4 Variables Relating Causal Reasoning to Scientific Thinking

In chapter 3, we suggested that rational constructivism, especially as instantiated with the causal graphical model framework, faced certain challenges and had certain theoretical insularities. Throughout the rest of the book, we want to address one issue that illustrates an aspect of that insularity—the relation between causal reasoning and scientific thinking. If children are so good at causal reasoning, as our views about cognitive development and learning suggest, why is scientific thinking challenging, even for adults? This is one of the key questions of this book.

Although not all scientific thinking involves causality, we want to suggest that causal reasoning is the basis of many aspects of the scientific process. Children's abilities to learn about the world through understanding causal relations could potentially be used when they need to interpret scientific content or predict the results of both formal experiments and informal actions on the world. While preschoolers may possess sophisticated causal reasoning abilities, as illustrated by the experiments we reviewed in chapters 2 and 3, it is more of a challenge for them to make inferences about experimental design, or, more generally, to solve scientific thinking problems in the classroom or in real life. Simply put, the fact that children can make inferences about blicket detector systems in a developmental laboratory does not mean that they can engage in the kind of systematic, content-rich inference-making required for scientific investigations.

But why should this be the case? Given that children are so good at causal reasoning, which underpins much of the scientific thinking they are asked to do, why do they struggle with explicit scientific thinking? Our main goal in this chapter is to begin to address this question. We review the similarities and differences between children's causal reasoning capacities and children's scientific thinking abilities, with the goal of beginning to reconcile these two bodies of work. We begin this reconciliation by accepting a few premises:

(1) Young children have sophisticated causal reasoning abilities by the time they enter preschool, and likely earlier. Those inferential abilities are evident in children's reasoning during laboratory-based tasks but are also evident in children's learning from play and social interactions, perhaps particularly in informal learning environments.

(2) Causal graphical models provide an initial description of how young children represent their causal knowledge. That is, algorithms from this modeling framework provide a useful (computational-level) description of how children engage in causal inference.

(3) These same children struggle with the scientific thinking problems presented to them both in the classroom and in the laboratory (although maybe less so in their play). They may lack the ability to demonstrate these kinds of reasoning abilities until late in elementary school or later.

Taken together, these premises imply that making scientific inferences is not the same process as making causal inferences. As reviewed in chapter 1, scientific thinking is a set of abilities that allows one to generate hypotheses, solve problems, and explain aspects of the world. Scientific thinking involves distinguishing between a hypothesis and evidence, an appreciation that theories that generate hypotheses can be wrong and can be revised with the appropriate evidence, and the metacognitive capacity to reflect on this process. All of these capacities have unique developmental trajectories (e.g., Bullock et al., 2009; Kuhn, 1989).

To narrow the scope of our investigation somewhat, here we focus on one crucial ability within the scope of scientific thinking: the ability to use the *control of variables strategy*. Briefly, as the name suggests, the control of variables strategy involves selecting or conducting unconfounded experiments so that one can learn a causal structure by changing exactly one variable and holding all else the same (e.g., Chen & Klahr, 1999; Inhelder & Piaget, 1958; Tschirgi, 1980). This is the method that underlies much of modern experimental practice because it allows for the isolation of particular events, providing a definitive answer to which factors are genuinely causal and which are merely correlated with a particular outcome.

For example, consider the slopes task used in Chen and Klahr (1999). Participants are introduced to two ramps and two different kinds of balls.

The ramps each have a starting gate where a ball can be placed. The angle of each ramp can be adjusted (making the slope steeper or gentler), as can the location of the starting gate (so that the ball can start at a higher or lower point on the ramp). Also, each ramp can be covered with one of two different surfaces, changing the amount of friction between the ramp and the ball as it rolls. Participants are told to learn which factors are important to determining how far a ball will roll down the ramp. Participants can set the angle of the slope, the starting position of each ramp, and the surface, and then can observe the two different balls roll down the ramps from the starting gate. The key idea to solving this task is to manipulate each variable independently in order to consider its efficacy in isolation.

Because of its importance, a large body of work has investigated whether children understand the control of variables strategy and whether and how they can learn this strategy (e.g., Chen & Klahr, 1999; Klahr & Nigam, 2004; Kuhn et al., 1992, 1995). In general, the young children tested in these studies often do not use this strategy without instruction, and even with instruction, they do not use it all the time. Even worse, they tend to misremember the evidence they generate and to misunderstand the relation between that evidence and the theories they might have (e.g., Amsel & Brock, 1996; Croker & Buchanan, 2011; Schauble et al., 1995). In a meta-analysis of studies on children's understanding of this strategy, age did not significantly moderate the outcomes of these kinds of studies (Schwichow et al., 2016). But the youngest children tested in that analysis were age 6, older than the participants in most of the research on causal reasoning discussed in chapter 2.[1]

Although the control of variables strategy is an important aspect of scientific thinking, as noted above, understanding this strategy (and other aspects of scientific thinking) has different conceptual requirements than understanding causality more generally, making direct comparisons between these reasoning processes difficult, if not impossible (see discussions in Kuhn, 2007a; Kuhn & Pearsall, 2000; Ruffman et al., 1993; Sobel et al., 2017; Sodian et al., 1991). Rather than try to align these two abilities directly, then, our approach in this chapter is to identify the main variables on which causal reasoning and scientific thinking—particularly as expressed in the control of variables strategy—differ.

As we make our arguments, we focus on two key example tasks, one from the literature on children's causal reasoning and one from the literature on children's scientific thinking. To represent the causal reasoning literature, we use the blicket detector study as reported in Gopnik et al. (2001), described in chapter 2. In this study, 3- and 4-year-olds observed the effects of two objects on the detector and recognized the difference between an object that activated the machine directly and one that only did so conditionally, dependent on the other object.

To represent the scientific thinking literature, we use computer-based inquiry tasks, such as Earthquake Forecaster as described by Kuhn and Dean (2005; see also Dean & Kuhn, 2007; Kuhn et al., 2009). In this task, children (usually 4th or 6th graders) play a computer game in which they try to predict the risk of earthquakes given particular features. For example, Kuhn (2007b; Dean & Kuhn, 2007) showed children that, on each turn of the game, they could select to observe cases that combined five potential causes of earthquake risk: soil type, S-wave rate, water quality, snake activity, and gas level (although other versions use other features, such as water pollution, water temperature, soil depth, and elevation). Each of these variables had two levels; for example, soil type could be either sedimentary or igneous. Students could query a database and find out how each combination corresponded to a geographic area that had one of four levels of earthquake risk: low, medium, high, or extreme. Students were charged with finding out whether and how each variable contributed to the outcome.

In general, the children tested in these studies rarely used the control of variables strategy when just given the program (e.g., see the "pilot assessment" reported by Dean & Kuhn, 2007 or the pretest performance in Kuhn, 2007b). After being exposed to the program (or a very similar one) over several sessions, children were able to make valid inferences about the potential causes of earthquake risk less than half of the time (Dean & Kuhn, 2007). Some children (19 of the 30 children tested) did show improvement in their reasoning with practice (Kuhn, 2007b), and children who had been given prompts to focus on only one variable at a time also performed better (Dean & Kuhn, 2007; see also Kuhn & Dean, 2005).

