

A study of students' reasoning about probabilistic causality: Implications for understanding complex systems and for instructional design

Tina A. Grotzer¹ · S. Lynne Solis¹ · M. Shane Tutwiler¹ · Megan Powell Cuzzolino¹

Received: 19 April 2015 / Accepted: 30 August 2016 / Published online: 10 September 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Understanding complex systems requires reasoning about causal relationships that behave or appear to behave probabilistically. Features such as distributed agency, large spatial scales, and time delays obscure co-variation relationships and complex interactions can result in non-deterministic relationships between causes and effects that are best understood statistically. Causal Bayesian Research (e.g. Gopnik and Schulz in *Causal learning: psychology, philosophy, and computation*, Oxford University Press, New York, 2007) suggests that summing across probabilistic instances is inherent to human causal induction, yet other research (e.g. Schulz and Sommerville in *Child Development* 77(2):427–442, 2006) suggests a human tendency to assume deterministic relationships. Classroom science learning often stresses the replicability of outcomes, putting this learning in tension with understanding probabilistic patterns in complex systems. This investigation examined students' reasoning patterns on tasks with probabilistic causal features. Microgenetic studies were conducted in multiple sessions over a school year with students in kindergarten, second, fourth and sixth grades ($n = 16$) to assess their assumptions when dealing with tasks from four domains: social; games; machines; and biology. Later sessions attempted to scaffold students' understanding using connection making and analogical reasoning. This paper reports on the overall patterns and trends in the data. Most students held a deterministic stance at the outset; however, at least one student at each grade level reasoned probabilistically from the start. All students except one eventually revealed at least one topic for which they held a primarily probabilistic stance. The results have implications for how students reason about complex systems and for how patterns of co-variation and evidence in science are discussed.

Keywords Probabilistic causality · Complex systems · Causal models · Deterministic causality · Students' reasoning

✉ Tina A. Grotzer
Tina_Grotzer@harvard.edu

¹ Harvard Graduate School of Education, Harvard University, 421 Longfellow Hall, 13 Appian Way, Cambridge, MA 02138, USA

Rationale

How people experience evidence for causality in our complex world includes discerning contingency patterns. In our everyday experiences we perceive co-variation relationships that help us to consider the existence of causal patterns. We may see moths gather on the windows on warm nights and assume that the warmth causes the moths to emerge. We may notice that when we eat a certain food, we feel ill later; this might lead us to investigate possible mechanisms, for instance, an underlying food allergy or microbes due to poor storage techniques.

Perceived contingency patterns can vary from deterministic relationships, in which a cause (or set of causes) always leads to an outcome—flipping a switch causes a light to come on—towards increasingly probabilistic relationships that are less reliable—calling a teenager periodically results in a return call. A well-established tenet in causal reasoning is that one-to-one covariation relationships between causes and effects lead to presumptions of causality. In fact, people often mistakenly assume that high levels of correlation indicate causality; a search for an underlying mechanism may reveal a third variable that causes both. How the patterns unfold also factor into our interpretations of the relationship. If every time we flip a switch, a light comes on and after many reliable instances, then it doesn't, we assume that something in the causal system is broken, while we accept the randomness of the teenager's return calls.

Causal reasoning or induction, broadly speaking, refers to thinking related to discerning causal relationships between events, entities, or processes. It aims to go beyond patterns of correlation or co-variation to establish that one side of the relationship is responsible for the outcomes on the other. Even though it extends beyond co-variation, it is often discerned through a combination of co-variation patterns (e.g. Shultz 1982), mechanism knowledge (e.g. Atran 1995; Keil 1994), and sometimes testimony from others (e.g. Harris 2012). It is essential to the pursuit of understanding in science—the recognition of patterns and the attempt to explain why those patterns exist. Gaining a sense of the regularities in our world and what accounts for them is central to the endeavor of science and enables prediction of future patterns. (See Grotzer 2012 for further information.)

Recognizing probabilistic causal patterns is essential to discerning causality in complex systems. Here, the term “complex system” is defined broadly to include systems with many interacting variables that may be spread across space and time. The relationship between an effect and its candidate causes may be highly complex with many mediating variables. It may include the phenomenon of emergence in which interactions at lower levels lead to outcomes at higher levels. For example, consider the complex system of interactions that lead to climate change. There are many and varied contributors including methane gas which has multiple sources—from fracking to the beef industry to carbon emissions from cars and factories. There are trigger points; below which the environment has the ability to absorb the impacts so that the same causal agents result in different outcomes later in the process than earlier. For instance, only beyond a certain level will emissions trigger global climate change. There are emergent outcomes resulting from collective actions that can then set into effect feedback loops that interact with further collective action and accelerate change, for instance, as summers get hotter, people's use of air conditioners and energy goes up, increasing greenhouse gases and melting ice.

Analyzing how such complex sets of variables interact necessarily involves reasoning about forms of causality that appear non-deterministic from the perspective of the

reasoner.¹ The reasoner cannot easily connect specific actions to specific outcomes. In distributed, emergent phenomenon, single actors cannot often discern the collective impact as causally connected to his or her actions. As systems effects are spread across space and time, the attentional frame of the reasoner does not include both the cause and the effect thus the covariation relationship escapes attention (Grotzer and Solis 2015). Features of the causal relationship, including delays or distance between causes and effects, threshold or trigger points, or non-obvious variables that obscure outcomes, all can contribute to the perceived irregularity. So despite the reliability of the causal relationship, the deterministic qualities may not be perceptually available to the reasoner (Grotzer and Tutwiler 2014).

