



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Cognitive Development



Which one made it go? The emergence of diagnostic reasoning in preschoolers

Philip M. Fernbach^{a,*}, Deanna M. Macris^b, David M. Sobel^b^a University of Colorado, United States^b Brown University, United States

ARTICLE INFO

Keywords:

Diagnostic reasoning
Probability
Uncertainty
Explanation
Causal reasoning
Inference
Abduction

ABSTRACT

We evaluate the hypothesis that children's diagnostic causal reasoning becomes more sophisticated as their understanding of uncertainty advances. When the causal status of candidate causes was known, 3- and 4-year-olds were capable of diagnostic inference (Experiment 1) and could revise their beliefs when told their initial diagnosis was incorrect (Experiment 2). In Experiments 3 and 4, only 4-year-olds made successful inferences when the causal status of candidate causes was uncertain. The results suggest that by age 3, children appreciate that an effect can have multiple candidate causes, but it is not until age 4 that they begin to reason correctly when the causal status of candidate causes is unknown.

© 2011 Elsevier Inc. All rights reserved.

Achieving a goal requires selecting an effective intervention. A doctor must choose a course of treatment, an auto mechanic must decide what to fix, and governments must institute public policy initiatives. All of these cases call for diagnostic inference, reasoning backward from an outcome (e.g., a cough, a noisy engine, a rise in violent crime) to its likely cause or causes (e.g., tuberculosis, a broken fan belt, a decrease in the price of narcotics). Once a cause has been identified with some measure of certainty, one can make predictions about the effectiveness of a given intervention.

Diagnostic inference is aimed at identifying the best cause among a set of possibilities. Philosophers and computer scientists refer to such inferences as 'abduction' or 'inference to the best explanation' (Harman, 1965; Josephson & Josephson, 1994; Lipton, 2001; Peirce, 1965). This form of inference is particularly challenging because it requires the reasoner to seek out and represent the set of possible causes. The search for candidate causes is difficult because it is 'global' in the sense that relevant

* Corresponding author at: University of Colorado, Leeds School of Business, 419 UCB, Boulder, CO 80309-0419, United States.
Tel.: +1 303 492 1311.

E-mail address: philip.fernbach@gmail.com (P.M. Fernbach).

considerations might be completely independent of the current discourse context (Fodor, 2000). Peirce (1965) describes abduction, therefore, as the only form of inference that “introduces something new (p. 171).” Representing and reasoning over multiple uncertain possibilities also requires substantial resources for causal inference. To make good judgments, one must understand the causal structure of the scenario, maintain a representation of multiple causal candidates, and weigh them against one other in a way that is sensitive to their base-rates and causal efficacy.

Studies with adults show that people do eventually come to these capabilities; they make diagnostic judgments of likelihood by using causal knowledge to set up an appropriate mental model (Fernbach, Darlow, & Sloman, 2011; Waldmann & Holyoak, 1992) and they evaluate the model by retrieving context-specific information from memory (Thomas, Dougherty, Sprenger, & Harbison, 2008). These abilities, however, are not easily acquired. A general finding of research on scientific reasoning is that school-age children often struggle to recover candidate causes from data (Klahr & Nigam, 2004; Kuhn, Garcia, Zohar, & Andersen, 1995; Kuhn, Pease, & Wirkala, 2009; Lehrer & Schauble, 2000; Ruffman, Perner, Olson, & Doherty, 1993; Sodian, Zaitchik, & Carey, 1991). Even 10-year-olds struggle with such inferences compared to adults (e.g., Schauble, 1996). This research suggests that children’s diagnostic reasoning is prone to many heuristics and biases (Schauble, 1990) and over-reliance on pre-existing beliefs at the expense of observed data.

In stark contrast to these results is a substantial literature suggesting that preschoolers possess sophisticated diagnostic reasoning abilities. For instance, Bullock, Gelman, and Baillargeon (1982) found that 3-year-olds used factors like temporal priority and spatial proximity to diagnose the cause of an event. Shultz (1982) demonstrated that children made such inferences on the basis of their mechanism knowledge (see also Buchanan & Sobel, 2011). Similarly, preschoolers can diagnosis whether objects have hidden properties based on their causal power, as opposed to other potential bases for such inference, such as perceptual similarity (Gottfried & Gelman, 2005; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007). Young children can also resolve ambiguous information as to which of two causes produced an effect by appealing to external information like the base-rate of events having such causal efficacy, their knowledge of the functional form of a causal relation and their existing knowledge of the specific causal mechanism (Griffiths, Sobel, Tenenbaum, & Gopnik, 2011; Kushnir & Gopnik, 2007; Lucas, Gopnik, & Griffiths, 2010; Schulz, Bonawitz, & Griffiths, 2007; Sobel & Munro, 2009; Sobel, Tenenbaum, & Gopnik, 2004).

What might explain such divergent findings? Some suggestive results come from a study in which children were given a relatively simple diagnostic reasoning task. Sodian et al. (1991) asked children to infer the size of a mouse (big or small) from evidence that was either ambiguous or conclusive for one of the options. They found that 6-year-olds often struggled with this diagnosis (although older children usually did succeed). Why should these children struggle in this case, but younger children succeed in others cases of diagnosis? Critically, in the Sodian et al. procedure, one pattern of evidence was conclusive while the other pattern was not. The presence of such uncertainty might have affected children’s inferences. In contrast, in the cases of preschoolers’ successful inferences, children are usually shown a small, exclusive set (usually two) of candidate causes, the causal relations children observe are deterministic, and there are no hidden or unknown causes. Children simply have to choose which of the alternatives produced the effect.

We hypothesize that what determines performance on diagnostic reasoning tasks is the requisite representational requirements for reasoning over alternative possibilities. Even young children will succeed when those requirements are small, for instance, when potential causes are readily available and unambiguous. In contrast, greater difficulty will emerge with diagnostic inferences that involve events that are not present or whose efficacy is unknown. Success under those conditions requires a more sophisticated understanding of uncertainty and broader thinking about possible causes.

