



The Development of Diagnostic Inference About Uncertain Causes

David M. Sobel, Christopher D. Erb, Tiffany Tassin & Deena Skolnick Weisberg

To cite this article: David M. Sobel, Christopher D. Erb, Tiffany Tassin & Deena Skolnick Weisberg (2017) The Development of Diagnostic Inference About Uncertain Causes, Journal of Cognition and Development, 18:5, 556-576, DOI: [10.1080/15248372.2017.1387117](https://doi.org/10.1080/15248372.2017.1387117)

To link to this article: <http://dx.doi.org/10.1080/15248372.2017.1387117>



Accepted author version posted online: 03 Oct 2017.
Published online: 03 Oct 2017.



Submit your article to this journal [↗](#)



Article views: 22



View related articles [↗](#)



View Crossmark data [↗](#)



The Development of Diagnostic Inference About Uncertain Causes

David M. Sobel^a, Christopher D. Erb^b, Tiffany Tassin^a, and Deena Skolnick Weisberg ^c

^aBrown University; ^bUniversity of North Carolina, Greensboro; ^cUniversity of Pennsylvania

ABSTRACT

Young children can engage in diagnostic reasoning. However, almost all research demonstrating such capacities has investigated children's inferences when the individual efficacy of each candidate cause is known. Here we show that there is development between ages five and seven in children's ability to reason about the number of candidate causes whose efficacy is unknown (Study 1). We also find development between ages six and seven in these abilities when children are presented with several uncertain candidate causes in an additive causal system (Study 2). These findings demonstrate how children's diagnostic reasoning abilities develop beyond the pre-school years and illustrate possible relations between children's developing diagnostic inference and scientific reasoning capacities.

How do we infer the causes of events? Such reasoning is commonplace: A nurse infers that a patient's elevated temperature is the result of an infection. An investigator attempts to recreate the events leading up to an accident. Both of these cases involve reasoning diagnostically—from effects to their underlying cause or causes—in situations of uncertainty. The nurse must await test results to confirm a diagnosis. The investigator must weigh contradictory accounts of the same event. Moreover, each of these cases requires the reasoner to reevaluate conclusions based on new evidence. New information about the patient's symptoms or about the reliability of the eyewitnesses could change the conclusion. As adults, we are capable of engaging in this kind of diagnostic reasoning (e.g., Fernbach, Darlow, & Sloman, 2011; Thomas, Dougherty, Sprenger, & Harbison, 2008). However, we do not know what the origins of these abilities are in childhood, nor do we know how children's abilities develop into full-fledged abilities to reason about causal systems, especially when the efficacy of some causes is unknown. The current paper investigates how children reason in two different situations of uncertainty and how this ability develops over the early elementary-school years.

One line of research has found that children possess early-emerging diagnostic reasoning capacities. Eight-month-olds use conditional probability information to make predictions (Sobel & Kirkham, 2006) and understand what underlying distribution in a population would likely produce an observed sample of data (Xu & Garcia, 2008). Slightly older infants can integrate these learning capacities with their existing knowledge

CONTACT David M. Sobel  dave_sobel@brown.edu  Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HJCD.

© 2017 Taylor & Francis

(e.g., Denison & Xu, 2010; Gweon, Tenenbaum, & Schulz, 2010), which is one way they come to conclusions about the nature of object properties (Wu, Gopnik, Richardson, & Kirkham, 2011) or individuals' preferences (Kushnir, Xu, & Wellman, 2010). By the preschool years, children can explicitly use temporal priority, spatial contiguity, and contingencies among events to diagnose whether certain kinds of causal relations are present (e.g., Bullock, Gelman, & Baillargeon, 1982). Preschoolers can also infer aspects of the causal structure of a set of events from conditional probability information (Gopnik, Sobel, Schulz, & Glymour, 2001; Schulz, Gopnik, & Glymour, 2007).

Although these studies show that children can engage in some forms of diagnostic reasoning before they begin formal schooling, what is critical about all of these cases is that children observe or can deduce the efficacy of all of the potential causes. A paradigmatic example is work by Gopnik et al. (2001). In this experiment, preschoolers were introduced to a machine that lit up and played music when certain objects were placed on it (a "blicket detector" based on Gopnik & Sobel, 2000). Two objects (A and B) were each placed on the machine individually. Object A activated the machine and object B did not. The two objects were then placed on the machine together, which activated. Children in this study judged object A, but not B, to be efficacious. In this case (and many similar procedures, e.g., Schulz & Gopnik, 2004), children observed the individual efficacy of both candidate causes before being asked to diagnose which one was responsible for the machine's activation when both objects were placed on the machine together. This suggests that children's early-developing diagnostic reasoning capacities may be robust only when they know the individual efficacy of each candidate cause.

In support of this suggestion, some prior research suggests that preschoolers have difficulty engaging in diagnostic inferences when they lack certain kinds of information about the causal system. For example, and in line with the work reviewed above, Bindra, Clarke, and Shultz (1980) found that four- and five-year-olds could hold multiple possibilities in mind when making diagnostic inferences about deterministic systems. When faced with multiple *uncertain* candidate causes (i.e., candidate causes whose individual efficacy was unknown), however, children's inferences were at chance levels until age eight.

These data suggest that children's reasoning about uncertainty undergoes significant development during the early elementary-school years. Uncertainty, however, can be conceptualized in a number of different ways. In addition to the issue of whether individual objects have known or unknown efficacy, recognizing uncertainty in causal reasoning can also involve appreciating that causal relations may be stochastic (i.e., a cause does not always produce its effect) or appreciating that a particular cause produces different effects randomly. Children's understanding of uncertainty about stochastic effects or the randomness of causal systems appears to improve between the ages of four and six (Beck, Robinson, Carroll, & Apperly, 2006; Hong, Chijun, Xuemei, Shan, & Chongde, 2005; Sobel, Sommerville, Travers, Blumenthal, & Stoddard, 2009). While we are not measuring these forms of uncertainty in the present study, it is interesting to note that children's ability to reason about such phenomena potentially develops around the same time—a point we will return to in the General Discussion.

Our focus in the current study is on children's abilities to reason about potential causes when they do not know whether each individual cause is efficacious. As noted above, this ability seems to develop between ages four and eight. The general hypothesis that motivates the first study presented here is that children are more successful at diagnostic

inference when they know whether each candidate cause is individually efficacious but are more likely to struggle with such inferences when one or more of the candidate causes has unknown efficacy.

This hypothesis may seem inconsistent with cases in which preschoolers do show some success at reasoning about this kind of uncertain data (e.g., Buchanan & Sobel, 2011; Griffiths, Sobel, Tenenbaum, & Gopnik, 2011; Kushnir & Gopnik, 2007; Lucas & Griffiths, 2010; Schulz, Bonawitz, & Griffiths, 2007; Shultz, 1982; Sobel, Tenenbaum, & Gopnik, 2004). But what all of these studies have in common is that children have some other piece of knowledge that they can use to infer the efficacy of all the potential causes. For example, Sobel et al. (2004) showed children two objects (A and B) that activated the blicket machine together. Children were then shown that object A did not activate the machine by itself, but they did not observe object B on the machine individually. Children, however, could infer that B was a blicket and A was not in this condition based on the logical possibilities, and they likely used that information as the basis for subsequent causal inferences: When asked to make the machine activate, children chose only object B (see McCormack, Butterfill, Hoerl, & Burns, 2009, for similar results). In each case where children seem to be reasoning about uncertainty, then, other considerations involving logical inference, base rate information, or knowledge of causal mechanisms allow children to infer the efficacy of each potential cause individually. Moreover, there is some development here: Three-year-olds tend to struggle with these inferences, while older children are more successful (see Bonawitz et al., 2011; Sobel & Munro, 2009, for potential domain-specific mechanisms for this age-related change).

