, and then that one of those objects alone has the same efficacy (backwards blocking), children were less likely to judge the other two objects as having efficacy compared to a control condition in which a fourth, independent object had such efficacy and compared to the case in which that object individually did not have efficacy (indirect screening off). However, in that latter case, the judgment that the other two objects did have efficacy did not differ compared to the same type of control condition, particularly for the younger children we tested.

**Model Fits**

We fit two competing computational models to these data. The first was a model based on Bayesian inference, described initially by Sobel et al. (2004) and in more detail in Griffiths et al. (2011). The second was a connectionist model following work by Rogers and McClelland (2014). The computational details of both models are described in the Supplemental Materials. Below, we briefly describe each model.

**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 hypothesis a learner has 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*).

The initial hypothesis space relies on assuming 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 told that the machine in front of them detects blickets, which is defined by the object having the efficacy to activate the detector, 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 about this form of structure learning). Griffiths et al. (2011) describe the formal parameterization of this hypothesis space and model that results in the hypothesis space shown in Figure 2. This figure represents the eight possible hypotheses reasoners might have about the experimental trials.

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 **. 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*) = ** Hypotheses 1, 2, and 4 posit exactly 2 blickets, so their *p*(*h*) = ****Hypotheses 3, 5, and 6 each posit 1, making their *p*(*h*) = ****Finally, Hypothesis 7 posits no blickets, making its *p*(*h*) = (1-**

Figure 2. Possible hypothesis for the experimental trials of the backwards blocking and indirect screening off conditions in which Objects A-C could each activate the detector (E). Note that this hypothesis space is similar for the control trials, but must include a fourth D object.

**Timeline

Description automatically generated with medium confidence**

Because the model assumes that objects with causal efficacy will act deterministically on detectors, 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 posit the causal relation between the object and the detector. 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.

Given the pretest, in which children observe two objects, one of which activates the machine and one which does not, it is possible to assume **= 0.5 as a starting point for the model. Table 2 presents the probability that each object is a blicket given the data presented to them in each type of trial. When 5- and 6-year-olds are considered as separate ages, this model fits the data the level of *r* = .86, with a Bayes Factor of 0.00000028, suggesting that the model is strongly related to these data.

*Table 2. Results of Bayesian Model, assuming* **= 0.5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Object A | Object B | Object C | Object D |
| Backwards Blocking Experimental | 1 | .5 | .5 | N.A. |
| Backwards Blocking Control | .57 | .57 | .57 | 1 |
| Indirect Screening Off Experimental | 0 | .67 | .67 | N.A. |
| Indirect Screening Off Control | .57 | .57 | .57 | 0 |

**Connectionist Network. DEON TO ADD**

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.

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

B)

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

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; 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 use a more valid measure of BB reasoning to determine whether 5- and 6-year-olds could engage in this form of reasoning for three and four objects. Second, it was designed to illuminate how—that is, by what cognitive mechanism—children reasoned about the present causal events. Our specific aim was to determine which of two cognitive mechanisms—Bayesian inference or associative learning—best explained children’s causal inferences in the present context.

**Current study**

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

Figure 5A-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 and Bayesian models 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 2 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 behavioral data than 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. Together, the connectionist model provided a better qualitative and quantitative fit to the behavioral data than any of the Bayesian model instantiations. We discuss the exact nature of this associative learning 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 connectionist 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. 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 the connections (which encode associations) 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 process. 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 nuance and 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 they are asked to reason.

These aims aside, some potential criticisms are worth noting. One such potential 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, over all 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, overall, 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, this study constitute one of the first systematic attempts to examine BB and ISO reasoning in human children in the context of three and four 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: associative learning is sufficient to explain 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.

1. Four-year-olds showed this response pattern more than 3-year-olds, a point we will return to subsequently. [↑](#footnote-ref-1)