Here we focus on two related abilities that must be present for successful diagnostic inference. First, children must understand that an observed event could have been brought about by more than one cause and that belief should be spread over the candidates. We refer to this kind of understanding as *first-order diagnostic uncertainty*. To illustrate, consider a doctor evaluating a patient who presents with a rash and reports being exposed to poison ivy and eating some bad shellfish. The doctor might choose which event is a more likely cause of the rash, but retain the other as a possibility. This would require an understanding of first-order diagnostic uncertainty. One measure of whether a child understands

first-order diagnostic uncertainty is success in belief revision. The ability to shift to a new hypothesis in the face of contrary evidence suggests that the child is not ‘stuck’ on a single hypothesis.

A more complex kind of understanding is that causes that are themselves uncertain can be candidates. That is, even if a cause is not definitively present, it could be the best explanation for the outcome. We refer to this kind of understanding as *second-order diagnostic uncertainty*. In the example above if the patient presented with a rash and may or may not have been exposed to poison ivy or bad shellfish, the doctor might still infer those are the best explanations for the rash. This would require an understanding of second-order diagnostic uncertainty.

It is important to note that understanding first and second-order diagnostic uncertainty is necessary but certainly not sufficient to make normative diagnostic inferences. In addition, a reasoner must consider the functional relation between causes and effects (e.g., conjunctive vs. disjunctive causes), base-rates of causes, and causal efficacy (i.e., when a cause is known to be only probabilistically related to an effect). Since our focus is on first and second-order diagnostic uncertainty, our experiments do not manipulate these additional factors.

There is some evidence that in cases where children must choose among multiple candidate causes, preschoolers have some understanding of first-order diagnostic uncertainty. For instance, 3–4-year-olds can judge whether a particular outcome will occur given a change to a candidate to the stage of a candidate cause, which is either efficacious or irrelevant to the occurrence of the outcome (Buchanan & Sobel, 2011; Bullock et al., 1982). Similarly, young children can reason about counterfactual inferences that involve predicting whether a specific outcome will occur given a hypothetical change to an antecedent event (Beck, Robinson, Carroll, & Apperly, 2006; German & Nichols, 2003; Sobel et al., 2004). In contrast, the literature suggests that firm understanding of second-order diagnostic uncertainty develops later. Beck et al. (2006) and Beck, Riggs, and Gorniak (2009) found that 4-year-olds struggled with what they called “open-ended counterfactuals,” cases in which they had to prepare for all the possible outcomes of an uncertain event; 6-year-olds, in contrast, had no trouble with these inferences. This evidence suggests that diagnostic reasoning in which children must recognize that an event whose efficacy is uncertain might be the best explanation for an outcome might be difficult or just beginning to emerge in preschoolers.

In the present experiments, children were taught about blocks that could potentially activate a novel machine and were subsequently asked to identify which block was the cause when the machine activated behind an occluder. By manipulating whether children were asked to revise their belief and whether the efficacy of the candidate causes was known, we were able to vary whether understanding first and second-order diagnostic uncertainty was required to make successful inferences. Fig. 1 depicts the experimental setting.

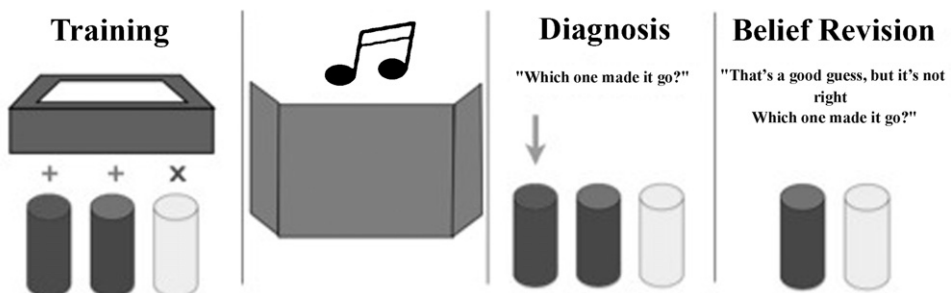


Fig. 1. Depiction of basic method for Experiments 1–4. Children observe blocks placed on the machine individually. Some activate it, some do not (Panel 1). In Experiments 1 and 2, all blocks are placed on the machine. In Experiments 3 and 4, one block's efficacy is uncertain. Children are asked to recapitulate the efficacy they observed (prediction questions). For diagnosis, a screen is introduced, and an object is used to activate the machine (Panel 2). Children are asked which object was used to activate the machine (Panel 3). In belief revision experiments (Experiments 2–4, Panel 4), children were told their original answer was incorrect, that block was removed, and children were asked again which of the remaining blocks were used to activate the machine.

1. Experiment 1

Three- and 4-year-olds were shown a machine that activates when objects are placed on it. They observed as three blocks were placed separately on the machine, and, depending on condition, either one or two of the three blocks activated the machine. After this training, the machine was occluded and then activated such that the child could not see which block was on it. The child was asked to guess which block had been used to activate the machine. This most basic kind of diagnostic test does not require the child to understand uncertainty at all because the child needs to only keep one effective block in mind. The first experiment was meant primarily to establish the method, and we expected both 3- and 4-year-olds to succeed on the task.

We also included predictive inference trials in which children were asked if the blocks would activate the machine if they were placed on it again. Asking predictive questions served as a control to verify that any diagnostic errors were not due to failures of memory. Success on the predictive task suggests that the child remembers which blocks are and are not effective. Although performance on prediction and diagnostic measures is often compared in adults (Fernbach, Darlow, & Sloman, 2010; Fernbach et al., 2011a; Waldmann & Holyoak, 1992), our diagnostic and predictive questions are in a different format. As a result, we do not compare them directly, and our data only provide suggestive evidence regarding the relative difficulty of prediction and diagnosis. We do, however, speculate on this issue in the general discussion.