To examine whether preschoolers struggle with diagnostic reasoning about uncertain candidate causes more directly, Fernbach, Macris, and Sobel (2012) compared how children engaged in such reasoning when the efficacy of all of the candidate causes was or was not known. They showed three- and four-year-olds three objects (A, B, and C) and a novel machine. Across two conditions, one object was shown to activate the machine (e.g., A_✓) and a second object was shown not to activate the machine (e.g., B_✗). What differed between the conditions was the status of a third object. In one condition, the third object was shown to activate the machine (C_✓); in the other, it was never placed on the machine, so its efficacy was not known (C_?). The machine and objects were then occluded from the children, and the experimenter activated the machine with one of the objects. The child was asked which object had just been used to activate the machine. Regardless of the object children chose, they were told they were incorrect and asked to make a second choice.

In both conditions, children should avoid choosing the object that had previously failed to activate the machine (B_✗) in response to the two test questions. When the efficacy of the third object was demonstrated (C_✓), children tended to avoid choosing the inefficacious (B_✗) object. When the efficacy of the third object was not demonstrated (C_?), however, children often chose the inefficacious object (B_✗). Three-year-olds' responses did not differ from chance in this condition. Four-year-olds did generate a correct sequence of choices more often than chance expectations, but far from ceiling levels; they only avoided the inefficacious object about 45% of the time. In follow-up work, Erb and Sobel (2014) documented that children's ability to treat the third (C_?) object as a potential cause developed between the ages of four and seven, with four- to five-year-olds only making

a correct sequence of responses about 69% of the time, while six- and seven-year-olds made a correct sequence of responses around 90% of the time.

Here we extend these findings in two ways. First, in Study 1, we examine how individual children make diagnostic inferences when all causes are known and when the number of unknown causes is varied. This way, we can more specifically map the developmental trajectory of children's abilities to make such inferences given that previous findings have only compared reasoning about causes with known and unknown efficacy at a group level. Investigating the role that uncertainty plays in causal reasoning can allow researchers to study causal reasoning not only as an abstract topic but also as one that is embedded in everyday life. For example, many studies suggest that young children, who are quite capable of the kinds of abstract causal inferences discussed so far, struggle with general scientific reasoning problems that require them to diagnose which of several potential causes with unknown individual efficacy produced a given set of results (e.g., Klahr, 2000; Klahr, Fay, & Dunbar, 1993; Koslowski, 1996; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Garcia, Zohar, & Andersen, 1995; Kuhn, Schauble, & Garcia-Mila, 1992; Lehrer & Schauble, 2000; Masnick & Klahr, 2003; Samarapungavan, 1992). While there are several crucial differences between the tasks used in those studies and in the body of work reviewed above (an issue we consider in more detail in the General Discussion), Study 1 investigates how the presence of uncertain candidate causes influences children's reasoning at different ages. This begins to address the gap between the literature on abstract causal inference and generic scientific reasoning problems.

Second, in Study 2, we investigate whether children can make diagnostic inferences about uncertain candidate causes in an interactive causal system. In most studies of children's causal reasoning, including our own Study 1, the parameterization of the causal model is disjunctive: An effect is produced if one or more candidate causes are present. These are the most commonly used type of model in experiments measuring children's (and adults') causal reasoning (as well as, potentially, in everyday life; see Lucas, Bridgers, Griffiths, & Gopnik, 2014, for a summary of this argument). In Study 2, we considered how children reasoned about a causal model with an additive element: Two candidate causes independently produced an effect but together combined to produce a different effect.

Such additive (in this case, conjunctive) causality is commonplace in everyday and scientific reasoning. One example is how a trebuchet (a slinged catapult) fires its payload. The distance a payload travels is an additive function of the mass of the counterweight, the mass of the payload, and the length of the sling. For example, the heavier the counterweight and the lighter the payload, the farther the payload travels. Such inferences are necessary not only for medieval warfare (and better appreciating certain battle scenes in *Game of Thrones*) but also in everyday reasoning, such as when one is trying to predict how a golf ball will travel when struck, which is based on the same additive model as a trebuchet.

Study 2 examines whether children could infer an additive causal model from data that do not disambiguate exactly how the causal system worked. Other studies have examined aspects of this question, but those studies have not tested children as young as those we investigate here, and they have also tended to present causal systems with some scientific content (another difference between our work and work in the field of scientific reasoning that we probe further in the General Discussion). For example, Schauble (1990, 1996) used

several examples of additive causal models in her investigations (e.g., examining predictions about how depth of a canal and mass of a boat relates to the speed with which the boat travels or how the presence of certain features on a car predicts its speed). Similarly, Kuhn (2007; see also Kuhn & Dean, 2005) presented an additive model of levels of earthquake risk, in which the presence of multiple candidate causes added together to increase the risk of earthquakes. In these examples, fourth and sixth graders often performed poorly on learning additive structures from their own investigation.

The goal of Study 2 is not to replicate these prior findings in scientific reasoning, as there are other critical differences among the present studies and those that measure scientific reasoning. Rather, we aim to examine whether children have the ability to diagnose causal efficacy from observed effects when the efficacy of candidate causes is uncertain. Further, taken together, these two studies shed new light on how children develop this crucial component of everyday reasoning, which is necessary both inside and outside of the lab.

Study 1

Study 1 used a paradigm similar to Fernbach et al. (2012). Children were introduced to a novel machine that lights up and plays music when objects are placed on it (the “blicket detector,” following Gopnik & Sobel, 2000) and sets of four objects. In one trial, all four objects were demonstrated on the machine, with three of them being efficacious ($A_{\checkmark}$, $B_{\times}$, $C_{\checkmark}$, $D_{\checkmark}$, in this order). In another trial, three objects were demonstrated and two were efficacious ($A_{\checkmark}$, $B_{\times}$, $C_{\checkmark}$, $D_{?}$, in this order). In the third trial, two objects were demonstrated and only one was efficacious ($A_{\checkmark}$, $B_{\times}$, $C_{?}$, $D_{?}$, in this order). In each of these three trials, following the demonstration, the machine and objects were then occluded from the child and the experimenter activated the machine. Children were asked which object was used to make the machine go. Regardless of which object was selected, children were told that they were incorrect and were asked to revise their answer. Children were then informed that their second selection was incorrect and were asked to revise their answer a second time. This method enabled us to assess whether differing degrees of uncertainty influenced how frequently children committed an error by selecting an object that they had observed fail to activate the machine. It also enabled us to identify whether the probability of such errors changed across children’s first and second selections.

Method

Participants

The final sample consisted of 90 children from five age groups: three-year-olds ($n = 18$; $M_{age} = 42.22$ months, range = 36.00–47.90 months; 8 boys, 10 girls), four-year-olds ($n = 18$; $M_{age} = 54.17$ months, range = 48.00–59.40 months; 7 boys, 11 girls), five-year-olds ($n = 18$; $M_{age} = 65.22$ months, range = 60.00–71.20 months; 10 boys, 8 girls), six-year-olds ($n = 18$; $M_{age} = 77.09$ months, range = 72.50–82.70 months; 10 boys, 8 girls), and seven-year-olds ($n = 18$; $M_{age} = 90.62$ months, range = 84.10–95.60 months; 9 boys, 9 girls). Data from four other children were collected but not included due to refusal to participate ($n = 1$) and experimental error ($n = 3$). Children were recruited through a list

of birth records and flyers posted at local preschools or at a local children's museum. Most children were Caucasian and from middle- to upper-middle-class families; however, no specific indicators of socioeconomic status were obtained.