Throughout the rest of the chapter, we use the blicket detector and Earthquake Forecaster tasks to illustrate what we take to be the main factors that differ between the literatures on causal reasoning and scientific thinking, acknowledging that these tasks are merely examples of their respective literatures and do not represent the entire state of the field.

Age

One of the clearest differences between most of the research on children's causal reasoning and most of the research on children's scientific thinking is the age of the participants. Studies of causal inference tend to recruit young participants, often toddlers and preschoolers. For instance, in chapter 3, we talked about the debate over whether the concept of "cause" was innate or developed from interactions with the environment. This dialogue places emphasis on research with infants and toddlers, which for the most part has demonstrated that these very young children have sophisticated causal reasoning abilities, at least after about 6 months of age (e.g., Denison et al., 2013; Saxe et al., 2005; Sobel & Kirkham, 2006, 2007). By contrast, studies of children's scientific thinking tend to focus on students in elementary school or even older (e.g., Chen & Klahr, 1999; Kuhn & Dean, 2005; Schwichow et al., 2016).

This difference in the age of the target populations is driven by the different (and equally valid) foci of the two literatures. Studies of causal reasoning have typically sought to find the basic building blocks for children's abilities, hence they have focused on young children and infants. By contrast, studies of scientific thinking have tended to be more concerned with understanding children's thinking within formal, educational contexts, aiming to connect this research to the practical problem of having to teach children science in schools.

A few studies have presented the same measures to young children and adults, or to children across a wide age range. Some of this research has argued for continuity in the mechanisms children and adults use to make inferences. For example, one of our studies presented adults with causal systems similar to the blicket detector (e.g., the "superpencil detector," which detects "superlead," a nonobvious property of golf pencils[2]) (Griffiths, Sobel, Tenenbaum & Gopnik, 2011). In this study, we replicated experiments previously done on preschoolers with adult samples and found similar (although more nuanced) results. We also generated new causal reasoning problems with this task and showed that both adults and children solved them in similar ways, again suggesting continuity between 4-year-olds and adults.

Other research, however, has found discontinuities in development. Lucas et al. (2014) presented causal reasoning tasks using the blicket detector to both adults and children, and they demonstrated that reasoning was relatively similar in some cases. But in other cases, children were better at learning causal structures than adults, particularly cases in which causal relations were conjunctive. The conclusion from this work could be that children are better than adults at causal reasoning. However, a more likely explanation is that this difference can be explained by different degrees or types of prior knowledge in these two populations. Specifically, although conjunctive causality is not rare in everyday life, conjunctive mechanisms are rarely discussed and are usually limited to enabling cases. For example, oxygen must be present with a spark to start a fire, but the absence of oxygen is not a counterfactual most will think about in order to explain how a fire could have been prevented. As a result, when a conjunctive cause is presented outside of an enabling paradigm, adults might be less open to registering this relation. Children, on the other hand, simply accept that this is the mechanism of this strange, novel machine. What changes in children's cognition is not the reasoning mechanism, but the prior knowledge that children use to instantiate initial hypotheses (analogous to the ideas in Bayesian inference that we discussed in chapter 2).

These studies suggest that preschoolers, older children, and adults may possess roughly equivalent abilities with respect to causal reasoning (just different amounts of prior knowledge). But there are some reasons to believe that there is more to the story. For one thing, these studies use tasks with similar (low) amounts of real-world scientific content, and that similarity might have led to the similarity in results. More importantly, the fact that studies in causal reasoning and scientific thinking have tended to recruit children of different ages has led to a variety of other design choices that then co-vary with age, including the issues that we discuss below: the causal systems' levels and types of complexity, whether these systems present scientific content, and whether children are asked to observe data about the system or generate their own. These confounds across the two literatures make it difficult to determine why young children often tend to succeed in causal reasoning tasks using methods like the blicket detector, while older children often respond differently on scientific thinking tasks like Earthquake Forecaster. One possibility is that children's early-developing causal reasoning skills are maintained as they grow older, but are masked by the increased task demands of systems like those in Earthquake Forecaster. A second possibility is that children's early-developing reasoning skills can be used only for simple systems or only for systems that lack recognizably scientific content, and children require different or additional skills in order to think appropriately about the kinds of systems presented by scientific thinking measures.

Complexity

Studies of children's causal reasoning capacities often focus on preschoolers, toddlers, or even infants. One reason for this is historical, because one of the fundamental questions behind these studies was whether young children could engage in causal inference, particularly in response to Piaget's claims that young children were "precausal." Because of this emphasis, the demonstrations and causal systems used in this work tend to be simple, and the procedures are designed to walk children through the demands of the study. These simplifications are designed to eliminate any extraneous factors that might get in the way of children's abilities to demonstrate their causal reasoning.

Just as choosing to test young participants introduces certain limitations to studies' designs and materials, choosing to test older children also guides the choice of methods and procedures. While blicket detectors are fascinating to 3-year-olds (and strange curiosities to adults), 9-year-olds or middle-school students are often bored by the simple causal systems presented in blicket detector tasks. Moreover, older children and adults (at least in Western, educated, industrialized cultures) tend to possess a great deal of existing causal knowledge about electronics, which allows them to make certain assumptions about how the machine works; these assumptions can lead them to do better on certain kinds of causal inference measures, and worse on others (as in the Lucas et al., 2014, example we described above). Tasks like Earthquake Forecaster (or the slopes task [e.g., Chen & Klahr, 1999], or playing with spring tension [e.g., Schauble, 1990], or experimenting with sinking and floating [e.g., van Schaik et al., 2020], or making inferences about planets [e.g., Panagiotaki et al., 2009[3]], or many other tasks in which scientific thinking has been tested) can be more engaging to children in this older group precisely because they are more complex. For the blicket detector, once you get over the initial surprise that the machine lights up, it's just not that interesting. Using more complex tasks provides a richer real-world context for children to explore, which makes them more interesting to do and which increases their resemblance to real tasks involving scientific thinking.

For these reasons, studies on children's causal inference and studies on children's scientific thinking tend to vary in the amount and type of complexity instantiated in the systems that these studies present. To make these differences more concrete, consider again the contrast between the blicket detector and Earthquake Forecaster. In the original blicket detector task, children saw two potential causes (objects A and B), each of which could be in one of two states (on or off the machine). These potential causes could have one of two effects: turning the machine on or failing to do so. By contrast, Earthquake Forecaster presented five potential causes (e.g., soil type, snake activity), each of which had two levels (e.g., igneous or sedimentary, high or low). This system also presented four possible effects (i.e., four levels of earthquake risk). The number of factors that a participant must keep track of is thus much larger in Earthquake Forecaster; this alone may contribute to older children's struggles in reasoning about such systems.