1.1. Method

1.1.1. Participants

Sixteen 3-year-olds (8 girls, M age = 40.10 months, range 37–44 months) and sixteen 4-year-olds (5 girls, M age = 53.60 months, range 48–59 months) were recruited from a list of hospital births, flyers posted at local preschools, or at a local children's museum. All children were tested in the laboratory or in a quiet room at the museum. Most children were Caucasian and from middle to upper-middle class families; however, no specific indicators of SES were obtained.

1.1.2. Materials

The machine was 5" × 7" × 3", made of wood (painted gray) with a red lucite top. The machine worked via a remote control (hidden from the child), which the experimenter used to control whether an object activated it. When the machine was active, it lit up red and played *Fur Elise* continuously. When a block was effective, the experimenter activated the machine as soon as the block made contact with it and turned it off as soon as the block was removed. This provided a strong impression that something about the block caused the machine to activate.

Six cylindrically shaped blocks of six different colors were grouped into two sets of three (Fig. 1). Additionally, a piece of cardboard (approximately 2' × 2') was used to occlude the detector from the child when demonstrating a diagnostic event.

1.1.3. Design and procedure

The two independent variables were direction of inference (*diagnostic or predictive*) and number of alternative causes (*one-cause* condition or *two-cause* condition), manipulated within participant. Participants therefore provided responses to four kinds of trials.

Children were tested by one of two experimenters (1 male, 1 female), with whom they were familiarized through a short free-play period. Children sat facing the experimenter across a table. The experimenter placed the machine on the table and introduced it to the child by saying, "This is my machine. Some things make it go and some things don't." The experimenter (E) then placed three blocks on the table in front of the machine (Fig. 1).

Blocks were placed on the machine one at a time for approximately 3 s, starting with the block on E's left. In the one-cause condition, one of the blocks activated the machine and two failed to do so. In the two-cause condition two blocks activated the machine and one failed to do so. The spatial location of effective and ineffective blocks was randomized. After demonstrating each block, E repeated the

Table 1

Percentage of diagnostic trials without errors by age in Experiments 1 and 2.

	Experiment 1 (chance = 50.0%)	Experiment 2 (chance = 33.3%)
3-year-olds	93.8%*	70.7%*
4-year-olds	90.1%*	93.8%*

Notes: Chance performance was determined by computing the probability of at least one error in a trial assuming uniform choice over options. Significance with respect to chance is denoted by *.

demonstration. The child therefore either saw each block activate the machine twice or fail to do so twice.

Next occurred a diagnostic and a predictive test trial, in counterbalanced order across children. On diagnostic trials E placed the cardboard occluder in front of the blocks and the machine so that they were not visible to the child and said, "I'm going to put one of the blocks on the machine, so pay attention." E placed a block on the machine and activated the detector so that it played music for approximately 3 s. The child could not see this, but could hear the music playing. E then placed the block back in its original position, removed the occluder and asked the child, "Which of the blocks did I put on the machine?" The trial was completed when the child pointed at one of the three blocks or gave a verbal response. On predictive trials, E pointed to each of the three blocks in turn and asked the child, "If I put this block on the machine, will it make the machine go?" After obtaining a verbal yes or no response, E asked about the next block until the child had made a prediction about each block.

After completing both the diagnostic and predictive trials, E proceeded to the other condition (one-cause or the two-cause, depending on which had been completed first) by putting away the first set of blocks and bringing out a new set. Order of the one-cause and two-cause conditions was counterbalanced.

1.2. Results and discussion

To verify that children could remember which blocks were effective, we coded the frequency with which children responded correctly to all three of the prediction questions. Performance was close to ceiling; most children responded to all three questions correctly and a Binomial test revealed that performance was much greater than chance, $p < .001$, with chance performance at 12.5%. There were no significant differences between age groups, Fisher's exact test, $p > .10$.¹

Children were coded as making an error on the diagnostic trials if they chose an ineffective block as the cause (one possibility in the two-cause condition, two possibilities in the one-cause condition). Table 1 shows the percentage of error-free trials for each age group. Children's performance was error free on 59 of the 64 trials (92.2%), superior to chance performance (50.0%, calculated based on uniform choice over the blocks), Binomial test, $p < .001$. Three children (two 3-year-olds and one 4-year-old) made a diagnostic error on one trial each, and one child (a 4-year-old) made an error on both diagnostic trials.

Experiment 1 showed that both 3- and 4-year-olds could make a basic kind of diagnostic inference. The procedure did not allow us to assess whether the children considered and chose among multiple hypotheses, however, and did not indicate whether they had an understanding of first or second-order diagnostic uncertainty. In both the one and two-cause conditions, children could have succeeded by considering just a single cause. We designed Experiment 2 to assess whether children can make successful diagnostic inferences that require an understanding of first-order diagnostic uncertainty.

¹ Twenty-two of the 32 children responded correctly to the prediction questions on both the one-cause and two-cause trials. Six children (three 3-year-olds and three 4-year-olds) responded 'yes' to all of the prediction questions across both the one-cause and two-cause conditions. Three children (two 3-year-olds and one 4-year-old) responded 'yes' to all the blocks in the one-cause-condition, but responded correctly in the two-cause condition. Only one child (a 3-year-old) predicted that an effective block would fail to activate the detector.

2. Experiment 2

We modified the diagnostic trials from Experiment 1 such that after the child made a guess, they were told that their guess was wrong and were asked to choose a different block. We only presented children with trials like the two-cause condition of Experiment 1, so that there was always a viable alternative cause. If children can shift to a novel hypothesis, it suggests that they understand first-order diagnostic uncertainty (i.e., that the other effective cause could also be a candidate). If they do not possess this understanding, they should choose at random between the effective and ineffective block when asked to revise belief.

2.1. Method

2.1.1. Participants

Seventeen 3-year-olds (10 girls, M age = 40.59 months, range 36–47 months) and sixteen 4-year-olds (8 girls, M age = 53.38 months, range 48–60 months) were recruited from the same population as in Experiment 1 and tested in the laboratory or in a quiet room at the Children's Museum. One additional child was tested but not included in the analysis because of an error in administering the procedure.