Materials

The machine was a $20.30 \times 15.20 \times 7.60$ cm black plastic box with a pressure-sensitive white plate on top. Under the white plate was a set of LCD lights that were visible through the plate when the machine was activated. A button on a remote control was used (out of sight of the child) to activate the machine. Whenever the button was pressed the machine would light up and play music. The machine was battery-powered, so no external cords ran to or from it. Four sets of four wooden blocks were also used. In each set, the blocks were the same shape (two sets of rectangular blocks with different dimensions, one set of cylindrical blocks, and one set of triangular blocks, all approximately 3 to 6 cm in height), but four different colors, so children saw 16 colored blocks. A 56.00×43.00 cm piece of cardboard was also used as an occluder.

Procedure

All children were tested by a male experimenter in the laboratory or in a quiet room at a children's museum. Children's caregivers were present during the testing. The machine was brought out to the table, and the experimenter told the child that they would play a game with the machine and some toys. They were told the machine was special because some toys made the machine go and some toys did not.

Children were given four trials. The first trial was always a *prediction* trial; the other three were *diagnostic* trials. After the prediction trial, the other three trials were presented in one of six random orders (counterbalanced).

On the prediction trial, two blocks, identical in shape but different in color, were initially placed on the table. The experimenter placed each block on the machine individually. One block activated the machine and the other did not (e.g., a blue cylinder made it go, but the black cylinder did not). This was demonstrated twice. Then, the experimenter placed two more blocks on the table. They were the same shape as the first two blocks, but unique colors (e.g., an orange cylinder and a brown cylinder). Each of the new blocks was placed on the machine one at a time and was shown to activate the machine. This was also demonstrated twice. The child thus saw four blocks total, three that successfully activated the machine (e.g., blue, orange, and brown) and one that did not (e.g., black).

The experimenter then placed the piece of cardboard between the child and the machine and objects, so that the child could not see the machine or objects, and said, "I'm going to put one of these objects on the machine." The experimenter mimed placing one of the objects on the machine behind the cardboard. He used the remote control, which he held in his hand to hide it from the child, to "activate" the machine. The child could hear the sound of an object going on the machine and the sound of the machine activating. The experimenter emphasized this by saying, "Do you hear that? The machine is going." Note that none of the blocks actually changed position on the table. The experimenter then mimed returning an object to its location, hid the

remote control, and removed the occluder so that the child could see the four blocks and the machine. The experimenter then asked the child whether each block would “make the machine go.” This trial served as a memory check for the three diagnostic trials, which all had a similar structure. We wanted to ensure that children were able to remember the efficacy of the four blocks, since the diagnostic trials relied on their ability to do so.

Following this prediction trial, children engaged in three diagnostic trials. These three trials differed in the number of uncertain objects that were present when the child was asked to infer whether each object could activate the machine. The *all-known* trial had an identical structure to the prediction trial, where children saw the same demonstration (three blocks made the machine go and one did not) with a new set of blocks. But instead of asking the child to predict whether each block would activate the machine, the experimenter asked, “Which one of the toys did I just put on the machine when I had this [indicating the occluder] between us?” After the child chose a block, the experimenter said, “That’s a really good guess, but it wasn’t this one.” He removed that block from the table and asked the first belief revision question: “Can you try again? Which one do you think it was?” After the child responded this time, the experimenter asked for a second belief revision. When the child made this third response, the experimenter did not tell the child whether she or he was correct and just moved on to the next trial. Note that for this trial, as in the prediction trial, children knew the efficacy of all four blocks (three had previously activated the machine and one had failed to do so).

The *one-unknown* and *two-unknown* trials had the same structure as the *all-known* trials except for what happened when the second pair of objects was brought out during the initial demonstration phase. Either one or both of those objects were not demonstrated on the machine. In the *one-unknown* trial, the one object that was demonstrated on the machine activated it (shown twice). This means that in the *one-unknown* trial, children knew the efficacy of only three of the four blocks (two activated the machine, one did not, and one was not demonstrated). In the *two-unknown* trial, children knew the efficacy of only two of the four blocks (one activated the machine, one did not, and two were not demonstrated).

Results

Following previous work (Erb & Sobel, 2014; Fernbach et al., 2012), we were primarily concerned with whether children would choose the blocks with unknown efficacy, rather than the block that had previously failed to activate the machine, when asked to guess again. We were also interested in the effects of age and trial type (i.e., how many blocks there were with unknown efficacy) on children’s choices. To measure these effects, we calculated how many children in each trial never chose the inefficacious block; we considered this pattern to be error-free performance. Specifically, for the prediction trial, we counted how many children said that the objects that previously activated the machine would do so again and that the object that did not activate the machine would not do so. For the three diagnostic trials, we counted how many children chose only the block(s) that had previously activated the machine and the block(s) with unknown efficacy, but not the block that they had observed fail to activate the machine. Table 1 shows the percentage of error-free performance on each trial among the five age groups.

Table 1. Proportion of error-free trials across the age groups and quartiles of ages who passed the prediction trial in Study 1.

All Data				
	Prediction Trial	All-Known Trial	One-Unknown Trial	Two-Unknown Trial
3-year-olds	.56 ^a (0.51)	.33 (0.49)	.39 (0.50)	.22 (0.43)
4-year-olds	.72 ^a (0.46)	.39 (0.50)	.33 (0.49)	.50 ^b (0.51)
5-year-olds	.94 ^a (0.24)	.78 ^b (0.43)	.33 (0.49)	.44 (0.51)
6-year-olds	1.00 ^a (0.00)	.50 ^b (0.51)	.50 ^b (0.51)	.56 ^b (0.51)
7-year-olds	.94 ^a (0.24)	.67 ^b (0.49)	.61 ^b (0.50)	.78 ^b (0.43)
Only Children Who Are Successful on the Prediction Task				
	All-Known Trial	One-Unknown Trial	Two-Unknown Trial	
First age quartile (<i>n</i> = 18; <i>M</i> _{age} = 46.67 months, <i>SD</i> = 4.79)	.39 (0.50)	.28 (0.46)	.39 (0.50)	
Second age quartile (<i>n</i> = 19; <i>M</i> _{age} = 62.39 months, <i>SD</i> = 4.31)	.68 ^b (0.48)	.32 (0.48)	.42 (0.51)	
Third age quartile (<i>n</i> = 19; <i>M</i> _{age} = 75.42 months, <i>SD</i> = 3.25)	.58 ^b (0.51)	.48 ^b (0.51)	.64 ^b (0.50)	
Fourth age quartile (<i>n</i> = 19; <i>M</i> _{age} = 89.61 months, <i>SD</i> = 4.31)	.63 ^b (0.50)	.68 ^b (0.48)	.74 ^b (0.45)	

Note. Standard deviations are shown in parentheses.

^aIndicates greater than chance-level performance on the prediction trials (based on binomial test, $p < .05$, with chance = 0.125).

^bIndicates greater than chance-level performance on any of the three diagnostic trials (based on binomial test, $p < .05$, with chance = 0.25).

Preliminary analyses suggest that gender and the order in which children received these trials did not affect responses (Chi-squared analyses, all p values $> .16$).

