Title

"Dancing on the ceiling": The role of different forms of thinking on retrospective reevaluation in children

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Beaton, Rebecca
Benton, Deon T.

Publication Date

2024

Peer reviewed

"Dancing on the ceiling": The role of different forms of thinking on retrospective reevaluation in children

Rebecca M. Beaton (rebecca.m.beaton@vanderbilt.edu)
Department of Psychology and Human Development, 230 Appleton Place
Nashville, TN 37203 USA

Deon T. Benton (deon.benton@vanderbilt.edu)
Department of Psychology and Human Development, 230 Appleton Place
Nashville, TN 37203 USA

Abstract

An open question in the developmental causal learning literature concerns how children's beliefs about causal systems impact their inferences. This study investigated how 4- and 5-year-olds' causal beliefs related to their "backwards blocking" abilities, as well as whether associative learning or Bayesian inference better explained their judgements. Children were taught either that two causes together produced a larger effect than that produced by each individually or that they produced the same size effect as that produced by either one. A third group received no training. Results indicated that 4-year-olds engaged in backwards blocking only after additivity training and that their inferences mainly matched an associative model. In contrast, 5-year-olds consistently engaged in backwards blocking and produced responses that largely matched a Bayesian model. These findings suggest that the effect of children's beliefs about causal systems on their inferences undergoes a developmental progression and implicate the role of multiple cognitive mechanisms.

Keywords: causal reasoning; developmental learning mechanisms; computational modeling; associative learning; Bayesian inference

Introduction

Causal reasoning is essential for understanding how the world works. For example, it enables individuals to make inferences (e.g., Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982), to intervene to generate novel effects (e.g., Butler, Gibbs, & Tavassolie, 2020; Schulz, Gopnik, & Glymour, 2007), and to reason counterfactually (e.g., Harris, German, & Mills, 1996; Walker & Nyhout, 2020). There is now consensus among researchers that causal reasoning emerges by and undergoes considerable development between 18 months and 5 years of age (cf., Sobel & Kirkham, 2006).

Of these findings, the one that has received the most empirical attention is the finding that children can engage in different forms of retrospective reevaluation or cue competition effects such as *backwards blocking reasoning*. Backwards blocking refers to the tendency to, after observing that two (or more objects) together produce some effect and then that one of the objects produces the effect alone, discount the first object as a cause. In one of the first studies on this topic, Sobel, Tenenbaum, and Gopnik (2004) introduced children to a simple causal system in which a machine called a "blicket detector" lights up and plays music only when objects called "blickets" are placed on it. Children were then shown that two competing cues or objects, A and B, activated the machine when they were placed on it together. Object A was then shown either to activate (the backwards blocking trial) or not activate (the indirect screening-off trial) the machine on its own. Following both demonstrations, children were asked whether objects A and B individually were blickets. Sobel et al. (2004) found that how children treated object B depended on whether A activated the machine by itself: Children considered B to be a better blicket candidate when object A failed to activate the machine. That children treated object B differently between the backwards blocking and indirect screening-off trials was interpreted as evidence of backwards blocking.

One criticism of this conclusion is that it is impossible to know whether children treated object B differently between both conditions because of backwards blocking reasoning, indirect screening-off reasoning (another cue competition effect), or both. It turns out that when a better measure is used—such as comparing children's treatment of object B in the backwards blocking trial to that in a control trial in which children first see that objects A and B together activate the machine (like in Sobel et al., 2004) but then see that an unrelated object C activates the machine—5- and 6-year-olds, but not 4-year-olds, engaged in backwards blocking. Considered together, this research indicates that how backwards blocking reasoning is measured determines when children engage in it.

An open question concerns what other factors impact children's ability to engage in backwards blocking reasoning. One factor that has been established to impact whether and by how much individuals engage in backwards blocking is their beliefs about how a causal system works. For example, participants who assume that two causes combine to produce a larger effect than that produced by either of the causes alone (i.e., outcome additivity) might be expected to discount or block object B as a cause after observing that object A alone produced the same size effect as objects A and B together.