In addition, and perhaps more importantly, the causal structure of the system presented by Earthquake Forecaster is more complex than the one presented by the original blicket detector task. This task presents an interactive (specifically, an additive) system, in which some of the five causes individually can be diagnostic of one level of risk for earthquakes, but jointly combine to produce different levels of risk. Previous work clearly shows that thinking about interactive systems is difficult, even for older children (e.g., Schauble, 1990, 1996; see also Novick & Cheng, 2004). By contrast, most studies in causal reasoning use disjunctive causality, in which the presence of any cause would lead to the same effect.

Finally, a crucial difference between the two types of task is the presence of causes whose efficacy is unknown. The blicket detector task showed children what happened when each object was placed on the detector by itself, and no other objects were present. This task thus presents no uncertainty about which things could potentially be causes or about what the effects of all potential causes are. Earthquake Forecaster (and most measures of scientific thinking), presents situations where the efficacy of individual causes is unknown. In many cases, this is because of the number of variables about which children have to reason.

Some of our previous work has shown that the presence of potential causes whose efficacy is unknown can compromise children's abilities to reason about a causal system (e.g., Erb & Sobel, 2014; Fernbach, Macris & Sobel, 2012; Sobel, Erb, Tassin & Weisberg, 2017). For example, Sobel et al. (2017, Study 1) presented 3- to 7-year-olds with a blicket detector and a set of four objects. In the first trial, children saw each object placed on the machine by itself. The first object activated the machine, making it light up and play music. The second did not. The third and fourth objects also activated the machine. The experimenter then brought out a large piece of cardboard and hid the machine and objects from the child. He told the child he was putting one of the objects on the machine, and that the machine lit up. Even though children could not see this, they could confirm it because they heard the music that went with activation. The experimenter then said he was taking the object off the machine and putting it back. (In reality, the experimenter never moved any of the four objects—he just mimicked doing so and activated the machine remotely, so that there would be no visual clues as to which object had been moved.)

The experimenter then removed the cardboard occluder and asked children a set of diagnostic reasoning questions. He first asked which object he had used to make the machine go. Whichever object children chose, the experimenter said, "That's a good guess, but it's not right." He removed that object and asked children to try again. Regardless of what they chose the second time around, he also indicated the choice was wrong, and asked for a third guess, this time with only two objects left on the table.

Responding correctly to these questions is straightforward: Just don't pick the object that you saw fail to activate the machine. However, 3- and 4-year-olds were equally likely to choose an object that had previously activated the machine as to choose the object that had not. What is important here is that they could (in a separate trial) correctly make predictions about what would happen if each object was placed on the machine, showing that they had remembered each one's effects (i.e., they could predict that the object that had failed to activate the machine would fail to activate it again, suggesting that their performance on the diagnostic reasoning question was not the result of poor memory). What was more of a challenge was applying that knowledge across numerous requests to revise their guess. Five-, 6-, and 7-year-olds, however, almost never got this trial wrong.

The more complicated finding comes from comparing performance on this trial with performance on two others (shown in figure 4.1). In the second trial, everything was same as the first, except at the initial demonstration phase. In this case, the first object activated the machine (same as above), the second object did not (also the same), the third did (also the

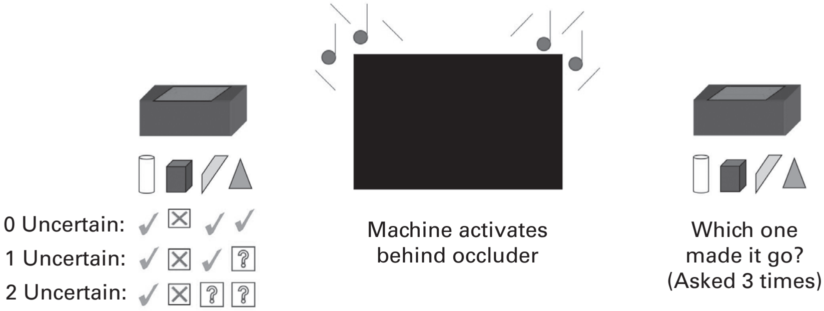

Figure 4.1
Schematic of procedure used by Sobel et al. (2017, Experiment 1).

same), but the fourth object was never placed on the machine (different from above). Now there is an object that could be a cause, but the child simply does not know. In the third trial, the first object made the machine go, the second did not, and the other two were never placed on the machine. In this case, there were two objects with unknown efficacy.

The critical finding of this study was that there was a significant correlation between age and the number of unknown causes children could accurately reason about diagnostically. Five-year-olds had no problem when all of the causes were known but fell to chance levels of performance in the other two trials. That is, children at this age sometimes mistakenly chose the object that had previously failed to activate the machine as a potential cause when they had not been able to observe the efficacy of all of the objects. As children got older, specifically between the ages of 5 and 7, they were able to handle to more and more uncertainty in their diagnostic inferences, becoming less and less likely to choose the inefficacious object, even in the presence of some uncertainty.

These results match up well with some of the differences that we have been emphasizing between young children's performance on tasks involving the blicket detector and older children's performance with more complex scientific thinking tasks like Earthquake Forecaster. The number of causes and causal relations, the complexity of these causal relations, and the amount of uncertainty presented by the system tend to be much greater for studies on older children's scientific thinking abilities than for studies on younger children's causal reasoning abilities. Again, this makes comparing them difficult;

it is not clear whether children perform less well on scientific thinking tasks because they lack some basic reasoning skills or because some aspect of the complexity of these tasks simply takes them out of their grasp.

Use of Scientific Content

As noted in chapter 2, causal reasoning measures like those that use the blicket detector do not reflect the structure of any particular real-world system. They present causal reasoning problems in an abstract way. Although some have described these tasks as *decontextualized*, lacking any reference to real-world causal structures (e.g., Kuhn, 2007a), it would be more accurate to say that such tasks are just *less* contextualized than other problems, in that they do not reflect any particular piece of real-world content knowledge. Making inferences about causal relations among objects and a machine requires children to possess knowledge about general causal factors such as temporal priority (causes precede their effects) and spatial contact (the machine probably works by having the blocks touch it, not merely come near it). We cannot imagine a task that is purely context-free (see Sobel & Munro, 2009).

Still, beyond these basic factors, causal reasoning measures like those that use the blicket detector do not require children to draw on much prior experience with blocks or machines or other real-world causal systems. This is a strength of the paradigm; children do not need much prior knowledge to make causal inferences about these systems. This is also a strength of the causal graphical model framework; it is a domain-general reasoning system. At an abstract level, a graphical model does not care what it is about; the nonindependence findings that we discussed with the cellphone example in chapter 3 potentially emanate from the fact that *participants* care. They might have been reasoning about models that were not intended by the researchers. Hierarchical modeling (e.g., as discussed by Goodman et al., 2011) as well as the edge replacement findings we discussed in chapter 3 may provide a way to conceptualize how this prior knowledge constrains children's reasoning and thinking, allowing the framework to take participants' tendencies into account.