We first considered performance on the prediction trial. All age groups were above chance on this trial, all binomial tests, all p values $< .05$. Further, error-free performance on the prediction trial significantly correlated with age, $\rho(87)^1 = .38$, $p < .001$. However, because the prediction trial served as a control measure, we chose to subsequently analyze data only from children who passed it. Fifteen children were excluded for this reason (eight 3-year-olds, five 4-year-olds, one 5-year-old, and one 7-year-old), leaving 75 children in the analyses reported below.

When we eliminated the children who failed the prediction question from the analysis, we did not find a significant correlation between age and performance on the all-known trial, $\rho(72) = .171$, $p = .15$. In contrast, error-free performance on the one-unknown and two-unknown trials did significantly correlate with age, $\rho(72) = .30$ and $.32$, $p = .01$ and $.008$, respectively. To analyze the role of age further, we used age quartiles to create four groups of equivalent sample sizes; this corrected for the fact that the majority of children who were removed because they failed the prediction task were the youngest in our sample. These data are also shown in Table 1. The youngest quartile ($M_{age} = 46.67$ months, range = 36.00–53.70 months) responded at chance levels on all

¹The parents of one six-year-old identified her as a six-year-old but refused to provide us with a birthday. As a result, we excluded this child from this and other analysis regarding age (hence the lower degree of freedom of this and subsequent analyses of age).

three diagnostic reasoning trials. The second quartile ($M_{age} = 62.39$ months, range = 54.20–68.40 months) responded above chance on the all-known trial, binomial test, $p < .001$, but no different from chance on the one- and two-unknown trials. The third quartile ($M_{age} = 75.42$ months, range = 69.20–80.40 months) and fourth quartile ($M_{age} = 89.61$ months, range = 81.60–95.60 months) responded above chance on all three diagnostic reasoning trials, binomial tests, all p values $< .05$.

We next considered how children erred on each of the diagnostic reasoning trials, that is, whether children chose the inefficacious block as their first, second, or third choice. This is critical to investigate because younger children might have generated more errors because they struggled to revise their beliefs multiple times. Contrary to this possibility, whenever children made an error on any of the diagnostic trials, that error was evenly distributed among the three questions, all $\chi^2(2)$ -values < 1.76 , all p -values $> .41$. That is, children did not just make an error on the last question. That all of these children responded correctly on the prediction question suggests that the younger children’s poorer performance was not a result of their inability to remember which objects were efficacious.

Because each child responded to all three diagnostic questions, we next considered individuals’ patterns of performance. If children go through a developmental progression based on the amount of uncertainty in the inference, we would expect four patterns of performance to occur more frequently: 1) cases in which children respond erroneously to all three diagnostic trials, 2) cases in which children are correct on only the all-known trial, 3) cases in which children are correct on the all-known and one-unknown trial, and 4) cases in which children are correct on all three trials (see Table 2). These four patterns of performance were indeed more likely than the other four patterns of performance,

Table 2. Patterns of response among the three diagnostic inference trials in Study 1.

Patterns Consistent With Developmental Trajectory	Percentage of Children Generating This Pattern	Mean Age (in Months) of Children Generating This Pattern
Errors on all diagnostic trials	20.0%	54.94 (13.39)
No error on all-known trial, errors on one- and two-unknown trials	15.5%	61.43 (16.54)
No errors on all-known and one-unknown trials, error on two-unknown trials	7.8%	64.01 (16.80)
No errors on all trials	20%	78.90 (12.33)
Total of consistent trials	63.3%	
Patterns Inconsistent With Developmental Trajectory		
No error on one-unknown trial, errors on all-known and two-unknown trials	6.7%	54.12 (20.14)
No error on two-unknown trial, errors on all-known and one-unknown trials	11.1%	64.62 (17.21)
No error on all-known and two-unknown trials, error on one-unknown trial	10%	68.86 (16.48)
No error on one-unknown and two-unknown trials, error on all-known trial	8.9%	66.68 (20.09)
Total of inconsistent trials	36.7%	

Note. Standard deviations of age are in parentheses.

occurring 63% versus 37% of the time, binomial test, $p = .02$.² Of the children whose responses fit these patterns (i.e., the top half of Table 2), there was a significant relation

between the number of error-free trials children produced and age, $p(45) = .47$, $p = .001$. There was no such relation for the children who did not fit this developmental profile (i.e., the bottom half of Table 2), $p(25) = .14$, $p = .48$.

Discussion

In order to investigate how young children reason under conditions of uncertainty, three- to seven-year-olds were asked to track multiple potential causes of an effect. We varied the extent to which children knew the efficacy of the potential causes. In one trial, the efficacy of all objects was known. In other trials, one or two objects had unknown efficacy. We observed change between ages four and six in children's ability to reason about potential causes with unknown efficacy. When all objects had known efficacy, the youngest children in our sample (those in the first age quartile) struggled to identify which objects could have caused the effect, while older children were able to select appropriate responses at above-chance levels. Children in the second age quartile still struggled when at least one of the objects had unknown efficacy. For children in the oldest two quartiles, reasoning on these trials was more accurate than for younger children and also more accurate than chance. Looking at individual patterns of performance, more children responded to the three diagnostic trials in a manner that was consistent with their developing the ability to consider increasing levels of uncertainty. That is, the degree to which children could navigate greater numbers of uncertain objects increased with age.

These data suggest that the ability to make diagnostic inferences about potential causes develops during and after the preschool years, dependent on the number of potential causes the efficacy of which is unknown. As children get older, they seem able to cope with more uncertainty in their inferences.

Equally important is that the youngest children examined here responded at chance levels in our diagnostic reasoning task even when the efficacy of each potential cause was known. This finding is inconsistent with much of the established literature suggesting that young children can engage in these kinds of causal inferences. But a critical difference between our measure and these previous findings is the overall number of objects and diagnostic inferences children had to track. For example, in paradigmatic examples of children's precocious causal inference, they must track the efficacy of only two potential causes (e.g., Bullock et al., 1982; Gopnik et al., 2001; Sobel et al., 2004), whereas here they had to track four. Moreover, other work has also found developmental differences in children's successful causal reasoning during this age range (e.g., Buchanan & Sobel, 2011; McCormack et al., 2009; Nazzi & Gopnik, 2000; Shultz, 1982; Sobel & Munro, 2009). This suggests that the complexity of the causal model that children must hold in mind affects their diagnostic reasoning capacities, which is consistent with the developmental trajectory we mapped in this study. This result also suggests that children's developing working memory and other cognitive control capacities influence their diagnostic reasoning. Working memory is also developing at this time and is necessary to hold in mind the efficacy of the various objects we presented (see e.g., Cowan, 1997; Luciana & Nelson, 1998; Towse & Hitch, 1995). Our

²We believe that this analysis takes the most conservative approach possible. It assumes that children who, for example, can only reason about the all-known condition do not succeed on either the one- or two-unknown trials simply by chance. The values reported in Table 2 potentially underestimate the number of children whose responses are consistent with this developmental trajectory.

decision to include only children who passed the prediction trial attempted to control for these factors; we suspect, for example, that children who failed the prediction trial might have been guessing on the diagnostic reasoning trials.

In Study 2, we explored a different aspect of children’s developing capacity to diagnose uncertain causes by changing the parameterization of the causal structure children had to learn. In Study 1 (and in most research on children’s causal reasoning), causal structures are inherently disjunctive: An effect occurs if at least one candidate cause is present. In real life, causal structures can be this simple, but they are often more complex: Effects occur if enough candidate causes are present, but some events individually are insufficient to produce the effect. To use the trebuchet example from the introduction, counterweights were often stones that also acted as payloads. One individual stone could not effectively launch the payload, so several had to be present in order to launch the payload a particular distance.