Although research with adults has shown that they are more likely to engage in backwards blocking when they are taught that the outcome is additive than when it is maximal such that there is no difference in effect when one or multiple causes are present (e.g., Beckers et al., 2005; De Houwer &

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Beckers, 2002; Lovibond et al., 2003), only one study by Simms, McCormack, and Beckers (2012) examined this in children. Four- to 5-year-olds and 6- to 7-year-olds were introduced to a toy robot and were told that its "tummy" lights up and makes noise when it consumes certain foods. Participants were then randomly assigned to one of two conditions: Additive or Nonadditive. In the Additive condition, children were shown that some combination of foods together made "a bit" of the robot's tummy light up with a noise (i.e., a smaller effect), whereas other combinations of foods made "all of" the robot's tummy light up and generate a bigger noise (i.e., a larger effect). In the Nonadditive condition, children were shown that combinations of foods that contained at least one food item that was efficacious caused the robot's tummy to light up; there was no distinction between pairs of foods that caused the robot's tummy to have a small or large effect. Children in both conditions then were shown backwards blocking Experimental and Control events, but with new food items. In the Experimental trials, children were first shown that foods A and B caused the robot's tummy to light up and make a noise and then that food A caused the robot's tummy to activate to the same extent as A and B together. In the Control trial, children were shown that foods C and D activated the robot's tummy to the same degree as A and B together, and then that food E failed to activate the robot's tummy. Participants were then asked whether it was food B or food C (the two "compound" objects—that is, the objects that were only ever presented in combination with A or D) that activated the robot's tummy. They found that backwards blocking increased with age and that the older children blocked (i.e., they were more likely to say that food C activated the robot's tummy than food B) in the Additive but not in the Nonadditive condition.

That the backwards blocking effect after additivity training compared to after nonadditivity training was stronger for the 6- to 7-year-olds is theoretically interesting because it challenges the idea that children's causal judgements can be captured by certain rudimentary associative models such as the Rescorla-Wagner (RW) model (Rescorla & Wagner, 1972). Models such as this predict that participants should treat objects B and C equivalently between the Experimental and Control test trials (though see Haselgrove, 2010, for associative accounts that potentially could explain these findings). As such, Simms et al. (2012) and others (e.g., Griffith et al., 2011; Sobel et al., 2004) have interpreted these findings to mean that children's causal inferences are best captured by effortful, rational processes that approximate Bayesian inference. The problem with using the study by Simms et al. (2012) to argue that rational processes underlie children's causal inferences is that the study itself is limited in several ways. First, most studies on backwards blocking in children have used the blicket detector paradigm. Given that Simms et al. (2012) used a different paradigm, it remains unknown whether the difference in the performance of children in Simms et al. (2012) and that of children in Sobel et al. (2004) was due to additivity training per se, to different paradigms, or to some combination of both.

Second, participants were grouped into either an Additive or Nonadditive condition. This means that there was no condition in which participants did not receive any training (i.e., a Control). For example, in the absence of training, do children assume that two causes create a larger effect together (i.e., outcome additivity), or do they assume that an observed effect is maximal and does not differ based on the number of causes present (i.e., outcome maximality)?

Third, the conclusion that data from this study supports a rational account rather than an associative one is speculative rather than quantitative; these researchers did not fit computational models that implement associative learning or Bayesian inference to children's causal judgements to determine which one provides a better account of children's inferences. By fitting quantitative models to participants' causal judgments, it is also possible to determine whether Bayesian inference and associative learning better capture different "aspects" of the same data. This could suggest the presence of multiple mechanisms.

The present study was designed to address these limitations and had four goals. First, we examined the effect of different types of training (e.g., Additive vs. Ceiling) on children's tendency to engage in backwards blocking. It is worth noting that the current Ceiling condition was equivalent to the Nonadditive condition in Simms et al. (2012). Second, we compared the causal judgments of participants who did not receive any training (the Control group) to participants who did receive training (Additive vs. Ceiling). This enabled us to determine which "mode" of thinking (i.e., additive or nonadditive) children defaulted to in the absence of any training. Third, we tested participants using the blicket detector paradigm. Fourth, we fit an associative model and a Bayesian model to participants' data to determine which provides a better account of participants' causal inferences.

Four- and 5-year-old children were introduced to the blicket machine and told how it worked. They were then assigned to one of three conditions: Additive, Ceiling or Control. By testing 4- and 5-year-olds, it is possible to determine whether the failure of the 4- and 5-year-olds in Simms et al. (2012) was due to the fact that children at this age are immune to additivity training or to methodological issues with the study. We also analyzed the 4- and 5-year-olds separately to determine whether there is a distinction between these two ages; Simms et al. (2012) merely grouped these ages together.