2.1.2. Materials

The materials were identical to those used in Experiment 1.

2.1.3. Design and procedure

The design was similar to Experiment 1 except that there were always two effective blocks and diagnostic and predictive inferences were elicited after separate training. Each child therefore completed one predictive and one diagnostic trial with separate sets of blocks. The order of predictive and diagnostic trials was counterbalanced.

The procedure was also similar to Experiment 1. The key difference was that on diagnostic trials, after the child made a guess, E removed the chosen block from the table and responded, "That's a good guess, but actually it's not right. That's not the one I put on the machine. Can you tell me which one I did put on the machine?" The trial concluded when the child chose one of the two remaining blocks.

2.2. Results

As in Experiment 1, we coded the frequency with which children responded correctly to all three prediction questions. Performance was again near ceiling.² Performance on diagnostic trials is shown in Table 1. Twelve of seventeen 3-year-olds picked the two efficacious blocks, which was better than chance, Binomial test, $p < .05$. Only one 4-year-old erred on the diagnostic questions, which was also above chance, $p < .001$. Four-year-olds performed marginally better than 3-year-olds, Fisher's exact test, $p = .10$, one tailed.

2.3. Discussion

Experiment 2 differed from Experiment 1 in that diagnostic trials required the child to shift from their initial hypothesis and think of the other effective block as an alternative explanation. Both age groups performed well above chance. This shows that by age 3 children can revise their beliefs in the face of contrary evidence. Four-year-olds performed near ceiling while 3-year-olds did make some errors. The difference between the age groups was marginally significant with a one-tailed test, providing some evidence for a developmental difference. But, taken as a whole, the results suggest that

² Thirteen of the seventeen 3-year-olds made no errors on the prediction trial, greater than chance, Binomial test, $p < .001$. All of the errors made by these children were cases in which the child responded 'yes' to all questions. Four-year-olds never erred on the prediction questions.

by age 3 most children understand first-order diagnostic uncertainty in the context of diagnostic reasoning.

3. Experiment 3

In Experiment 3 we asked whether children are capable of diagnostic inferences that require understanding second-order diagnostic uncertainty. We extended the Experiment 2 procedure by asking preschoolers to make diagnostic inferences about blocks whose efficacy was unknown. After training with the first three blocks as in Experiments 1 and 2, E introduced a fourth, novel block whose efficacy was not demonstrated. This created uncertainty in the set of potential causes, making an understanding of second-order diagnostic uncertainty necessary for success on the diagnostic trials. We also reintroduced the one-cause vs. two-cause manipulation from Experiment 1 because adding the novel block provided an alternative cause even in the one-cause condition. That is, in the one-cause condition, children had to keep in mind an uncertain cause and reason that it was a more likely cause than the ineffective blocks. In the two-cause condition, children had to keep a second effective block and the uncertain block in mind as candidate causes. Contrary to Experiment 2, we speculated that preschoolers might perform poorly based on evidence reviewed above that second-order diagnostic uncertainty is relatively late to develop.

3.1. Method

3.1.1. Participants

Fifteen 3-year-olds (8 girls, $M = 39.85$ months, range 36–47 months) and eighteen 4-year-olds (7 boys, 11 girls, $M = 54.67$ months, range 49–60 months) were recruited from birth records and tested in the laboratory. They were from a similar population to that drawn from in Experiments 1 and 2. An additional three children were tested but not included in the analysis because of errors in administering the procedure.

3.1.2. Materials

Materials were identical to those used in Experiment 1 except that there were now four groups of four blocks each. The blocks within each group were the same shape but different colors. Each group was a different shape and all 16 blocks were different colors.

3.1.3. Design and procedure

The design was the same as in Experiment 1 except that the predictive and diagnostic trials were separated into different conditions with different block sets and training phases. The predictive and diagnostic trials for a given number of causes always occurred one after the other. The order of the one-cause vs. two-cause conditions was counterbalanced, as was the order of predictive and diagnostic trials.

The procedure for the training phases was similar to that used in Experiments 1 and 2 except that after demonstrating the three blocks, E brought out a fourth block and placed it on the table to the right of the other blocks. E then proceeded with a predictive or diagnostic trial. Predictive trials were the same as in Experiments 1 and 2 except that E also asked a predictive question about the novel block. Diagnostic trials were similar to Experiment 2 except that the child's first choice was among four blocks instead of three and the second choice – after removal of the block chosen first – was among three blocks rather than two.

3.2. Results

Performance on predictive trials was again near ceiling for both age groups.³ Diagnostic trials were coded as errors if the child responded with an ineffective block on the first guess, the second guess, or

³ Collapsing across the two age groups, performance on prediction trials was better than chance, Binomial test, $p < .001$. This was also true when the 3- and 4-year-olds were analyzed separately, both p -values $< .001$. When children erred on the

Table 2
Percentage of diagnostic trials without errors by condition and age in Experiment 3.

	One-cause (chance = 16.7%)	Two-causes (chance = 50.0%)	Both conditions (chance = 33.3%)
3-year-olds	13.3%	66.7%	40.0%
4-year-olds	44.4*	55.6%	50.0*

Notes: Chance performance was determined by computing the probability of at least one error in a trial assuming uniform choice over options. Significance with respect to chance is denoted by *.

both. The data are shown in Table 2. Collapsing across age groups, children responded without error on 44% of the diagnostic trials, just higher than chance performance (33%), Binomial test, $p < .05$. However, when the two age groups were analyzed separately, only 4-year-olds exceeded chance performance on the diagnostic trials (50%), Binomial test, $p < .05$. Three-year-olds responded at chance levels (40%), Binomial test, *ns*. Three-year-olds and 4-year-olds, however, did not differ on overall performance on diagnostic trials when compared to each other, Fisher's exact test, $p > .10$.