However, a lack of context is also a weakness of the blicket detector paradigm. Although the causal systems instantiated in the blicket detector could be abstract representations of real-world causal systems, that fact is likely opaque to young children. More importantly, children's responses to such systems, in which their reasoning abilities can be neither helped nor hindered by their prior knowledge about specific scientific domains, may not provide a true reflection of children's scientific thinking abilities. These tasks might show how children think about minimally contextualized systems, but they could have little or no bearing on our understanding of children's thinking about real-world causal systems. That is, the abilities that children need in order to solve a causal reasoning problem with a blicket detector might be different from the reasoning processes used in scientific investigations in the lab, the classroom, and everyday life. In support of this argument, Dunbar (2002) suggests that "causal reasoning in science is not a unitary cognitive process . . . but a combination of very specific cognitive processes that are coordinated to achieve a causal explanation such as an unexpected finding" (p. 157).

To expand on this statement, designing experiments, integrating data with existing theory, revising beliefs, and communicating both the conclusions and the process by which one arrives at those conclusions are all part of causal reasoning in science. So far, the blicket detector and other measures of children's causal reasoning have really focused on only one cognitive process: making appropriate causal inferences. In part II of the book, we explore several ways that blicket detectors can be used to probe different aspects of children's abilities. But, for now, we still must deal with the objection that research using novel and artificial causal systems like the blicket detector have no bearing on children's scientific thinking in real-life contexts because they do not adequately resemble the work that real scientists do. If this is the case, then it becomes vitally important to understand why tasks that do resemble the work that real scientists do tend to be more difficult. We can again use tasks like Earthquake Forecaster, which present more richly contextualized systems that instantiate their variables in real-world problems, to identify the source of these problems. Instead of blocks that can be either on or off the machine, children are asked to think about potential causes that might conceptually relate to causes of earthquakes (e.g., what type of rock is more abundant, whether there is water pollution, the elevation of the land mass).

This choice to contextualize the causal system is both a strength and a weakness. As noted above, a major strength of this task is its realism. When children think about the structure of the real world, they need to grapple with data that come with meaningful labels and that produce effects in a real-world context, because most reasoning tasks that children will have to solve will involve some kind of real-world content. Plausibly, in order to solve such tasks, children would need to strip away the specific labels and effects and construct more abstract representations of the variables that they represent. The blicket detector and other laboratory-based measures eliminate the need for this step. As a result, they might be overestimating children's abilities. Children will need to develop the skills to deal with both the real-world content and the abstract causal structures that underlie it to become fully mature scientific thinkers.

But richly contextualized systems could be making the underlying reasoning task too difficult for children, and hence may be underestimating their abilities. These tasks not only ask children to solve the abstract problem of how different variables contribute to an effect, they also require children to navigate potentially unfamiliar vocabulary, like "igneous" and "sedimentary," or to think carefully about whether a general variable like elevation relates to earthquakes. Additionally, different children may have come to this task with more prior knowledge about earthquakes and their causes, which might put them at an unfair advantage. For instance, there is a complex relation between the activity of nonhuman animals (including snakes) and the probability of earthquakes (Woith et al., 2018). Despite this, many laypersons seem to believe that animal activity is a risk factor. However, we doubt whether young children believe this initially, and thus might dismiss this variable as potentially irrelevant, much like infants ignore certain correlations (e.g., Madole & Cohen, 1995, as discussed in chapter 2).[4] Prior knowledge may work against them in this case, and their performance on this task may be poor not because they were unable to think about how combinations of potential causes lead to an effect, but because the surface features of the task threw them off the trail.

This distinction relates to the idea of the "seductive allure" of particular types of scientific content that we have investigated in a series of studies on adults' understanding of science (e.g., Hopkins, Weisberg & Taylor 2016; Weisberg et al., 2008; Weisberg, Hopkins & Taylor, 2018). Briefly, this effect occurs because adults tend to prefer explanations of scientific phenomena that involve reducing those phenomena to a more fundamental science, such as an explanation of a chemical phenomenon in terms of physics. This is especially the case for psychological explanations, which seem more satisfying when they include neuroscientific terms, even when the neuroscience information does not add any explanatory value.

One potential explanation for this preference is that adults believe that the language in reductionist explanations sounds smarter and more complex, hence more "scientific." But that's adults. For children, those "smarter" explanations—the ones that use more complex vocabulary—might sound too smart. Children might get so caught up in trying to understand these new concepts that they fail to reason appropriately, or they might get so turned off by needing to understand the strange vocabulary that they do not try to reason appropriately at all.

More generally, the way in which a measure of scientific thinking is designed might affect children differently based on what prior knowledge they bring to the reasoning task. Parents who talk to their children endlessly about sinking and floating while in the bathtub might have children who approach sinking and floating tasks with a better (or at least a different) understanding of the role of density than parents who talk to their children more about the importance of hygiene. But these latter children might perform better in laboratory tasks about the role of germs in disease transmission (e.g., Conrad et al., 2020).

This is a potentially important explanation for why studies have found sophisticated causal reasoning when using the blicket detector and other less contextualized measures, and less robust reasoning when tasks are more contextualized: Asking children about blicket detectors may make it easier for them to demonstrate their reasoning about causal systems per se, while asking children about more contextualized problems with scientific content, like predicting the risk of earthquakes, may make it harder for them to understand how to bring those capacities to bear on the problem they are presented with. Similarly, this might be why scientific inference in the laboratory or the classroom is hard to demonstrate, but seems more evident in free play either in the lab (e.g., Cook et al., 2011) or in museum settings (e.g., Callanan et al., 2020), as we documented in chapter 3.

In the end, the extent to which one agrees with the findings of one set of studies over the other may be a matter of perspective: Is it more important to emphasize the thinking skills, regardless of context? Or is it more important to test children's abilities in the context of real-world information, because that is where they will tend to need to display these abilities? We address these issues directly in chapter 6, where we compare children's performance with more and less contextualized causal systems, and in chapter 11, where we discuss children's understanding of learning in different learning environments.

Observing versus Generating Data

As noted above, scientific thinking often involves experimentation—doing things and observing their results. For example, the Earthquake Forecaster program asks children to figure out how five two-level variables combine to reflect four different levels of earthquake risk. The program provides them with a database of cases that they can query. Their main task is to use the program to design controlled experiments—changing only a single parameter while keeping everything else fixed—because only this method of testing will allow them to draw conclusions about which variables are causally relevant and how the different variables are causally related.

However, results with the Earthquake Forecaster program and similar systems show that children do not always apply the control of variables strategy, even late in elementary school. They tend to not test hypotheses explicitly and generally do not design unconfounded interventions to learn new causal relations. Instead, they tend to change many variables at once. Naturally, because their experiments are generally ineffective at disentangling different causal factors, children in these studies often do not discover the causal structure of the system—at least not without practice or direct instruction, and even then, this thinking is still difficult.