We predicted that if participants engaged in backwards blocking reasoning, then they should be less likely to say that the compound object(s) is a blicket following the Backwards Blocking (i.e., Experimental) test trial than following the Control test trial. Crucially, if this effect was moderated by the kind of training that participants received, then the backwards blocking effect should be larger in the Additive condition than in the Ceiling condition. If this effect is additionally moderated by age, we might expect the effect to differ between the two age groups.

Experiment

Method

Subjects. Participants were 26 4-year-olds (9 boys and 17 girls; M = 54.40 months) and 33 5-year-olds (15 boys and 18 girls; M = 65.32 months). This number of participants mirrored that tested in previous work on this topic. Participants were either tested in a quiet room in the lab or at a local children's science center. Participants tested in the lab were recruited from a shared registry.

Materials. This experiment took place on a 13" Dell laptop. The stimuli used in the experiment were created using Macromedia Director MX 2004 and played through VLC Media Player. Experimenters were instructed to read from a script, which was built into the videos. Experimenters were further instructed to pause the videos whenever they asked questions. This ensured that participants were given enough time to answer. The videos consisted of a computer-animated "blicket detector" and several "toys" (e.g., Benton et al., 2023). The machine consisted of a beige rectangular base with a black outline that was approximately 17.5 cm wide x 5 cm tall. This rectangular base was adjacent to a rectangular "meter", which measured 2.5 cm wide x 10 cm tall. Two to 4 circles (the number being dependent on trial type) were located above the rectangular base. The circles were different colors, equally spaced apart, and measured 3 cm in diameter. The colors that were used were burgundy, olive green, gray, teal, orange-yellow, light-pink, brown, yellow, black, red, green, purple, and orange. Different colors were used to ensure that participants' responses were not based on color. Whenever an object that was predetermined to be a blicket descended onto the machine, the machine activated by gradually changing from a flurry of rainbow-like colors to dark blue. The rectangular meter—which was adjacent to the rectangular base—simultaneously "filled up" by also changing from a flurry of rainbow-like colors to dark blue.

Procedure. During the experiment, children were seated at a table next to their caregiver and the experimenter. Participants were randomly assigned to one of three conditions: Control, Additive, or Ceiling. Participants in the Additive and Ceiling conditions experienced the following sequence of events: training, manipulation check, and test. Participants in the Control condition did not experience the training or manipulation check trials. Instead, these children were shown a separate video prior to the test trial that was unrelated but was similar in length to the amount of time that it took participants in the other conditions to complete the training and manipulation check.

Training Trials. Participants in the Additive and Ceiling conditions were presented with compound-effect and element-effect training (see Table 1 for an example of what a participant viewed during the entire training trial). The order in which these trials were presented was counterbalanced across participants, and each trial was shown twice.

During the element-effect training, participants viewed a video that began with red, green, purple, and orange circles (i.e., the "toys") positioned above the machine. Each toy (A, B, C, and D) independently descended from its starting position until it contacted the machine. Two of the objects activated the machine and were called blickets. The remaining two objects failed to activate the machine and were not called blickets. The position and causal efficacy of the objects were counterbalanced across participants in the Additive and Ceiling conditions.

The same four toys were presented during the compound-effect training. Participants observed what happened to the machine when two blickets were placed on it, when one blicket and one non-blicket were placed on it, and when two non-blickets were placed on it. The objects maintained their efficacy from the element-effect trial.

In the Additive condition, when the two blickets descended onto the machine, the machine lit up, and the adjacent meter filled up completely. The words, "Look, these two make the machine go a lot because they are both blickets" also appeared. However, if only one of the two objects was a blicket, the machine continued to light up by changing colors, but the meter became only half-filled with color. The words, "Look, these two make the machine go because only one of them is a blicket" appeared on the screen. Finally, if neither object was a blicket, neither the machine nor the meter activated, and the words, "Look, these two do not make the machine go because they are both not blickets" appeared. This manipulation instantiated an "additivity rule".