3.2.1. One-cause vs. two-cause condition

To assess the effect of number of alternative causes, we examined each condition separately. Table 2 shows the percentage of correct diagnostic trials by condition for each age group. Overall, performance was relatively poor in both conditions. However, performance in the one-cause condition was slightly above chance, Binomial test, $p < .05$. This difference was driven by above-chance performance by the 4-year-olds, Binomial test, $p < .01$. Three-year-olds did not respond differently from chance in the one-cause condition and made more errors in this condition than the 4-year-olds, Fisher's exact test, $p < .05$. In contrast, both age groups responded at chance levels on the two-cause trials. Performance on the prediction questions did not vary as a function of number of causes; there were more errors in the two-cause condition (20% of trials) than the one-cause condition (10% of trials), but this difference was not significant.

To understand the differential diagnostic performance between the one-cause and two-cause conditions, we looked at whether the diagnostic errors in each condition were made in response to the first or second question. In the two-cause condition, children were more likely to make an error on the first question (9 of the 13 errors were on the first question) while in the one-cause condition errors on the second guess were more common (18 of 23 errors were on the second question). This difference was significant, Fisher's exact test, $p = .01$. We speculate about this difference in the discussion.

3.2.2. Novel block choice

One might ask whether children's tendency to make diagnostic errors reflects a failure to understand that the novel block could activate the machine. This is a particular concern in the one-cause condition where the novel block is the only viable alternative. At least two findings speak against this interpretation. First, children often chose the novel block in diagnostic trials; 22 of the 33 children chose the novel block at least once. This suggests that they did understand that it was a possible cause. The prevalence of novel block choices makes it unlikely that diagnostic failures resulted from a misunderstanding of the instructions. Participants treated the novel block similarly to the other blocks. Second, on the prediction trials, children were asked whether this block activated the machine (even though they had not seen its efficacy). Participants tended to predict that the novel block would activate the detector; this did not vary across the number of causes (26 of 33 trials in the two-cause condition and 28 of 33 in the one-cause condition), both response patterns greater than chance

prediction question, they often responded 'yes' to all questions across both conditions (three children, two 3-year-olds and one 4-year-old, showed this pattern of response). When their data were excluded, 3-year-olds were error-free on 24 of 30 trials (80%) 4-year-olds were error-free on 33 of 36 trials (92%) – response levels greater than chance performance, Binomial tests, both p -values < 0.001 . Although 4-year-olds were slightly better than 3-year-olds on predictive trials when all responses were included, $p < .05$, this difference disappeared when 'yes bias' responses were removed from the sample.

predictions (50%), Binomial tests, both p -values $< .001$. This suggests that if anything, there was a bias to interpret the novel block as effective in the absence of definitive evidence.

4. Discussion

In contrast to Experiments 1 and 2, the children made many diagnostic errors in Experiment 3. Three-year-olds performed at chance on the diagnostic questions in both the one-cause and two-cause conditions. Four-year-olds performed above chance in the one-cause condition but performed at chance in the two-cause condition. Children's success on predictive trials makes it unlikely that diagnostic errors were due to simple memory failures. Participants were able to remember which blocks were effective and which were ineffective, but they had trouble using that knowledge to their advantage in diagnostic inference. Diagnostic failures also were not due to misunderstanding the role of the novel block; participants treated it much like the other blocks.

An informative trend emerged by looking at whether diagnostic errors occurred on the first or second question in each diagnostic trial. On the one-cause trial, errors tended to occur on the second question. This suggests that both 3- and 4-year olds were capable of keeping the single effective cause in mind, but only 4-year-olds were able to then revise belief to the uncertain novel block. In other words, only 4-year-olds demonstrated an understanding of second-order diagnostic uncertainty. The fact that the 4-year-olds performed above chance – but not at ceiling as in the previous experiments – suggests that this capability is just beginning to emerge at this age.

On the two-cause trial, errors tended to occur on the first guess. This implies that the novel block interfered with children's ability to identify even a single good explanation. This difficulty emerged for both age groups suggesting that the presence of uncertainty combined with the requirement to consider more than one well-established cause leads to particular difficulty in diagnosis.

One concern with this interpretation is that the procedure used in Experiments 1 and 2 required children to keep track of the efficacy of three blocks while the procedure in Experiment 3 required them to keep track of the efficacy of four blocks. It is possible that the difficulty was driven by this increased information processing demand. Success on the predictive trials speaks against this interpretation. Moreover, in the one-cause condition, the absolute number of potential causes was the same as in Experiment 2 (i.e., two: the effective block and the novel block). Nonetheless in Experiment 4 we tested this alternative explanation directly.

5. Experiment 4

We modified the procedure used in Experiment 3 to ensure that the results from that experiment were due to the addition of second-order-uncertainty, and not due to an increase in the number of blocks used. Children completed four diagnostic trials in which we varied number of blocks (3 vs. 4) and the presence of uncertainty (all blocks known vs. one block unknown). If errors in Experiment 3 were due to the presence of uncertainty we should observe a main effect of uncertainty. Alternatively, if the errors were due to the number of blocks there should be a main effect of number of blocks, but no effect of uncertainty. Thus, we made the somewhat counterintuitive prediction that children would perform poorly under uncertain conditions even with only three blocks but could be successful with four blocks if the efficacy of each block was known. Also, because the unknown conditions required an understanding of second-order diagnostic uncertainty we expected 4-year-olds to outperform 3-year-olds on the uncertain trials.

5.1. Method

5.1.1. Participants

Sixteen 3-year-olds (8 girls, $M = 40.88$ months, range 36–47 months) and sixteen 4-year-olds (7 girls, $M = 52.56$ months, range 48–57 months) were recruited from the same population as in Experiment 1 and tested in the laboratory or in a quiet room at the Children's Museum. An additional three 3-year-old children were tested, but not included in the final analysis because of experimental error ($n = 1$), failure to engage with the experimenter ($n = 1$), and equipment failure ($n = 1$).

Table 3
Percentage of diagnostic trials without errors by condition and age in Experiment 4.