Although this is partially due to the other variables discussed above, the fact that children have to generate their own data may be a major contributor to their poor performance with this program, particularly when they are not given instructions to focus on individual variables. Indeed, paradigms in scientific thinking, contextualized with rich scientific content, present children with two distinct challenges. First, as noted above, these tasks ask children to construct appropriate experiments using this system to tease apart which combinations of variables lead to the effects. This poses a challenge because most children either do not know how to use the control of variables strategy or do not know how to apply the strategy to this situation. Second, children must draw conclusions from a set of data. This may be a difficult task in itself, even when the set of data has been properly constructed (Kuhn, 2007b).

Causal inference measures that use less contextualized paradigms, on the other hand, almost always present children with sets of data and do not ask children to design their own experiments. These sets of data are constructed to lead children to discover the underlying causal structure of the system.

Indeed, children tend to view these situations pedagogically, understanding the experimenter's presentation of the data as being designed to teach them something about the system (e.g., Gweon et al., 2010; Yang et al., 2013). This removes the requirement for children to be able to generate productive data, allowing them to focus only on the second task: drawing conclusions from the data they observe.

This is not to say that children cannot learn from their own actions, particularly when those actions are set up so that they do not conflict with the efficacy of the objects. Preschoolers do engage in productive exploratory behavior when faced with ambiguous evidence, and that exploratory behavior can lead to them generating information that allows them to resolve the ambiguity (e.g., Cook et al., 2011; Legare, 2012; Schulz & Bonawitz, 2007). And preschoolers also can learn a novel set of causal relations better from exploring their environment than from observing the actions of another, at least under some circumstances (e.g., Sobel & Sommerville, 2010). Further, as we discussed in chapter 3, there are cases of children learning from their own actions in museum settings, particularly when those actions are construed as play (e.g., Callanan, Legare et al., 2020; Sobel, Letourneau et al., 2021). However, the ability to produce a genuine control of variables strategy emerges slowly over the course of elementary school, and even adults find aspects of such reasoning hard (e.g., Klahr et al., 1993; Schauble, 1996). Small wonder, then, that children struggle more to solve tasks like Earthquake Forecaster than tasks like the blicket detector: Even with some degree of understanding of how to explore a causal system, it is far less taxing to their cognitive resources to merely draw conclusions from a set of data that has been designed to help them draw the correct conclusions than to have to additionally generate this set of data.

There's one other important difference. Most measures of causal reasoning, like those using the blicket detector, involve generative causality. Children are shown a set of potential causes (usually objects), which individually or together have the efficacy to activate the machine. During the experiment, those objects are placed on the machine in an active intervention. In contrast, Earthquake Forecaster is more diagnostic. Children select a certain combination of features and then observe a case that had those features. But selecting an observation is not the same as making an intervention. That is, when children choose to observe a case in which there are more igneous rocks and lower snake activity, they are not creating a new case with that combination of variables. Selecting observations of data that have occurred in the past does not necessarily provide the same conditional probability information necessary for causal learning as conducting interventions on a generative system (Lagnado & Sloman, 2004). While many measures of scientific thinking, like the slopes task, involve children generating their own novel data from their interventions, other measures rely on selecting observations, which could interfere with children's learning.

Recognizing and Using the Control of Variables Strategy

So far, we have suggested that an important finding of a variety of studies of children's scientific thinking is that the control of variables strategy has a prolonged developmental trajectory, extending beyond the preschool years (e.g., Klahr & Nigam, 2004; Kuhn et al., 1995). Second graders can learn the strategy, particularly through direct instruction (Chen & Klahr, 1999). But children do not seem to use this strategy spontaneously until they are older, if at all (e.g., Klahr et al., 2011; Kuhn et al., 2008). And, as we have discussed, even some middle school students can struggle with constructing meaningful and well-controlled interventions without instruction (e.g., Kuhn & Dean, 2005).

But in some cases, younger children seem to have nascent understanding of the strategy and the ability to recognize its efficacy. One example comes from work by Sodian et al. (1991), who showed that first and second graders could recognize which interventions would be effective in solving a particular problem. Children in this study were presented with a story in which two characters knew there was a mouse in their house but did not know whether the mouse was big or small. The characters designed two mousetraps. One mousetrap had a large opening, which either mouse could get into; the other had a small opening, which only a small mouse could get into. Children were asked which mousetrap they wanted to use to figure out the size of the mouse. Most of the second graders (86%) and about half of the first graders (55%) in this study recognized that the mousetrap with the small opening should be used to figure out the size of the mouse: If the mouse was caught, it was the small mouse; if not, it was the large one. These children also correctly recognized that the mousetrap with the large opening would be useful if the character's goal was to feed the mouse, because either mouse could access it.

Piekny and colleagues (Piekny et al., 2014; Piekny & Maehler, 2013) replicated and extended this work to a larger age range (4- to 12-year-olds), and they also looked at development in a longitudinal sample between the ages of 4 and 6. In general, 4-year-olds were not different from chance responding on this measure. Five-year-olds were better on some of the measures (for example, they were better than chance on choosing which mousetrap to use if the goal was to feed the mouse), but there were clear differences between 5-year-olds and 6-year-olds.

These data suggest that children can choose an intervention that would provide them with the appropriate evidence to make a causal conclusion or to achieve a goal. The ability to evaluate evidence to draw explicit conclusions has a similar developmental trajectory. For example, Ruffman et al. (1993) introduced children between the ages of 4 and 7 to correlational evidence (e.g., that children who ate food of a certain color had fewer teeth than those who ate food of another color). After children reported on the correlation, the experimenter rearranged the data, so that it now presented a “fake” relation—suggesting that food of the other color promoted tooth loss. Over a number of studies, they showed that children by the age of 6 could unambiguously infer that one might come to a particular conclusion from a set of data, even if those data did not represent the actual state of the world.

These studies suggest that children can determine what would be an appropriate test of a hypothesis and can come to appropriate causal conclusions about hypotheses by approximately age 6 or 7. Additionally, evidence reviewed earlier suggests that preschoolers (and possibly even younger children) might be able to use evidence to make causal inferences when those data are presented to them, and they might even perform informative interventions in their exploratory play. But none of these abilities are sufficient for scientific thinking as we have defined it. For instance, hypothesis testing requires more than just causal inference; it requires the kind of strategic thinking involved in appreciating what kinds of data one needs to observe, and what conclusion to draw given observing those data or alternative results that speak against a given hypothesis (Morris et al., 2012). This understanding seems to have a more prolonged developmental trajectory.

But is it the case that young children lack these capacities entirely? We think that the answer is no. For example, van der Graaf et al. (2015) demonstrated that 4- to 6-year-olds could generate unconfounded interventions with ambiguous data using a scientific thinking task that presented only two potential causes, which were specifically explained to the children. While this suggests children might be able to design their own interventions, children in this study were also given much scaffolding and direct instruction. Can children engage in these behaviors more spontaneously?