In the Ceiling condition, the meter was never more than halfway full. This means that the meter filled up to the halfway mark both when one blicket and one non-blicket descended onto it, as well as when two blickets descended onto it. The words "make the machine go" were only used in this condition (compared to the words "make the machine go a lot" in the Additive condition). This manipulation instantiated a "ceiling rule".

Table 1: Schematic of the task structure for each condition.

Condition	Training Trial	Events (each 2x)
Additive	Element Compound	A+ / B- / C+ / D- BD- / AB+ / AC++
Ceiling	Element Compound	A+ / B- / C+ / D- BD- / AB+ / AC+
Control	Filler Task	Not Applicable

Note. - = no response; + = smaller effect; ++ = larger effect.

Manipulation check trial. Following the Additive or Ceiling manipulations, participants were given a manipulation check. The purpose of the manipulation check trial was to determine whether children learned the rules they were taught. Four new circles (i.e., "toys") were used during the manipulation check; two were blickets and two were not.

Participants were then asked three types of questions. First, to ensure the participants encoded each object's label, they were asked whether each object was a blicket. Second, the participants were asked what the machine would do if two of the toys were put on the machine. Third, participants were asked how far up the adjacent meter would go when two of the toys were on the machine. This last question allowed us to determine whether participants had encoded the additivity or ceiling rules. One pair of objects consisted of two blickets; the other pair consisted of one blicket and one non-blicket; and the final pair consisted of two non-blickets.

Training for the Control Group. Participants in the Control group were shown an unrelated video after the pretraining trial but before the test trials that consisted of five different shapes (i.e., a star, cloud, lightning bolt, triangle, heart) that moved from the bottom left of the screen to the top right, at which point they were then asked to name the shape. The rationale for including this unrelated activity for participants in the Control condition was to ensure that differences in performance between participants in the Control condition and those in either the Additive or Ceiling conditions were not due to the fact that participants in the Control condition spent less time in the study.

Test trial. Following training, all participants were given two Backwards Blocking test trials and two Control test trials. The Backwards Blocking or Experimental test trial began with two new objects (i.e., an olive-green circle and a maroon circle) above the machine. The two toys then simultaneously descended until they contacted and subsequently caused the machine to activate and the meter to fill up to the halfway mark. Both objects then returned to their original positions, and the first object then descended by itself until it contacted and subsequently activated the machine and caused the meter to fill up to the halfway mark. Participants were then asked whether each toy was a blicket.

Participants were also shown a Control test trial which consisted of three new objects (i.e., a gray circle, a teal circle, and a yellow-orange circle) that were positioned above the rectangular base. The first two objects then simultaneously descended until they contacted and subsequently caused the machine to activate as well as the meter to fill up to the halfway mark. Both objects then returned to their starting positions. The third object then descended until it contacted and caused the machine to activate and the meter to fill up to the halfway mark. Participants were then asked whether each of the three objects was a blicket. The presentation order of Experimental and Control test trials was randomized across participants, and each participant was shown each test trial twice. Note that prior to the training and test trials, all participants completed a pretraining trial to determine that they could reason and answer questions about blickets as well as knew what a blicket was prior to the manipulation or test.

Results

The main analysis examined whether the 4- and 5-year-olds separately engaged in backwards blocking reasoning. In a supplementary analysis, we examined whether children's responses during the test trials differed by condition. Given that the main analysis focused on children's backwards blocking performance, we only considered participants' treatment of the objects that were never put on the machine by themselves (i.e., the compound objects); that is, we only considered children's yes/no responses to object B in the Experimental test trial and the average of their yes/no responses to objects A and B in the Control test trial. The rationale for this was twofold. First, this is the approach taken in previous work on backwards blocking in children (e.g., McCormack et al., 2009). Second, the backwards blocking effect is defined as greater treatment of the compound objects in the Control than in the Experimental test trial.

To examine this question in the 4-year-olds (see Figure 1), we used a mixed-effects logistic regression model, with condition (Additive, Ceiling, or Control) as the between-subjects effect, test trial (Experimental or Control) as the within-subjects effect, and participant as the random effect. This analysis revealed a significant main effect of test trial, $\chi^2(1) = 4.91$, $p = .03$ which was qualified by a significant interaction between test trial and condition, $\chi^2(2) = 8.14$, $p = .02$. Follow-up mixed-effects logistic regression models for each condition, with test trial as the only fixed effect and participant as the random effect, revealed that participants engaged in backwards blocking in the Additive condition, $\chi^2(1) = 7.22$, $p = .01$, but neither in the Ceiling condition, $\chi^2(1) = 2.71$, $p = .10$, nor in the Control condition, $\chi^2(1) = 0.31$, $p = .58$: 4-year-olds in the Additive condition were more likely to choose the compound objects as blickets in the Control test trial than the compound object in the Experimental test trial, odds ratio = 62.03, 95% CI [6.67, 18732.11], $p < .01$.