	3-blocks (chance = 33.3%)		4-blocks (chance = 16.7%)	
	Known	Unknown	Known	Unknown
3-year-olds	56.3%*	43.4%	62.5%*	25.0%
4-year-olds	87.5%*	68.8%*	75.0%*	75.0%*
Total	71.9%*	56.3%*	68.8%*	50.0%*

Notes: Chance performance was determined by computing the probability of at least one error in a trial assuming uniform choice over options. Significance with respect to chance is denoted by *.

5.1.2. Materials

Materials were identical to Experiment 3 except that there were two groups of three blocks each and two groups of four blocks each. As before, the blocks within each group were the same shape but different colors. Each group was a different shape and the 14 blocks were all different colors.

5.1.3. Design and procedure

The two independent variables were number of blocks (*three-block* condition or *four-block* condition) and presence of a block whose efficacy was unknown (*unknown* condition or *known* condition), manipulated within participants. Children therefore responded to four trials, which were counterbalanced. There were no predictive trials in this experiment.

The procedure for the training phases was similar to that used in Experiments 1–3, except that only two blocks were present on one type of trial (three-blocks with an unknown block), three blocks were present on two trials (three-blocks with no unknown block and four-blocks with an unknown block), and four blocks were present on one type of trial (four-blocks with no unknown block). On known trials, two blocks activated the machine, and on unknown trials, only one block activated the machine. For the unknown trials, the experimenter brought out an additional block after training and placed it to the right of the other blocks. The procedure for asking the diagnostic questions was the same as used in Experiments 2 and 3.

5.2. Results and discussion

We coded responses by counting any trial on which a child chose an ineffective block (on the first or second guess or both) as an error. The percentage of error-free trials by condition is shown in Table 3. Our critical prediction was that there would be a main effect of uncertainty but no effect of number of blocks and no interaction. A repeated measures analysis of variance, with number of blocks and presence of uncertainty as factors,⁴ confirmed our prediction, showing a significant main-effect of uncertainty, $F(1, 31) = 5.52, p < .05$, but no effect of number of blocks and no interaction, both F -values $< 1, ns$.

To analyze effect of age, we compared number of errors made by 3- and 4-year-olds. Three-year-olds made an average of 2.13 errors over the 4 trials while 4-year-olds made only 0.94 errors, a significant difference, $t(30) = 2.82, p < .01$, Cohen's $d = 1.03$. The partial correlation between number of errors and age (in months), controlling for effect of age in years, was negative and significant, $r = -0.48, p < .01$. Together these results show that 4-year-olds performed better than 3-year-olds and performance improved from the younger to older children within each age group.

To understand performance by age and condition further, we compared performance in each condition to chance for both ages (Table 3). Binomial tests revealed that 3-year-olds performed at chance on both unknown trials, both p -values $> .10$, but exceeded chance in both known conditions, both p -values $< .05$. Four-year-olds performed above chance in all conditions, all p -values $< .05$. This finding replicates the Experiment 3 finding that only 4-year-olds performed above chance in the one

⁴ Our dependent measure is dichotomous, but this analysis is appropriate given sufficient degrees of freedom (Lunney, 1970).

cause condition. These results provide more evidence that 4-year-olds have mastered second-order diagnostic uncertainty but 3-year-olds find it difficult.

6. General discussion

Across four experiments we explored the developmental trajectory of diagnostic inference. In Experiment 1, both 3- and 4-year olds succeeded on a simple diagnostic inference task that did not require any understanding of uncertainty. In Experiment 2, success in belief revision required mastery of first-order diagnostic uncertainty. Both 3- and 4-year olds performed above chance, but only 4-year-olds performed near ceiling and there was a marginal difference between the age groups. Taken together these results suggest that 3-year-olds have mastered first-order diagnostic uncertainty although there may still be some development occurring between the ages of 3 and 4.

The addition of uncertainty in the set of possible causes in Experiment 3 yielded many more diagnostic errors by both 3- and 4-year-olds. However, 4-year-olds performed above chance in the one-cause condition, suggesting that mastery of second-order uncertainty is beginning to emerge. Experiment 4 corroborated this result; 4-year-olds, but not 3-year-olds, were successful when the set of causes was uncertain. The memory demands of the task made no difference; both 3- and 4-year olds performed above chance when the set of causes was fully determined, regardless of how many candidate causes they had to track. The correlation between age and performance further suggested active development in the understanding of uncertainty during these ages. The two-cause condition of Experiment 3 was the only condition in any of the experiments in which 4-year-olds completely failed. Accomplishing diagnostic inference when there are multiple well-established alternatives *and* uncertainty appears to emerge later in development. This implies that even by age 4, a relatively simple diagnostic task can prove challenging. This last finding is potentially related to why many measures of scientific reasoning show particular inferences to be difficult for even school-aged children and adults (Kuhn et al., 2009), while causal reasoning studies find preschoolers quite capable of similar inferences (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Sobel et al., 2004). Future research should address this issue.

6.1. A proposed trajectory for diagnostic reasoning

Our results show that the capacity for diagnostic reasoning is constrained by the child's developing mastery of diagnostic uncertainty. The fact that both 3 and 4-year-olds perform at ceiling in Experiment 1 but that there was some evidence for a development difference in Experiment 2 suggests the possibility that children begin with a basic ability to reason diagnostically even absent an understanding of uncertainty. The idea is that before a child understands that multiple causes are candidates, he or she may be able to reason to a single cause that is inferred with certainty. This is an appropriate strategy only in the case that there is a single cause; it leads to overconfidence (and irrationality) otherwise. It provides a good starting point because it affords a basic kind of diagnostic inference with no requirement to reason about multiple possibilities. An obvious downside is that such an inference does not allow one to revise belief in the face of contrary evidence.

A more sophisticated capability emerges once the child has mastered first-order diagnostic uncertainty. This allows the child to reason appropriately when there is a small set of potential causes definitively known to be present. The child chooses among them while appreciating the possibility that an alternative cause could be responsible for the effect. This kind of inference requires understanding that multiple possible causes could produce an outcome, but the values and identities of each of those causes are known. Based on the result of Experiments 2 and 4, it is clear that by age 3 most children have this capacity.