Recent research suggests that they can. For example, Walker et al. (2019) presented 4- and 5-year-olds with information about a new causal system: A block that had both a dot on its front and an antenna on its top made a machine play music. Children then observed two people try to figure out what made the machine work. One of these people used a control of variables strategy, choosing a block with a dot but no antenna to put on the machine (i.e., changing only one variable while keeping the others fixed). The other person did not use the strategy, choosing a block with neither a dot nor an antenna to put on the machine (i.e., changing two variables at once). In both cases, the machine failed to activate, and both people drew the conclusion that the antenna was causally responsible for the music. When children were asked to endorse one of the experiments, they chose the one that used the control of variables strategy.

Lapidow and Walker (2020) followed up on this work by showing 4- to 6-year-olds a novel causal system: a gear machine in which two gears could be interlocked and made to activate. Across several experiments, they introduced children to the gear machine and different gears that were either "working" (could spin, and possibly cause other gears to spin) or "broken" (would not spin or could only be spun by another gear). In one experiment, children were told that the two gears would spin together if at least one was a "working" gear. They saw two gears spin and were told to determine whether both gears were working or whether only one was working and the other was broken. Children were given the choice of two interventions to try. One was informative and would tell the child which causal structure was accurate. The other was confounded, so observing the results would not allow children to reach the appropriate causal conclusion. Children mostly chose to make the informative intervention. Those who chose this intervention were more likely to draw the correct causal conclusion from the data that they subsequently saw. This indicated that these children could both select an unconfounded intervention designed for them and could learn from the results of this intervention.

A similar study was performed by Moeller and Sodian (2019). They presented 4- to 6-year-olds with cases in which a set of candidate causes produced an effect, but the evidence about these candidate causes' efficacy was confounded. Children were asked whether they knew which of the candidate causes was the actual cause. The majority of children (around 70%) said that they did know, but a significant minority routinely stated that they did not and justified their responses appropriately. The same children also saw trials in which two or three candidate causes produced an effect. Children were asked to choose which experiment they would want to do in order to figure out whether a particular candidate cause was the actual one. Like in Lapidow and Walker's study, children were shown confounded or unconfounded interventions. Around half of the children correctly chose the experiment that was unconfounded. Further, about 20% of the children not only chose that option but also justified their choice by appealing to the control of variables strategy. There were also significant effects of age, with the older children in the sample (6-year-olds[5]) outperforming the younger children (3-year-olds), and of complexity, with all age groups performing better when there were only two candidate causes than when there were three.

We expanded on these results in a follow-up experiment, which produced similar results, but which also considered how well children could determine which of the candidate causes was the actual cause based on the intervention they chose and the data they observed (Moeller, Sobel & Sodian, 2021). Critically, children who chose the unconfounded intervention came to the appropriate causal conclusion about 70% of the time, significantly more often than chance levels and significantly more often than children who chose to make the confounded intervention.

These data suggest that, during the preschool years, children might have some nascent understanding of the control of variables strategy and of what is involved in constructing experiments that will allow one to learn about a system's causal structure. This is quite impressive for children who have not yet had any formal training in science. However, these studies do not show that children can design their own experiments that use the control of variables strategy. In all of these studies, children observed experiments that had been designed by someone else, and they had to choose which intervention was appropriate. This is a far cry from designing a successful experiment oneself (as those of us who mentor students in our labs know well).

Learning from Exploration and Play

The studies we reviewed in the previous section suggest that children can distinguish confounded from unconfounded interventions if those actions are designed by another person. Further, if they observe unconfounded data, they are more likely to learn a system's causal structure. But this work does not demonstrate that young children can design those unconfounded interventions on their own. Indeed, Bullock and Ziegler (1999) showed that children could select interventions consistent with the control of variables strategy in a scientific thinking context, but they struggled to produce such interventions (although see Cook et al., 2011, for a demonstration that they can do so).

A number of studies that examine how children learn from exploration could be interpreted as consistent with the possibility that young children are not always designing unconfounded interventions in their exploration and learning (e.g., McCormack et al., 2016; Meng et al., 2018; Nussenbaum et al., 2020; see also Reiber, 1969, for similar findings[6]). In these studies, children are asked to learn about a causal system through actions that they take on the system. For example, McCormack et al. (2016) showed 5- to 8-year-olds a causal system of three possible events and possible models for how the events were related to one another (for example, event A could cause both events B and C as a common cause, or event A could cause event B, which in turn caused event C, as a chain). They then asked children to learn the causal structure through intervening on the system. Children could learn some structures better than others. For example, common causes seemed to be easier to learn than chains. More interestingly, children in these studies often tried to make events occur. Younger children in particular more often intervened on a "root" node—an event that causes the most outcomes (for example, in the $A \rightarrow B \rightarrow C$ model described above, they mostly intervened on event A). Results like these have led some to describe children not as little scientists but rather as little engineers, whose primary goal is to produce as many effects as possible (following Schauble et al., 1991).[7] Contrary to rational constructivism (which views children as little scientists, as reviewed in chapter 2), these data suggest that children are not naturally motivated to generate data that facilitates learning.

Nussenbaum et al. (2020) proposed a mixed model that integrates aspects of both of these approaches. They suggest that young children have two exploratory strategies, one that involves maximizing information gain in their exploration and one that displays what they call a "positive testing bias" (following work by Coenen et al., 2015, on adults), which motivates children to make as many effects occur as possible. As children move into adolescence, the balance of strategy use shifts away from a positive testing bias and toward more of an optimal information-gain strategy for self-guided learning.

This approach is sensible, but it might tell only part of the story, particularly when contrasted with the exploration and learning that children engage in when they are in more authentic learning environments like a children's museum. As we discussed in chapter 3, when looking at parent-child interaction at a set of gear exhibits, children's systematic exploration of exhibits was related to their causal knowledge of the exhibit components (Callanan, Legare, Sobel et al., 2020). Moreover, numerous studies that we have already discussed suggest that children can learn from their play, both on their own and collaboratively (see Weisberg, Hirsh-Pasek & Golinkoff, 2013, for a review). One of the reasons that play can be such a powerful mechanism for learning is that play involves self-generated actions; children decide what actions to take and then observe the results of those actions.

But self-generated action has some drawbacks, particularly for younger children. As noted in chapter 3, preschoolers can be more influenced by the data they themselves generate than by the data they observe, even if this leads them to make the wrong conclusions. For example, Kushnir and Gopnik (2005) asked 4- to 6-year-old children to make inferences about the efficacy of objects that had probabilistic relations with the blicket detector. In their critical experiment, children observed two objects (A and B) that were each placed on the machine twice. Object A activated the machine both times, and object B did not activate the machine both times. The children were then given both objects and asked to put each on the machine. When the child put A on the machine, it did not activate, but when the child put B on the machine it did. When children were asked which of the two objects had "more special stuff inside" (which they had established was responsible for the machine's activation), children in this condition were more likely to choose object B—the one that worked for them, even though it did not work the majority of the time. In a control condition, children observed A and B being placed on the machine with the same efficacy.