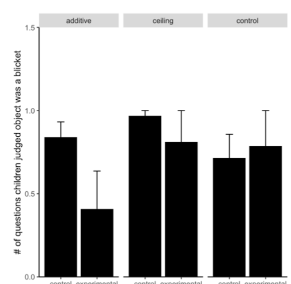

Figure 1: The frequency of 4-year-olds' responses to the compound objects in the Control and Experimental test trials in each condition separately.

To examine the same set of questions in the 5-year-olds (see Figure 2), we used a mixed-effects logistic regression model, with condition (Additive, Ceiling, or Control) as the between-subjects effect, test trial (Experimental or Control) as the within-subjects effect, and participant as the random effect. This analysis revealed only a significant main effect of test trial, $\chi^2(1) = 25.24$, $p < .001$. This result reflected the fact that 5-year-olds were more likely to choose the compound objects as blickets in the Control test trial than the compound object in the Experimental test trial, irrespective of the condition to which they were assigned, odds ratio = 7.78, 95% CI [2.15, 35.34], $p < .005$.

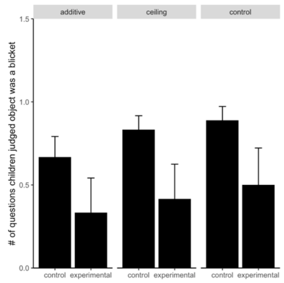

Figure 2: The frequency of 5-year-olds' responses to the compound objects in the Control and Experimental test trials in each condition separately.

In a supplementary analysis, we examined which "mode" of thinking the 4- and 5-year-olds defaulted to in the absence of specific training; that is, in the absence of specific training, do participants believe that causes combine additively to produce larger effects or that it cannot be determined whether causes combine additively because of an imposed ceiling? We fit separate mixed-effects logistic regression models for the 4- and 5-year-olds, with condition (Additive vs. Ceiling vs. Control) as the sole between-subjects fixed effect and participants as the random effect. This analysis involved comparing participants' responses to the five test trial objects in the Control condition to the same five objects in the Additive and Ceiling conditions. The analysis indicated that the main effect of condition was neither significant for the 4-year-olds, $\chi^2(2) = 3.42$, $p = .18$, nor was it significant for the 5-year-olds, $\chi^2(2) = 3.68$, $p = .16$. This result indicated that 4- and 5-year-olds treated the objects equivalently across the three conditions; participants' responses after Additive or Ceiling trainings did not differ significantly from each other and when compared to the Control condition. That participants' responses to the objects in the Control condition did not differ from their responses in the other two conditions suggests that 4- and 5-year-olds may have defaulted to an "intermediate" mode of thinking that was midway between "additive" and "ceiling".

Discussion

In this experiment, children were assigned either to the Additive, Ceiling or Control condition. The Additive and Ceiling conditions were designed to instantiate different "rules", and the Control condition was designed to assess the mode of thinking to which children defaulted in the absence of specific training. The key result was that the 4-year-olds showed backwards blocking only after receiving additivity training; the 5-year-olds showed backwards blocking in all conditions regardless of the training they received. One question that this experiment does not address is how—or by what mechanism—participants reasoned about the events. To answer this question, we fit two computational models to the data—one based on associative learning; the other based on Bayesian inference.