Finally, once a child has mastered second-order diagnostic uncertainty, he or she will succeed on diagnostic inferences when the status of the possible alternative causes is uncertain. Experiments 3 and 4 suggest that 4-year-olds, but not 3-year-olds, have this capability in some form, though even at age 4, it is tenuous; not all 4-year-olds succeed and introducing a single additional known cause leads to chance-level performance (as in the two-cause condition of Experiment 3). Moreover, by age 4 the development of diagnostic causal reasoning is far from complete. Added complexity arises as the set

of causes is larger, when causes are probabilistic and when they interact with one another (Kuhn & Pease, 2008).

6.2. Diagnosis vs. prediction

One question that emerges from our studies is whether they shed light on the relative development of different kinds of inferences such as forward inferences (like predictions) and backward inferences (like diagnoses or explanations). Some have argued that the counterfactual inferences involved in diagnosis are more difficult for preschoolers than prediction about future hypotheticals (Riggs, Peterson, Robinson, & Mitchell, 1998) while others have suggested that children might come to possess certain explanatory abilities prior to their being able to engage in prediction (e.g., Bartsch & Wellman, 1989; Wellman & Liu, 2007).

A substantial literature on adult causal reasoning suggests that diagnostic reasoning is relatively difficult compared to prediction. Recent evidence suggests that one contributor to the ease of predictive reasoning is that people focus on the cause they have in mind and ignore the contribution of relevant alternatives to the likelihood of the effect (Fernbach et al., 2010; Fernbach, Darlow & Sloman, 2011). This finding fits into a much larger body of work that suggests that people reduce effort by focusing narrowly, displaying a sort of ‘cognitive myopia.’ For instance, people tend to base judgment on whatever comes to mind most easily (Tversky & Kahneman, 1973), and they test hypotheses by focusing on the truth of the hypothesis they have in mind (Klayman & Ha, 1987), neglecting relevant alternatives and displaying ‘pseudodiagnostic’ data selection (Doherty, Mynatt, Tweaney, & Schiavo, 1979). Such myopic tendencies are so prevalent that the failure to consider alternatives sufficiently is a basic assumption of descriptive theories of reasoning and judgment (Dawes, 2001; Evans, Over, & Handley, 2003). There are good reasons for myopia is human cognition. Thinking of alternatives is costly in time and resources, and myopic inference can often yield reasonable judgments with little effort. For instance, a “positive test strategy” turns out to be an optimal way to test hypotheses under fairly general conditions (Crupi, Tentori, & Lombardi, 2009; Klayman & Ha, 1987).

In contrast to these myopic cases, people tend to think more broadly when reasoning in a diagnostic context. Although people do not think of a fully comprehensive set of possible alternatives when reasoning diagnostically (Fischhoff, Slovic, & Lichtenstein, 1978), they do a good job of thinking of the most important ones without prompting. Moreover, adults’ diagnostic judgments are sensitive to other relevant causal variables, such as the base-rate of the focal cause (Fernbach et al., 2011a; Griffiths et al., 2011). Fernbach et al. (2010, 2011a) argued that this pattern emerges because diagnostic inference is inherently comparative. Ignoring alternative causes would result in incoherence. The downside of considering alternative causes is that diagnostic inference is relatively slow; reaction time studies suggest that the number of possible alternative causes directly affects response times (De Neys, Schaeken, & D’ydevalle, 2002; Fernbach & Darlow, 2010). Further, diagnostic inference is associated with metacognitive difficulty (Medin, Coley, Storms, & Hayes, 2003; Tversky & Kahneman, 1982). Taken together, these results suggest that diagnostic reasoning requires a relatively effortful cognitive process that operates over alternative causes. The connection to the present work is to show the developmental trajectory of these processes and representations. Moreover, it makes sense given the prevalence of myopic judgment among adults that children should begin by construing the diagnostic task too narrowly (e.g., ignoring relevant but ambiguous causes), and only come to a fuller capacity as the requisite concepts and computational abilities emerge.

The present studies also inform previous research in the developmental literature that has investigated the development of predictive and diagnostic reasoning. For instance, Bindra, Clarke, and Shultz (1980) asked preschoolers to predict the appearance of a light based on the position of two switches and to diagnose the position of the switches based on the light. Performance on their predictive questions was superior to their diagnostic questions in general. This is consistent with our findings in that the predictions did not require reasoning about uncertain causes at all. The status of the switches was known when children predicted the lights. Bindra et al. also asked children two kinds of diagnostic questions – ones that required understanding what we call first-order uncertainty only and ones that required understanding what we call second-order uncertainty. They found that 4-year-olds were more likely to answer the first kind of question correctly than the second, and that

the latter questions were quite difficult. Only 9–10-year-olds answered these questions reliably above chance levels. These results are also consistent with the proposed developmental trajectory because the difficulty of the questions increased as the required notion of uncertainty increased in complexity.

In a more recent study, Hong, Chijun, Xuemei, Shan, and Chongde (2005) assessed preschoolers' predictive and diagnostic reasoning using a variant of the 'ramp task' (Frye, Zelazo, & Palfai, 1995). They tested whether preschoolers could predict the outcome of a causal system, given particular rules (of varying complexity) governing the system's behavior, as well as whether they could infer the start state of the system given the same rules and the outcome of the system. They found that children performed better on predictive than diagnostic questions. Moreover, the rules they manipulated involved reasoning about a single possible alternative cause (i.e., requires no understanding of uncertainty) or another possible, known alternative cause (i.e., requires an understanding of first-order uncertainty). Three-year-olds were more accurate on the former inference than the latter, while 4-year-olds did not differ between these two types of inferences. These results also square with our findings that some development of competence in first-order diagnostic uncertainty is still ongoing at age 3.