To investigate these sorts of systems, Study 2 presented children with a causal system that featured an additive mechanism: More than one block was necessary to activate the machine in a certain way. Although additive models can be more complex than simple conjunctive causes like this one, we used this structure because it is the simplest way to introduce this type of uncertainty into a reasoning problem. Specifically, we showed children a situation in which combinations of four blocks were placed on a novel machine. Two of those blocks were not efficacious. The other two worked in an additive (in this case, conjunctive) manner. This structure is shown in Figure 1. When each of the two efficacious blocks was placed on the machine individually, the machine would activate red

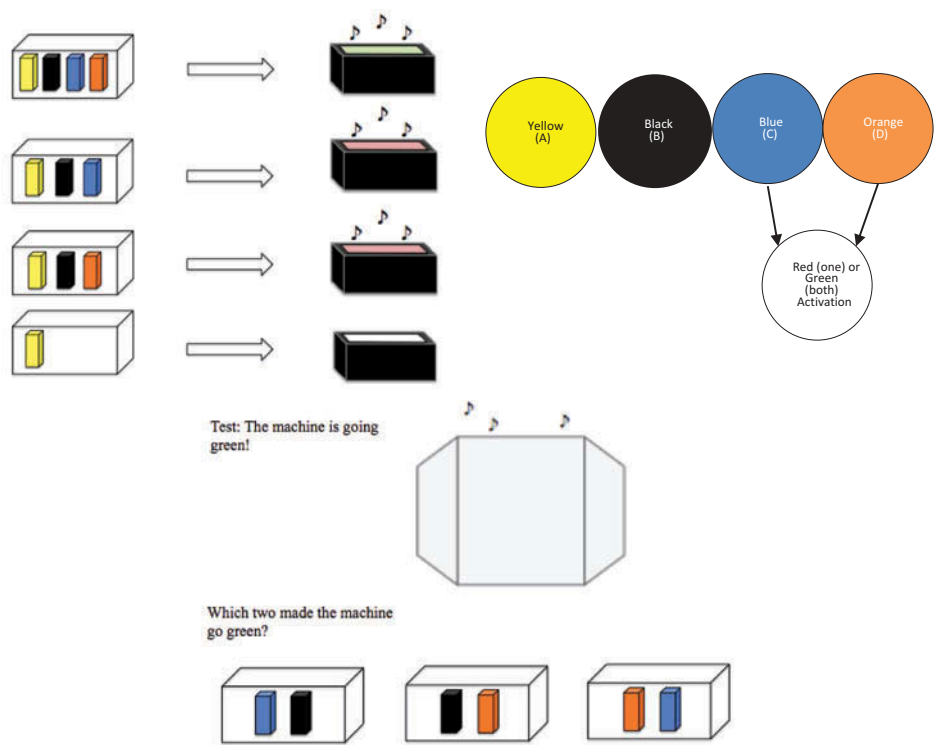


Figure 1. Causal reasoning procedure used in Study 2 and model of this structure.

and play a particular song. If both were placed on the machine together, the machine would activate green and play a different song. Children observed data that unambiguously allowed them to make this diagnostic inference but always presented the efficacious blocks in combination with other blocks. Uncertainty in this system was thus built into the way that the blocks were presented, which did not automatically disambiguate the role of each block. Children thus needed to track not only which blocks (i.e., potential causes) were present and the efficacy of those causes but also how that efficacy related to the presence of other potential causes. Because this task is more difficult than that in Study 1 and because the majority of three- and four-year-olds responded at chance in Study 1, we investigated a slightly older cohort in Study 2, focusing on five- to eight-year-olds.

Study 2

Study 2 examined a more complex case of diagnostic reasoning, in which children had to infer not only the efficacy of individual objects but also the causal relations among them. We introduced children to a nonobvious additive causal model using the blicket detector paradigm. Children were shown the machine and a set of four objects. One object was shown to be ineffective. Children had to infer the efficacy of the other three objects from the pattern of data, in which two objects were efficacious and the third one was not. They also had to learn that presenting the two effective objects together created a different effect. This kind of additive model is used in some measures of scientific reasoning, in which children have to diagnose the causal structure of a set of variables in some domain (e.g., Chen & Klahr, 1999; Kuhn & Dean, 2005; Schauble, 1990, 1996). Although there are many differences between our decontextualized causal reasoning problem and these investigations in scientific reasoning (which we will review in the General Discussion), we believe that causal systems of this type generally bear a closer resemblance to those those in real-world reasoning problems.

Method

Participants

The final sample consisted of 72 children, 36 between the ages of five and six (17 girls, 19 boys; $M_{age} = 72.88$ months, range = 60.90–83.70 months) and 36 between the ages of seven and eight (16 girls, 20 boys; $M_{age} = 95.68$ months, range = 84.50–119.70 months). Eight additional children (six from the younger group and two from the older group) were tested but were not included in the final analysis due to experimenter error ($n = 1$), a loss of video footage ($n = 1$), and failing the control procedure outlined below ($n = 6$). Participants were recruited through birth record lists and a local children's museum. Although specific demographic information was not obtained, most children were Caucasian and from middle- or upper-middle-class families.

Materials

The machine was a battery-powered black plastic box measuring $20.30 \times 15.20 \times 7.60$ cm. The top of the machine was covered with a white pressure-sensitive plate. LCD lights were

oriented under the plate and were visible upon activation. A hidden remote control activated the LCD lights either red or green, and the box played one of two songs depending on which color was activated.

Two sets of four wooden blocks were used: yellow, black, blue, and orange cuboids and blue (a different shade from the first set), pink, light gray, and dark gray cylinders. A transparent plastic box measuring approximately $9.00 \times 9.00 \times 11.50$ cm was used to hold these blocks when they were placed on the machine. There were also 14 photographs (approximately 10.50×15.50 cm) depicting various combinations of the blocks used in the experiment (8 for the demonstration phase and 6 for the test trials), and one green and two red circular paper dots (1.00 cm in diameter). A piece of cardboard with dimensions approximately 36.00×91.00 cm was used as an occluder. The occluder was folded so that the child could not see around it.

Procedure

Children were tested either in the laboratory or in a quiet room at a children's museum. The procedure is shown in [Figure 1](#). The experimenter placed the machine on the table and informed the child that they would be playing a game with a machine that could light up and play music.

There were two trials, the first of which used the set of cuboid blocks and the second of which used the set of cylindrical blocks. The effect that the four blocks (ABCD) had on the machine was demonstrated in one of two orders: ABCD, ABC, ABD, and then A or the reverse order (counterbalanced among participants). We will use the first order to illustrate the procedure.

First, the experimenter took all four blocks (ABCD), placed them in the plastic box, and put the plastic box on top of the machine. The machine activated green and played the song associated with this activation. She asked the child, "What color is the machine going?" She waited for a response (e.g., "green") and then tried the four blocks on the machine again. She took the blocks off the machine and repeated that all four blocks on the machine made the machine light up green. The experimenter then showed the child a photograph of the four blocks and placed it on the table next to the machine. She placed a green circular chip next to the picture for the child to reference, indicating that the machine turned green in this scenario. This same procedure was repeated for the following combinations: blocks A, B, and C together turned the machine red; blocks A, B, and D together turned the machine red (and in both cases, red chips were placed next to the photograph to serve as an explicit memory aid); and block A did not activate the machine and no chip was placed next to that photograph (see [Figure 1](#)).