Computational Models

Bayesian model. The model we adopted was used previously (Benton et al., 2023; Griffiths et al., 2011; Griffiths & Tenenbaum, 2005; Tenenbaum & Griffiths, 2001). The model assumes that learners possess a set of hypotheses, h, about which objects are causal and which objects are not. Each hypothesis is assigned a prior probability, $p(h)$, which specifies the learner's belief about that hypothesis before any data is encountered. Learners must then compute the "posterior probability" of each hypothesis (i.e., determine which hypothesis is most consistent with the observed data) by considering each model's prior probability and its likelihood, $p(d|h)$. A model's likelihood was set to 1 if the observed data was consistent with that model; if the observed data was inconsistent with the model, its likelihood was set to 0. Posterior probabilities were computed according to Bayes' rule:

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

The probability that an individual object was a blicket was then computed based on those posterior probabilities:

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

where $p(o \rightarrow E \mid d)$ corresponds to the probability that an object activates the machine given the data, $p(o \rightarrow E \mid h)$ corresponds to the probability that there is a link between an object and the machine's activation given hypothesis h, and $p(h|d)$ corresponds to the posterior probability of a hypothesis calculated using Bayes' rule. We fit the model with the following prior probabilities: 0.5, 0.65, 0.8, 0.95 and 1. We chose to fit the model with these prior probabilities because we did not manipulate prior probabilities in the study. We only report the results of the model whose prior probability best fit the behavioral data.

Connectionist model. We also built a connectionist model that was trained with the Delta Rule (Gluck & Bower, 1988; Widrow & Hoff, 1960), a learning rate of 0.05, and no momentum. The model consisted of an input layer, which was directly connected (via a set of "weights") to the output layer. The input layer consisted of (at most) three units, and the output layer consisted of one unit. Each input unit corresponded to one of three objects, and the single output unit corresponded to the blicket detector. If an object that was a blicket was "placed" on the machine, the activation value of its corresponding input unit was set to 1, and the network was trained to "activate" the single output unit by setting its activation to 1. We trained 100 models from anywhere between 200 and 1000 epochs (i.e., weight updates), in increments of 200. We varied five different numbers of epochs because we varied five different prior probabilities in the Bayesian model. We trained models for different numbers of epochs to ensure that the results were not due to the number of training epochs. Each model run—which corresponded to a different "participant"—involved initializing the weights to small random values (distribution range = ± 0.1). Of the models that were trained for different numbers of epochs, we only present the results of the best-fitting model.

Results

We first looked at children's responses to the five objects in the Experimental and Control test trials for each condition. We averaged the responses for objects A and B (the compound objects) in the Control test trial. We then fit those responses to each of the two models (i.e., associative vs. Bayesian). To determine model fit, we calculated the root mean square error (RMSE) between each model's prediction and the behavioral data. As can be seen in Table 2, we found that the associative model best accounted for the 4-year-olds' judgements in the Ceiling and Control conditions, whereas the Bayesian model best accounted for their responses in the Additive condition. In contrast, the Bayesian model best accounted for the 5-year-olds' responses in all conditions.

Table 2: Fitting Experimental and Control test trial responses for each condition separately.

Condition	Age	Bayesian	Connectionist
Additive	4-year-olds	RMSE = 0.101* Prior = 0.5	RMSE = 0.221 Epochs = 200
	5-year-olds	RMSE = 0.112* Prior = 0.5	RMSE = 0.244 Epochs = 200
Ceiling	4-year-olds	RMSE = 0.075 Prior = 0.8	RMSE = 0.066* Epochs = 1000
	5-year-olds	RMSE = 0.102* Prior = 0.5	RMSE = 0.211 Epochs = 200
Control	4-year-olds	RMSE = 0.10 Prior = 0.8	RMSE = 0.079* Epochs = 200
	5-year-olds	RMSE = 0.119* Prior = 0.65	RMSE = 0.176 Epochs = 200

* Indicates the better fitting model based on RMSE

We next examined which model best accounted for children's backwards blocking responses across conditions; that is, we compared each model's treatment of object B in the Experimental trial and the average of A and B in the Control trial to children's treatment of the same objects. As shown in Table 3, 5-year-olds' backwards blocking responses were best explained by the Bayesian model in all conditions. In contrast, the Bayesian model provided the better description of 4-year-olds' backwards blocking responses only in the Additive condition. This latter result suggests that the 4-year-olds may have relied on Bayesian inference *and* associative learning, with greater weight given to associative learning.

Table 3: Fitting the backwards blocking responses among test trials for each condition separately.