Our conclusions also share similarity with recent work on children's developing counterfactual inferences about conditionals. Guajardo and Turley-Ames (2004) suggested that children's ability to generate multiple antecedents that would change a causal consequence in an open-ended counterfactual reasoning task improved between the ages of 3 and 5. Similarly, Beck et al. (2006, 2009) found that preschoolers are able to reason about counterfactual conditionals about known antecedents during the preschool years, but struggle preparing for all possible outcomes of an uncertain event. The development we observed in preschoolers' difficulty with diagnostic inference to uncertain alternatives parallels these findings on counterfactual conditional reasoning.

To conclude, the requirements of Experiments 3 and 4 are the closest we came to those faced by doctors, mechanics, and public policy officials, but they are clearly minimal in comparison. Candidate causes (even uncertain ones) were right in front of the children, and there was no ambiguity in whether a known cause was effective. In contrast, real world inference (medical diagnosis for instance) requires much broader thinking, more possible alternative causes, and a much greater level of uncertainty. We have highlighted just the first few steps on the road to fully competent adult inference. Nonetheless, we believe that these steps are vital because they represent general conceptual abilities that must undergird more sophisticated diagnostic inference.

Our data suggest that the development of diagnostic reasoning is constrained by the child's developing understanding of diagnostic uncertainty. The more basic inferences that develop earlier require less computational and representational resources and may be effective at fulfilling needs in early childhood. Such reasoning does not bear the hallmark of full diagnostic inference however; children must learn how to consider multiple hypotheses and reason about them even when they are uncertain. This capability is likely more useful when a child begins to reason about more complex causal systems and has more nuanced goals. It may only convey a benefit later in development.

Author's note

This work was supported by a Galner Dissertation Fellowship and an APA Dissertation Research Award to PMF and NSF grant DLS-0518161 to DMS. We are especially grateful to David Buchanan for assistance with data collection and for many helpful conversations.

References

- Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, 60, 946–964.
- 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.
- Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating development in children's counterfactual thinking and executive functions. *Thinking and Reasoning*, 15, 337–354.
- 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.
- Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development*, 82, 2053–2066.
- 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: Academic Press.

Please cite this article in press as: Fernbach, P. M., et al. Which one made it go? The emergence of diagnostic reasoning in preschoolers. *Cognitive Development* (2011), doi:10.1016/j.cogdev.2011.10.002

- Crupi, V., Tentori, K., & Lombardi, L. (2009). Pseudodiagnosticity revisited. *Psychological Review*, 116, 971–985.
- Dawes, R. M. (2001). *Everyday irrationality: How pseudoscientists, lunatics and the rest of us fail to think rationally*. Boulder, CO: Westview Press.
- De Neys, W., Schaeken, W., & D'ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory and Cognition*, 30, 908–920.
- Doherty, M. E., Mynatt, C. R., Treeney, R. D., & Schiavo, M. D. (1979). On pseudodiagnosticity. *Acta Psychologica*, 43, 111–121.
- Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2003). A theory of hypothetical thinking. In D. Hardman, & L. Maachi (Eds.), *The psychology of reasoning and decision making*. Chichester: Wiley.
- Fernbach, P. M., & Darlow, A. (2010). Causal conditional reasoning and conditional likelihood. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, 21, 329–336.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011a). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, 140, 168–185.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011b). When good evidence goes bad. The weak evidence effect in judgment and decision-making. *Cognition*, 119, 459–467.
- Fischhoff, B., Slovic, P., & Lichtenstein. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human, Perception, and Performance*, 4, 330–344.
- Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge: MIT Press.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*, 6, 514–523.
- 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.
- Gottfried, G. M., & Gelman, S. A. (2005). Developing domain-specific causal-explanatory frameworks: The role of insides and immanence. *Cognitive Development*, 20, 137–158.
- 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, 1407–1455.
- Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development*, 19, 53–80.
- Harman, G. (1965). The inference to the best explanation. *The Philosophical Review*, 74, 88–95.
- 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.
- Josephson, J. R., & Josephson, S. G. (1994). *Abductive inference: Computation, philosophy technology*. New York: Cambridge University Press.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661–667.
- Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228.
- Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction*, 26(4), 512–559.
- Kuhn, D., Garcia, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, 245(60), 1–157.
- Kuhn, D., Pease, M., & Wirkala, C. (2009). Coordinating the effects of multiple variables: A skill fundamental to scientific thinking. *Journal of Experimental Child Psychology*, 103, 268–284.
- 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, 186–196.
- Lehrer, R., & Schauble, L. (2000). The development of model-based reasoning. *Journal of Applied Developmental Psychology*, 21, 39–48.
- Lipton, P. (2001). *Inference to the best explanation*. London: Routledge.
- Lucas, C. G., Gopnik, A., & Griffiths, T. L. (2010). Developmental differences in learning the forms of causal relationships. In *Proceedings of the 32nd annual conference of the Cognitive Science Society*.
- Lunney, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Education Measurement*, 7, 263–269.
- Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, 10, 517–532.
- Peirce, C. S. (1965). Lectures on pragmatism. In C. Hartshorn, & P. Weiss (Eds.), *Collected papers of Charles Sanders Peirce* (pp. 14–212). Cambridge, MA: Harvard University Press (original work published 1931).
- Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13, 73–90.
- 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.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49, 31–57.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32, 102–119.
- Schulz, L. E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared make your tummy ache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology*, 43, 1124–1139.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, 47(1) (Serial No. 194).

- Sobel, D. M., & Munro, S. A. (2009). Domain generality and specificity in children's causal inferences about ambiguous data. *Developmental Psychology*, 45, 511–524.
- 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.
- Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, 8, 159–182.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62, 753–766.
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, 115, 155–185.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Tversky, A., & Kahneman, D. (1982). Causal schemata in judgements under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117–128). Cambridge: Cambridge University Press.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121, 222–236.
- Wellman, H., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik, & L. Schulz (Eds.), *Causal learning: Philosophy, psychology and computation* (pp. 261–279). New York: Oxford University Press.