Children were then shown a test sequence. The experimenter said she would try two blocks on the machine without the child being able to see. She placed the piece of cardboard in front of the machine to occlude the child's view, moved two blocks into the plastic box, and placed it on the machine. The machine activated green and played the song associated with green activation. Even though the child could not see what blocks were on the machine, she or he could hear the song and verify the activation. The activation was demonstrated twice, and then the experimenter placed the two blocks back on the table.

The experimenter then removed the cardboard, revealing the four blocks in their original locations. The experimenter then introduced three photographs with pairs of blocks: blocks B and C, blocks B and D, and blocks C and D. The order in which the choices were introduced was randomized. The experimenter asked the child to choose which of these combinations had just activated the machine green (the correct answer being C and D, as the machine only activated green in the presence of both C and D). After the child responded to the first trial, the experimenter repeated this procedure with the cylindrical set of blocks.

After the second trial, children were asked two control questions about what happened in that trial. With the photographs of the second set of blocks and the green and red dots on the table, children were asked to remind the experimenter what happened when all four blocks were placed on the machine (green activation) and to remind the experimenter what happened when only block A was placed on the machine (nothing). Children had to answer correctly in order to be included in the analysis; failure to do so might indicate that the child did not understand the procedure. Six children failed this check and were removed from analysis (five from the younger group and one from the older group).

Children were also asked to define the word “science,” but those data were collected as part of a different investigation and will not be discussed here.

Results

Correct responses (i.e., children choosing the CD block combination) to the two trials significantly differed, McNemar $\chi^2(1, n = 72) = 3.84, p = .05$, and this difference was particularly profound for the older half of the sample, who responded correctly 61% of the time on the first trial and 22% on the second, McNemar $\chi^2(1, n = 36) = 7.04, p = .007$. Although this difference was unexpected and its cause is unclear, the effect may have been driven by a factor unrelated to children’s causal reasoning: One of the ineffective objects on the second trial was pink, and a majority of children (69%) chose one of the responses with that object on the second trial. It is possible that children’s color preferences (preferences for pink in particular) influenced their judgment. A more compelling reason is that because no feedback was given in response to the first trial, children who responded correctly might have interpreted the absence of explicit positive feedback as indicating that their response was incorrect, leading them to switch their answer on the subsequent trial. Indeed, children who responded correctly on the first trial switched to an incorrect response on the second trial 81% of the time, suggesting a strategic response to the lack of feedback. Children who responded incorrectly on the first trial switched to a correct response on the second trial only 42% of the time, which was significantly different, $\chi^2(1, n = 72) = 11.46, p < .001, \Phi = 0.40$.

As a result of this unanticipated difference between trials, we will restrict our main analysis to just the first trial. Preliminary analyses suggested no difference between the genders of the participants or between the orders in which data were presented, both $\chi^2(1, n = 72) < 2.03$, both p values $> .15$.

Overall, the five- to six-year-olds chose the correct pair of blocks (C and D) on 39% of the trials, no different from chance performance (33%), binomial test, $p = .29$. The seven- to eight-year-olds chose the correct pair on 61% of the trials, marginally more often than

the younger children, $\chi^2(1, n = 72) = 3.56, p = .059, \Phi = 0.22$, and significantly more often than chance, binomial test, $p = .001$.

Discussion

Study 2 was designed to further investigate how children make diagnostic inferences in situations of uncertainty. Unlike Study 1, which introduced uncertainty by never demonstrating the effects of some of the objects, this study introduced uncertainty by only ever demonstrating the efficacious objects' effects in combination with each other. We found that children started to be able to diagnose this simple additive structure between the ages of six and seven. Building on Study 1, these data suggest that children in the early elementary years are capable of diagnosing not only the efficacy of individual causes when they are unknown but also how those causes relate to each other.

One concern with the present conclusion is that the conjunctive inference required to respond correctly in Study 2 could be explained via an associative inference. If one simply calculates the ratio of each object's presence on the machine and it activating green, objects C and D have the highest ratios. We do not believe, however, that children are responding based solely on this association. If they were, we would likely have not seen the difference between the two trials in the study; that children were potentially influenced by other factors such as object color suggests that when they did respond correctly it was not on the basis of another association. Moreover, there are numerous examples of causal reasoning in young children that are not explained easily by general associative inference (e.g., Griffiths et al., 2011; Lucas et al., 2014; Sobel et al., 2004). It seems unlikely that children would make an associative inference here, while using other reasoning mechanisms for similar patterns of data.

General discussion

The present studies illuminate how diagnostic reasoning develops past the preschool years. Although many studies have examined children's diagnostic reasoning capacities, an open question concerns whether and how this capacity changes across development. The current study targets an important factor that seems to mediate successful diagnostic inference at different ages: whether the efficacy of candidate causes is certain or uncertain. While preschoolers have sophisticated diagnostic reasoning abilities when they are aware of the efficacy of all objects individually, when the individual efficacy of some objects is not known, their reasoning abilities undergo significant developmental change throughout and after the preschool years.

Study 1 illustrates that it is not until age five that children show evidence of successful diagnostic reasoning when the efficacy of individual candidate causes is unknown. This ability improves during the sixth and seventh years. Study 2 demonstrates that between the age of five and eight, children are developing the ability to reason diagnostically about the additive effects of multiple causes when these causes are not demonstrated individually. Although it is not surprising to see children's inference improve with age, we posit that a particular capacity—the ability to make inferences when candidate causes' efficacies are uncertain—plays a significant role in the development of diagnostic reasoning. In both real-life settings and in laboratory science, it is rare to find cases in which the efficacy of all

causes is known (and even cases where all possible causes are known). This means that mature diagnostic reasoning abilities must allow one to grapple with uncertainty. Further, the particular kind of development uncovered in Study 1 suggests that this is a hard-won achievement, as the ability to handle increasing amounts of uncertainty improves with age, even though children's success on the prediction task suggests that they possess the working memory required to make this inference. This in turn sheds light on the development of causal reasoning more generally: Despite some impressive early abilities, this set of skills faces a protracted developmental process.

Two important questions emerge from these findings. The first concerns what cognitive mechanisms are responsible for the age-related change that we report. The second concerns how these findings relate to the broader literature on children's diagnostic and scientific reasoning. We discuss each of these questions in turn.

What mechanisms are responsible for this change?

There is a tension in the literature between two relevant but distinct abilities that relate to success on diagnostic reasoning tasks. First, diagnostic reasoning requires that children hold probabilistic hypotheses in mind and reason over a hypothesis space; they must infer which hypothesis among a set of plausible hypotheses correctly reflects how a causal system works. Many studies suggest that children have the rudiments of this ability at young ages: They can track statistical regularity among events and use that regularity to build hypotheses about the physical and psychological world (e.g., Denison & Xu, 2010; Kushnir et al., 2010; Sobel & Kirkham, 2006; Walker & Gopnik, 2014; Xu & Garcia, 2008).

Second, diagnostic reasoning often requires children to appreciate various forms of uncertainty. The uncertainty that we considered here focused on recognizing that unknown potential causes are not necessarily inefficacious (Bindra et al., 1980). Understanding this kind of uncertainty allows children to keep potential hypotheses that have not been disconfirmed as potential explanations for observed data and relevant for predictions about the future. This kind of uncertainty is related to appreciating that efficacious candidate causes can have different outcomes and may or may not be efficacious at a certain moment in time (e.g., Beck et al., 2006; Perner, 1979; Sobel et al., 2009). It is interesting to consider that understanding both kinds of uncertainty potentially have similar developmental trajectories. Moreover, this latter form of reasoning may be more explicit than the first: In Study 1, children were able to articulate different candidate causes after being told they were incorrect. Macris and Sobel (2017) similarly found that explicit abilities to engage in belief revision are also still developing after the preschool years, again contributing to the possibility that children develop a broader understanding of uncertainty and use it when making diagnostic inferences.