The data were the same, but in this condition, children never touched the objects. Now, they (correctly) chose object A more often. This result suggests that there are drawbacks to self-generated interventions: They might lead children to overvalue what they themselves do as opposed to objectively inferring general patterns of efficacy.

In general, preschoolers overestimate the importance of action for learning (Sobel & Letourneau, 2018) and to some extent discount actions that they generate that are not efficacious (Gweon & Schulz, 2011). This overreliance on self-generated action indicates that how children learn from their own actions might depend on the results of those actions.

We investigated this issue in our lab in two ways. First, we asked parents to play together with their 3- or 4-year-old to learn the rules of how a blicket detector worked. Those rules had previously been shown to be difficult for preschoolers to learn (Walker et al., 2017), but were fairly straightforward for parents. It turned out that the actions that children made during their play did not predict how well they learned. What did predict their learning was the nature of the parent-child interaction. Children whose parents allowed them to play freely with little guidance learned the rules the least, whereas children whose parents guided their play or engaged in more direction by setting goals for their play learned the rules better. However, children in this latter group (the one that was more parent-directed) were the least engaged by the task (Medina & Sobel, 2020). That is, children can learn from both directed and guided play, but are more engaged when parents do not explicitly tell them what to do.

Second, we looked at how 4- to 7-year-olds played with a blicket detector on their own after being shown ambiguous data (Sobel, Benton, Finiasz, Taylor & Weisberg, 2021). In this case, we restricted children to a certain type of intervention: They could not simply test individual potential causes one at a time, but rather could only test pairs of potential causes together. This required them to use Tschirgi's (1980) strategy of varying one thing at a time if they wanted to resolve the ambiguity.

We found that, independent of children's age, their first action and its results influenced how children played. If children's first action varied only one thing relative to the initial demonstration, they were more likely to generate the information necessary to disambiguate the initial data, even if that first action did not itself provide unconfounded information. Moreover, if children's first action was efficacious, they were more likely to continue to activate the machine on subsequent trials. This suggests that children might sometimes be motivated to uncover causal structure in their play, but what happens when children start playing might influence their later goals. That is, young children *can be* systematic in their exploration, even if they are not using the most efficient control of variables strategy and even if they do not engage in such behavior all of the time.

This study, together with the findings reviewed in this section and the last one, shows that children in preschool and in early elementary school seem to understand something about the logic of the control of variables strategy. They can recognize its utility in forced-choice tasks and can sometimes interpret data generated from this strategy. In turn, this suggests that children's difficulties with tasks like Earthquake Forecaster may lie in that system's causal complexity or its contextualization and not necessarily its requirement to generate unconfounded experiments, though more research is needed on these issues.

Diagnosis and Belief Revision

Mature scientific thinking involves a suite of skills, importantly including the ability to diagnose what underlying causal structure was likely responsible for a currently observed set of data (often in the service of making predictions or inferences about the system) and the ability to continually revise and update one's beliefs in light of new evidence. Although the causal graphical model framework discussed in chapter 2 can provide mathematical formalizations for these processes, studies of children's early causal reasoning abilities have not tended to measure whether children can engage in them.

Consider again most experiments using a blicket detector. Children presented with this paradigm are tasked only with saying which objects are blickets (or efficacious in some other way). That is, they need to make predictions about which block or blocks will turn the machine on. These predictions might have come from a process of diagnostic inference, but that process was minimal at best. More importantly, most blicket detector studies do not ask children to revise their articulated beliefs in light of new evidence, so they provide no evidence that children are able to update their hypotheses explicitly. Given that these processes are crucial to scientific thinking, these arguments strongly suggest that the skills that allow children to succeed at young ages in causal reasoning tasks might be different from the skills that they will eventually need for mature scientific thinking.

Indeed, a small number of recent studies suggests that these kinds of reasoning skills continue to mature well after the preschool years. When preschoolers see systems that do not provide them with full information about a system's causal relations, they are less able to construct causal models and to draw appropriate conclusions than are older children (Sobel et al., 2017). In addition, children's diagnostic reasoning capacities interact with some of the variables discussed above, particularly the ability to reason about uncertainty (Fernbach et al. 2012; Kuhn, 2007a), making it even harder to determine when and how these abilities develop.

With respect to this kind of belief revision in young children, the literature is scarce. It is well known that even adults have difficulty considering information that contradicts their existing beliefs, and they will often distort the new information to conform to their beliefs or reject altogether in favor of seeking out confirming evidence (*confirmation bias*; see Nickerson, 1998). While it seems likely that children are implicitly forming and revising hypotheses over the course of a blicket detector experiment, no experiment that we know of has examined this directly. To address this issue, our labs are currently conducting research to examine these processes in children. In one study (Macris & Sobel, 2017), 4- and 5-year-olds were shown that a cube with a particular kind of internal property activated the blicket detector. We asked children whether they thought cubes made the machine go or whether other objects with that same kind of inside made the machine go. Crucially, no matter what children chose, they were provided with evidence that this initial guess was wrong. In some cases, children were shown weak counterevidence—one demonstration that their initial guess was wrong. In other cases, children were shown stronger evidence—three demonstrations that their initial guess was wrong. In still other cases, children were given verbal feedback about that counterevidence. For example, an experimenter told them, "I've seen these objects before, and I know that the cubes make the machine go."

These experiments found that children have some tendencies to change their belief on the basis of counterevidence (see also Bonawitz et al., 2012;

Kimura & Gopnik, 2019; Young et al., 2012). But they are much more likely to succeed when they are directly told that their initial guess was wrong than when they observed counterevidence (although the stronger counter-evidence was more effective than the weaker counterevidence). In the cases where they were given verbal testimony, they changed their belief about 80% of the time. In the cases where they observed data, they changed their belief more frequently than chance levels, but not more than half the time. There is definitely room for children to improve.

One facet of these studies is that children's initial belief is based on a guess. A different way of testing children's belief revision is to look at how they generate data to test a more firmly held belief. To do this, Köksal-Tuncer and Sodian (2018) introduced children between the ages of 4 and 6 to a blicket detector and four blocks, two that were heavy and two that were light. The heavy blocks activated the machine and the light blocks did not, suggesting that weight was the underlying factor that caused the machine's activation. Children then observed two new blocks that contradicted that belief (the light block made the machine activate and the heavy one did not). At this point, they showed children four new objects, and revealed that one object but not a second one had a sticker on the bottom, hidden from view. Critically, among these four objects, one heavy and one light object had a sticker, and the objects with a sticker activated the machine, thus invalidating the "weight" hypothesis by offering a hidden, alternate cause. Children were allowed to explore these objects while the researchers measured two aspects of their behavior. First, the researchers considered what objects children placed on the machine and the order of that placement (i.e., what they did first). They also measured what children said was the cause of the machine's activation after exploring these objects.