Condition	Age	Bayesian	Connectionist
Additive	4-year-olds	RMSE = 0.139* Prior = 0.5	RMSE = 0.274 Epochs = 200
	5-year-olds	RMSE = 0.118* Prior = 0.5	RMSE = 0.492 Epochs = 200
Ceiling	4-year-olds	RMSE = 0.096 Prior = 0.8	RMSE = 0.079* Epochs = 750
	5-year-olds	RMSE = 0.132* Prior = 0.5	RMSE = 0.268 Epochs = 200
Control	4-year-olds	RMSE = 0.085 Prior = 0.8	RMSE = 0.056* Epochs = 200
	5-year-olds	RMSE = 0.149* Prior = 0.65	RMSE = 0.218 Epochs = 200

* Indicates the better fitting model based on RMSE

Discussion

These results indicate that the mechanism(s) that children use to reason about backwards blocking events undergoes a developmental progression between 4 and 5 years of age: 4-year-olds' causal responses were largely captured by associative learning with minimal evidence of Bayesian inference; 5-year-olds' responses were exclusively captured by Bayesian inference. Bayesian inference involves determining which of a set of hypotheses is most consistent with the observed data. It may be the case that the reason the Bayesian model captured the 5-year-olds' causal judgements, but not that of the 4-year-olds', is that the 5-year-olds may have possessed greater information processing than the 4-year-olds and were better able to "pick out" the winning causal hypothesis. However, this idea is speculative and should be explored further in future research. Another issue that should be examined in future research concerns the fact that the prior probability and the number of epochs for the best-fitting models was not consistent across the different model fits.

Conclusion

The present study is among the first to assess how different ways of thinking about a causal system affected children's backwards blocking inferences. These findings differ and extend previous work (Simms et al., 2012) given the way we operationalized blocking, the simple paradigm used to instantiate additivity training, the use of a baseline or control condition, and our use of computational models to elucidate underlying cognitive mechanisms. We found that although different kinds of training did influence whether 4-year-olds engaged in backwards blocking, 5-year-olds engaged in backwards blocking irrespective of the training they received in the blicket detector paradigm. This finding also extends the study by Simms et al. (2012): They found that as a group 4- and 5-year-olds did not engage in backwards blocking reasoning. In contrast, we found not only that 4- and 5-year-olds showed evidence of backwards blocking reasoning, but that the nature of that reasoning differed between the two ages. We then fit an associative learning model and a Bayesian model to children's causal inferences and found that the Bayesian model provided the best explanation for the 5-year-olds' responses, whereas a combination of Bayesian inference and associative learning provided the best explanation for the 4-year-olds' responses. This finding suggests that children, especially at 4 years of age, may rely on multiple mechanisms to reason about causal events and that the notion that either Bayesian inference or associative learning underlies children's causal reasoning may be misguided; both processes are likely at play, albeit to different degrees.

References

Beckers, T., De Houwer, J., Pineño, O., & Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 238–249.

Benton, D. T., Kamper, D., Beaton, R. M., & Sobel, D. M. (2023). Don't throw the associative baby out with the bayesian bathwater: Children are more associative when reasoning retrospectively under information processing demands. *Developmental Science,* Article e13464.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. The developmental psychology of time, 209-254.

Butler, L. P., Gibbs, H. M., & Tavassolie, N. S. (2020). Children's developing understanding that even reliable sources need to verify their claims. Cognitive Development, 54, 100871.

De Houwer, J., & Beckers, T. (2002). Second-order backward blocking and unovershadowing in human causal learning. *Experimental Psychology, 49*(1), 27–33.

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

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive science*, *35*(8), 1407-1455.

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

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

Haselgrove, M. (2010). Reasoning rats or associative animals? A common- element analysis of the effects of additive and subadditive pretraining on blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 296–306.

Lovibond, P. F., Been, S., Mitchell, C. J., Bouton, M. E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition, 31,* 133–142.

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.

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.

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

Shultz, T. R. (1982). Rules of Causal Attribution. *Monographs of the Society for Research in Child Development*, *47*(1), 1–51.

Simms, V., McCormack, T., & Beckers, T. (2012). Additivity pretraining and cue competition effects: developmental evidence for a reasoning-based account of causal learning. *Journal of experimental psychology. Animal behavior processes*, *38*(2), 180–190.

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

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

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629–640.

Walker, C. M., & Nyhout, A. (2020). Asking "why?" and "what if?": The influence of questions on children's inferences. The questioning child: Insights from psychology and education, 252-280.

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