This ability may relate to children's developing metacognitive and working memory skills. Metacognition is related to uncertainty, as it allows children to question the nature of the inferences they are making. Metacognitive abilities have a similar developmental trajectory to what we have found here, both in terms of how confidence judgments predict accuracy (Hembacher & Ghetti, 2014) and whether children can monitor their own judgments of learning based on the amount of time they studied material (Destan, Hembacher, Ghetti, & Roebers, 2014); both abilities develop between ages five and seven. In addition, working memory allows children to remember that there are multiple

potential causes. This capacity also undergoes significant development during the early elementary school years (Cowan, 1997; Luciana & Nelson, 1998; Towse & Hitch, 1995).

The results of the two studies reported here suggest that children's understanding of uncertainty per se also undergoes significant development during the early elementary-school years. This ability allows children to reason that candidate causes with unknown individual efficacy should be treated as potential causes, as opposed to being ignored or assuming that they specifically are or are not efficacious. This capacity allowed children to reason appropriately when faced with such uncertain candidate causes in Study 1. It also allowed children to make the appropriate inference in Study 2: When they were told at the test trial that only two objects activated the machine green, they had to recognize that the C and D objects were the cause even though they had only ever observed green activation from all four objects. This recognition required children to appreciate the uncertainty that C and D could do something other than make the machine turn red.

Speculations on how these findings inform our understanding of causal and scientific reasoning

There is a disagreement in the literature about when children are capable of engaging in complex causal reasoning tasks. For the most part, prior work on causal reasoning has suggested that young children have sophisticated and precocious reasoning abilities (see Gopnik & Wellman, 2012; Sobel & Legare, 2014, for reviews). In contrast, the scientific reasoning literature suggests that only by age seven or eight do children start to form explicit hypotheses, engage in evidence assessment, or draw conclusions from data (e.g., Kuhn & Pearsall, 2000; Ruffman, Perner, Olson, & Doherty, 1993; Sodian, Zaitchik, & Carey, 1991). An open question concerns whether these sets of capacities are linked.

The results of the present experiments suggest that this ability is beginning to develop between ages five and eight. We speculate that one step toward reconciling the differences between the causal reasoning literature (which argues for early success at diagnostic inference) and the scientific thinking literature (which argues for a more protracted developmental process) is to demonstrate that one facet that influences children's performance is the nature of the task, specifically the number and type of causes with unknown efficacy. Tasks used to demonstrate early success at causal reasoning typically present systems in which the efficacy of all causes are known, whereas tasks used in the scientific thinking literature often present systems in which students must infer the efficacy of some causes without the ability to test each candidate cause independently (as in our Study 2).

However, we do not want to conclude that uncertainty is the only difference between measures of causal and scientific reasoning. We suggest that the presence or absence of a scientific context also plays a major role. The present experiments were designed to be a simple and decontextualized investigation of diagnostic reasoning and, as such, to provide us with a first estimate of when children develop these abilities. Contextualizing causal inference problems can even lead adults to have difficulty with reasoning, often because intuitive theories interact with the interpretation of extant data. For example, Shtulman and Schulz (2008) suggested that one reason adults have such difficulty reasoning about certain evolutionary concepts is that they conflict with intuitive essentialist theories (as described in Gelman, 2003; see other examples in Shtulman, 2017).

The current results are only a first attempt at bridging the gap between the building blocks of scientific reasoning, such as making appropriate diagnostic inferences given data constructed by another person, and the more explicit scientific inferences that mature scientists make, such as engaging in experimental design and theory formation. That said, we posit that the underlying capacities necessary for successful performance in the present experiment and some scientific investigations are similar in structure. Learning about an additive causal model, like the one used in Study 2, is typically used in measures of scientific reasoning. But another critical difference between the present studies and previous investigations of scientific reasoning is that we designed experiments for children and did not require children to construct data themselves, something that is clearly difficult for elementary-school children (e.g., Chen & Klahr, 1999). Thus, blinket detector tasks, like those used here, usually test only a single metric of children's ability to evaluate observed data. True scientific thinking requires more than just the kind of diagnostic reasoning we measure here (see, e.g., Schauble, Glaser, Duschl, Schulze, & John, 1995), and there is usually not a knowledgeable informant telling scientists that their initial inference is incorrect or specifically designing experiments for them.

Further, an important limitation of the present research is that we examined only one kind of uncertainty: the extent to which children know whether candidate causes independently are efficacious. But, as mentioned in the Introduction, there are other types of uncertainty. For example, we assumed that children inferred a deterministic relation between objects and the machine, based on numerous investigations consistent with this hypothesis (e.g., Bullock et al., 1982; Griffiths et al., 2011; Kushnir & Gopnik, 2005; Sobel & Munro, 2009). But in many real-world investigations, stochastic causal relations are common. Estimating signal from noise is another critical step toward mature scientific and everyday diagnostic reasoning (e.g., Masnick & Klahr, 2003). In general, the present data only test children's reasoning about one form of uncertainty, and how children integrate other forms of uncertainty into diagnostic inference is an open question.

To conclude, the present results suggest that it is not until after the preschool years that children can generate error-free responses in diagnostic reasoning tasks that feature candidate causes with unknown efficacy. We suggest that this capacity to engage in diagnostic reasoning about multiple uncertain potential causes is necessary for belief revision in numerous real-world contexts, including certain types of scientific investigations. Children's early successes with causal reasoning suggest that they have the potential to be "scientists in the crib" (Gopnik, Meltzoff, & Kuhl, 1999). But, like practicing scientists, their abilities to diagnose the appropriate causal structure from evidence and reason about all the possible causes of an event develop over time.

Acknowledgments

We would like to thank all of the parents and children who participated in this research. We wish to thank Allison Chernov, Julia Franckh, Natalie Kleeman, Deanna Macris, Andrea Wister, and Rachel Zolno for help with data collection and analysis and Phil Fernbach for insightful discussion about this topic.

Funding

This research was supported by NSF1223777.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

ORCID

Deena Skolnick Weisberg  <http://orcid.org/0000-0002-4000-4941>

References

- Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, 77, 413–426. doi:10.1111/cdev.2006.77.issue-2
- Bindra, D., Clarke, K. A., & Shultz, T. R. (1980). Understanding predictive relations of necessity and sufficiency in formally equivalent "causal" and "logical" problems. *Journal of Experimental Psychology: General*, 109, 422–443. doi:10.1037/0096-3445.109.4.422
- Bonawitz, E. B., Fischer, A., & Schulz, L. E. (2011). Teaching 3.5-year-olds to revise their beliefs given ambiguous evidence. *Journal of Cognition and Development*, 13(2), 266–280. doi:10.1080/15248372.2011.577701
- Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development*, 82, 2053–2066. doi:10.1111/cdev.2011.82.issue-6
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York, NY: Academic Press.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70, 1098–1120. doi:10.1111/cdev.1999.70.issue-5
- Cowan, N. (1997). *The Development of Memory in Childhood*. New York, NY: Psychology Press.
- Denison, S., & Xu, F. (2010). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science*, 34, 885–908. doi:10.1111/cogs.2010.34.issue-5
- Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5-to 7-year-old children. *Journal of Experimental Child Psychology*, 126, 213–228.
- Erb, C. D., & Sobel, D. M. (2014). The development of diagnostic reasoning about uncertain events between ages 4–7. *PLoS One*, 9(3), e92285. doi:10.1371/journal.pone.0092285
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, 140, 168–185. doi:10.1037/a0022100
- Fernbach, P. M., Macris, D. M., & Sobel, D. M. (2012). Which one made it go? The emergence of diagnostic reasoning in preschoolers. *Cognitive Development*, 27, 39–53. doi:10.1016/j.cogdev.2011.10.002
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford.
- Gopnik, A., Meltzoff, A. N., & Kuhl, P. (1999). *The scientist in the crib*. New York, NY: Harper Collins.
- Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, 71, 1205–1222. doi:10.1111/cdev.2000.71.issue-5
- 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, 620–629. doi:10.1037/0012-1649.37.5.620
- Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138, 1085–1108.

- Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes andblickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, 35, 1407–1455. doi:[10.1111/cogs.2011.35.issue-8](https://doi.org/10.1111/cogs.2011.35.issue-8)
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107, 9066–9071. doi:[10.1073/pnas.1003095107](https://doi.org/10.1073/pnas.1003095107)
- Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science*, 25, 1768–1776.
- Hong, L., Chijun, Z., Xuemei, G., Shan, G., & Chongde, L. (2005). The influence of complexity and reasoning direction on children's causal reasoning. *Cognitive Development*, 20, 87–101. doi:[10.1016/j.cogdev.2004.11.001](https://doi.org/10.1016/j.cogdev.2004.11.001)
- Klahr, D. (2000). *Exploring science*. Cambridge, MA: MIT Press.
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, 25, 111–146. doi:[10.1006/cogp.1993.1003](https://doi.org/10.1006/cogp.1993.1003)
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Kuhn, D. (2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education*, 91, 710–726. doi:[10.1002/\(ISSN\)1098-237X](https://doi.org/10.1002/(ISSN)1098-237X)
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- Kuhn, D., & Dean, J. D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16(11), 866–870. doi:[10.1111/j.1467-9280.2005.01628.x](https://doi.org/10.1111/j.1467-9280.2005.01628.x)
- Kuhn, D., Garcia, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, 245, 60.
- Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, 1, 113–129. doi:[10.1207/S15327647JCD0101N_11](https://doi.org/10.1207/S15327647JCD0101N_11)
- Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, 9, 285–327. doi:[10.1207/s1532690xci0904_1](https://doi.org/10.1207/s1532690xci0904_1)
- Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, 16(9), 678–683. doi:[10.1111/j.1467-9280.2005.01595.x](https://doi.org/10.1111/j.1467-9280.2005.01595.x)
- Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers' use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, 43, 186. doi:[10.1037/0012-1649.43.1.186](https://doi.org/10.1037/0012-1649.43.1.186)
- Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, 21, 1134–1140. doi:[10.1177/0956797610376652](https://doi.org/10.1177/0956797610376652)
- Lehrer, R., & Schauble, L. (2000). The development of model-based reasoning. *Journal of Applied Developmental Psychology*, 21, 39–48. doi:[10.1016/S0193-3973\(99\)00049-0](https://doi.org/10.1016/S0193-3973(99)00049-0)
- Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, 131, 284–299. doi:[10.1016/j.cognition.2013.12.010](https://doi.org/10.1016/j.cognition.2013.12.010)
- Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, 34, 113–147. doi:[10.1111/j.1551-6709.2009.01058.x](https://doi.org/10.1111/j.1551-6709.2009.01058.x)
- Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four-to eight-year-old children. *Neuropsychologia*, 36, 273–293.
- Macris, D. M., & Sobel, D. M. (2017). The role of evidence diversity and explanation in 4-and 5-year-olds' resolution of counterevidence. *Journal of Cognition and Development*, 18(3), 358–374.
- Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development*, 4(1), 67–98. doi:[10.1080/15248372.2003.9669683](https://doi.org/10.1080/15248372.2003.9669683)
- 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. doi:[10.1037/a0017408](https://doi.org/10.1037/a0017408)
- Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science*, 3(4), 389–396. doi:[10.1111/desc.2000.3.issue-4](https://doi.org/10.1111/desc.2000.3.issue-4)

- Perner, J. (1979). Discrepant results in experimental studies of young children's understanding of probability. *Child Development*, 50, 1121–1127. doi:[10.2307/1129339](https://doi.org/10.2307/1129339)
- Ruffman, T., Perner, J., Olson, D., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617–1636. doi:[10.2307/1131459](https://doi.org/10.2307/1131459)
- Samarapungavan, A. (1992). Children's judgments in theory choice tasks: Scientific rationality in childhood. *Cognition*, 45, 1–32. doi:[10.1016/0010-0277\(92\)90021-9](https://doi.org/10.1016/0010-0277(92)90021-9)
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49(1), 31–57. doi:[10.1016/0022-0965\(90\)90048-D](https://doi.org/10.1016/0022-0965(90)90048-D)
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32, 102–119. doi:[10.1037/0012-1649.32.1.102](https://doi.org/10.1037/0012-1649.32.1.102)
- Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, 4(2), 131–166. doi:[10.1207/s15327809jls0402_1](https://doi.org/10.1207/s15327809jls0402_1)
- Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, 43, 1124. doi:[10.1037/0012-1649.43.5.1124](https://doi.org/10.1037/0012-1649.43.5.1124)
- Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, 40, 162. doi:[10.1037/0012-1649.40.2.162](https://doi.org/10.1037/0012-1649.40.2.162)
- Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, 10, 322–332.
- Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York, NY: Basic Books.
- Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, 32(6), 1049–1062. doi:[10.1080/03640210801897864](https://doi.org/10.1080/03640210801897864)
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, 47, 1(1, Serial No. 194). doi:[10.2307/1165893](https://doi.org/10.2307/1165893)
- Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, 42, 1103–1115. doi:[10.1037/0012-1649.42.6.1103](https://doi.org/10.1037/0012-1649.42.6.1103)
- Sobel, D. M., & Legare, C. (2014). Causal learning in children. *WIRE-Cognitive Science*, 5, 413–427. doi:[10.1002/wcs.1291](https://doi.org/10.1002/wcs.1291)
- Sobel, D. M., & Munro, S. E. (2009). Domain generality and specificity in children's causal inference about ambiguous data. *Developmental Psychology*, 45, 511–524.
- Sobel, D. M., Sommerville, J. A., Travers, L. V., Blumenthal, E. J., & Stoddard, E. (2009). Preschoolers' use of others' beliefs to make causal inferences from probabilistic data. *Journal of Cognition and Development*, 10, 262–284. doi:[10.1080/15248370903389416](https://doi.org/10.1080/15248370903389416)
- 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, 303–333.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62, 753–766. doi:[10.2307/1131175](https://doi.org/10.2307/1131175)
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, 115, 155–185. doi:[10.1037/0033-295X.115.1.155](https://doi.org/10.1037/0033-295X.115.1.155)
- Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *The Quarterly Journal of Experimental Psychology*, 48, 108–124.
- Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, 25(1), 161–169. doi:[10.1177/0956797613502983](https://doi.org/10.1177/0956797613502983)
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47, 1220–1229. doi:[10.1037/a0024023](https://doi.org/10.1037/a0024023)
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, 105(13), 5012–5015. doi:[10.1073/pnas.0704450105](https://doi.org/10.1073/pnas.0704450105)