For the most part, children tested blocks that provided confounded data (heavy with the sticker and light without the sticker) as often as blocks that provided unconfounded data (heavy without the sticker, and light with the sticker). Moreover, children's first actions were equally likely to test a block that provided unconfounded data as confounded data. But when they were asked directly what activated the machine, the majority of children (76%) reported that only the objects with the sticker would do so. Toward the end of the study, a second experimenter articulated the belief that heavy objects made the machine go. Children were given the opportunity to argue against this belief, which they did in many cases by showing the adult disconfirming evidence (i.e., the stickers). So while children in this study could articulate changes to their belief (and contradict others who held what they believed to be an incorrect belief), children were not systematic in the way they went about gathering evidence for their belief changes (see also Ronfard, Chen & Harris, 2018, 2021).

These studies can provide insight into why children struggle with scientific thinking tasks like Earthquake Forecaster. The aspects of scientific thinking that are tested in Earthquake Forecaster require children to predict which variables will be associated with earthquake risk on the basis of their current diagnosis of how the causal system works, which in turn is based on their past observations of the system. Crucially, they must continually form and revise hypotheses about how the system works in order to get to the right answer. Given the findings reviewed above, showing that children struggle to revise their beliefs even when given clear instances of counterevidence, it is not surprising that children struggle with this system, although it does provide a more realistic view of their performance with scientific thinking tasks.

Metacognition

One vitally important process that develops during the elementary-school years is metacognition: an explicit awareness of one's own thought processes. Although this process is used in a wide range of contexts, it is also crucial for fully mature scientific thinking, as we argued in chapter 1. Specifically, metacognition allows one to explicitly recognize that a currently held belief does not accord with a new piece of evidence. This ability is thus closely tied to the process of belief revision described in the previous section. Once one understands that one holds beliefs and that these beliefs could be false, then one is in a better position to understand the need to revise these beliefs on basis of incoming evidence. Similarly, in terms of testing hypotheses, one must also understand how evidence could bear on that hypothesis and that the evidence could demonstrate that the hypothesis is incorrect—that is, that one could have a false belief.

Arguably, one reason that younger children respond differently on measures of belief revision than older children could be because of the development of metacognitive capacities. Understanding that others' behavior can be motivated by false beliefs is standardly said to develop around age 4, while the more advanced abilities characteristic of metacognition (sometimes called "interpretive" or "advanced" theory of mind) develop between the ages of 5 and 8 (Carpendale & Chandler, 1996; Flavell et al., 1995; Lagattuta & Wellman, 2002; Osterhaus et al., 2016; Tang et al., 2007).

Following this work, Kuhn, Cheney, and Weinstock (2000) documented a developmental progression in children's abilities to coordinate multiple people's beliefs. Children start out with an absolutist view of beliefs, in which only one interpretation of a situation can be right; there is no room for subjectivity. Around age 7, they shift to a multiplist view, in which two people can have different interpretations, but any interpretation has equal merit (sometimes called relativism). A later developmental achievement, which might not occur until adulthood, is to attain an evaluativist view: Different interpretations of a situation are not all equally valid. This view recognizes that these interpretations must be evaluated in light of other knowledge and theories to determine which is closest to the truth, and that this process of evaluation can be subjective (see also Barzilai & Weinstock, 2015). While navigating disagreements is not quite the same as aligning one's own beliefs with the world, both processes involve some understanding of how beliefs work and how they are formed, which in turn can allow one to recognize that one is in a situation where one's beliefs could potentially change.

None of the studies of children's causal reasoning or of children's scientific thinking show that children can engage in explicit thinking about their reasoning processes or explicit reflection on how their actions can produce the data necessary for learning. As noted above, the preschoolers recruited in studies that use the blicket detector are generally too young to have this kind of metacognitive understanding and hence are not able to explicitly reflect on how their knowledge can change. Although older children are sometimes asked to reflect on their thought processes as they engage in control of variables tasks like Earthquake Forecaster, it is unclear whether this kind of metacognitive reflection is too complex for children in general. That is, we know little about how children's developing metacognitive abilities might bear on their abilities to think scientifically, particularly in cases that involve hypothesis generation and belief revision.

In chapter 7, we describe a set of studies that begins to address this gap by examining how children's developing understanding of conflicting beliefs might bear on their scientific thinking abilities.

Building Bridges

We have hoped to make clear in this chapter that there are many points of disagreement between the literature on infants' and preschool-age children's precocious causal reasoning abilities and the literature on elementary-age children's scientific thinking abilities. Briefly, studies of children's causal reasoning tend to recruit young children, present them with simple and minimally contextualized causal systems, and ask them direct questions to see whether they can make predictions about those systems' structures. Studies of children's scientific thinking tend to recruit children in elementary school (and beyond), present them with richly contextualized and causally complex systems, and allow them to explore these systems in order to draw conclusions on their own. Small wonder, then, that preschoolers tend to succeed in the former tasks while older children are often not successful on the latter.

While we do not claim to have identified all the points of disagreement that exist between these two bodies of literature, we hope to have made clear that there are many differences in the methods used to test these two groups of children. This is more than just a problem of inconsistent methods, however. The issue is that, because of these inconsistencies and their interrelations, we are not yet in a position to understand how scientific thinking develops. Do older children genuinely lack scientific thinking abilities, or do the complexities of the tasks with which they are presented mask their skills? Do younger children genuinely possess the abilities to reason scientifically, or do the tasks with which they are presented oversimply the situation to the point of being too unrealistic to test those abilities? Do early causal reasoning abilities develop into scientific thinking abilities, and if so, how? These are the questions that we aimed to tackle in the empirical work that we present in the next section of the book. Our approach has been founded on the need to take seriously the results from both sets of studies: Young children really may possess important causal reasoning skills, but they still have much to learn regarding certain facets of scientific thinking. Further, some aspects of scientific thinking may have a more prolonged developmental trajectory, and developmental trajectories may differ in different cultures. Given this, we aimed to find a way to put these results into closer dialogue. In the next few chapters, we review these studies and their results, then discuss how this work can help to paint a fuller picture of the development of scientific thinking.

This is a section of doi:10.7551/mitpress/11939.001.0001

Constructing Science

Connecting Causal Reasoning to Scientific Thinking in Young Children

By: Deena Skolnick Weisberg, David M. Sobel

Citation:

Constructing Science: Connecting Causal Reasoning to Scientific Thinking in Young Children
By: Deena Skolnick Weisberg, David M. Sobel
DOI: 10.7551/mitpress/11939.001.0001
ISBN (electronic): 9780262370615
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Weisberg, Deena Skolnick, author. | Sobel, David M., author.
Title: Constructing science : connecting causal reasoning to scientific thinking in young children / Deena Skolnick Weisberg and David M. Sobel.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021045987 | ISBN 9780262044684 (paperback)
Subjects: LCSH: Science—Methodology. | Reasoning in children. | Scientific ability. | Science—Study and teaching—Psychological aspects. | Constructivism (Education)
Classification: LCC Q175.32.R45 W45 2022 | DDC 501—dc23/eng/20211214
LC record available at https://lccn.loc.gov/2021045987