Reasoning about complex phenomena also involves considerable cognitive load. In contrast to summing across a few experiences in a serial manner, as in the above example of feeling ill after eating, one needs to collect a substantial amount of data over time. Holding it in mind quickly overwhelms our ability during everyday causal reasoning and we resort to other methods such as considering salient cases (Kahneman 2011; Grotzer and Tutwiler 2014). In formal scientific inquiry, data is typically downloaded and sources of variance are modelled supporting the ability to reason about them statistically.² Consider smoking and lung cancer, for instance. Attempting to trace the cause of lung cancer in a given case would lead to a long and, most likely, unproductive reductionist search. Modeling the relationship reveals a higher incidence of cancer amongst those who smoke, however, it is mediated by other variables such that the relationship between smoking and lung cancer is not deterministic.

This raises questions related to probabilistic causality and instruction. Students are learning to reason about causality everyday by discerning covariation between candidate causes and effects and yet reasoning about complex systems involves reasoning about probabilistic forms of causality best captured by data charts and statistical models. Modeling programs such as Star Logo or Net Logo (e.g. Resnick 1996; Wilensky 1999) are a means of helping students to reason about complex causal interactions. However, the nature of how students connect their everyday notions of causality to statistical models—a question to which probabilistic causal reasoning is central—is an important area for inquiry, whether the relationship is one of support, contradiction, or isolation. The Next Generation Science Standards (Achieve 2013) require middle school students to develop the understanding that “some cause and effect relationships in systems can only be described using probability.”

In considering models of scientific explanation relevant to science education, Braaten and Windschitl (2011) discuss the different explanatory attributes of statistical-probabilistic models and causal models as well as the instructional implications of each. They focus on reasoning from large data sets and examining trends as key aspects of statistical/probabilistic models of explanation and induction from patterns and attempts to seek an underlying causal mechanism for a phenomenon as aspects of causal ones. From an instructional stance, they argue that explanation as causation can build upon students' curiosity and can help them come to understand mechanistic principles that operate in the world. Shaughnessy (1992) has argued that deterministic models of causality are the lowest

¹ Regardless of whether or not a causal pattern is probabilistic from the stance of an objective and fully informed perceiver (the possible existence of which is a matter of debate), how people experience evidence for causality in our complex world includes contingency patterns. (See Eerola 1994).

² Probabilistic causation relates to the larger body of research on reasoning about probability (e.g. Konold et al. 2011; Shaughnessy 1992) and about the heuristics that people invoke in instances involving risk (e.g. Kahneman 2011). However, it diverges in its extension beyond the mathematical outcomes in its focus on understanding co-variation patterns (For further explanation, see Grotzer 2012).

level conceptions of stochastics given their non-statistical nature. However, as elaborated below, developmental research on causal induction has argued more recently that statistical models are at the heart of all causal induction (e.g. Gopnik and Schulz 2007). An empirical understanding of the relationships between causal and statistical models can help educators better connect these bodies of intuitive and formal knowledge.

The concept of probabilistic causation intersects statistical models and causal models of explanation. How students understand and reason about probabilistic causation can interact with how they grasp and learn to reason about causal contingencies within complex systems and what information they perceive or discount.

The developmental research, as elaborated below, makes the case that students implicitly reason using a probabilistic schema that sums across causal instances (e.g. Gopnik et al. 2004), however, that when reasoning explicitly, they have a tendency towards determinism (e.g. Schulz and Sommerville 2006). The research in this paper seeks to advance our understanding of how children's causal models incorporate notions of probabilistic covariation relationships. It examined tasks that involve probabilistic causation, but focused on tasks that are not part of a large complex system in order to minimize the confounding effects related to other aspects of complexity (spatial and temporal aspects, for instance). This limits the immediate applicability to complex systems reasoning, but is viewed here as an essential step in discerning how children's reasoning about probabilistic causality may interact with their broader systems understandings. This paper considers whether and at what ages children are able to reason about probabilistic causal relationships in contrast to deterministic causal reasoning. It starts at the youngest grades before children have been formally exposed to concepts of statistical models or reliability in science.

Background

Research on probabilistic causation and its' relationship to causal induction has a long history due to interest in what could be learned about the thresholds related to noticing covariation relationships and how variables related to time and space interact with detecting physical causality (e.g. Einhorn and Hogarth 1986). It found that children expected reliable cause-effect relationships (e.g. Bullock 1985; Bullock et al. 1982; Shultz 1982). Older students showed a greater expectation for reliability than preschoolers; it was argued that this was because older students were better at tracking the information to discern the reliability of the relationship (Shultz and Mendelson 1975; Siegler 1976; Siegler and Liebert 1974). These outcomes contributed to the argument that determinism was one of a set of fundamental principles that learners applied in their causal reasoning (e.g. Bullock et al. 1982).

However, at the heart of more recent research in cognitive development known as Bayesian Cognitive Science is the notion of a probabilistic mind that challenges these findings (e.g. Chater and Oaksford 2008). It suggests that even preschoolers follow Bayesian rules in summing across experiences in their causal reasoning. Gopnik et al. (2004) contrasted conditions in which a detector activated in response to a deterministic or probabilistic cause. Most preschoolers considered both to be causal (Gopnik et al. 2004) and in related research (Kushnir and Gopnik 2007), were able to make inferences that fit different statistical probabilities suggesting that they could track the differences. This research on causal induction suggests that summing across probabilistic instances may be a

key part of the human causal repertoire from an early age. Even so, learners may invoke a default assumption of deterministic causation. Schulz and Sommerville (2006) found that preschoolers prefer deterministic over probabilistic causality when considering a machine-like toy box mechanism in which a switch caused the top to glow. When the relationship appeared stochastic, preschoolers sought a hidden cause. In another task, preschoolers accepted that two events could be stochastically associated but that they resisted the belief that direct causes could be stochastic.

The domain of the task and children's mechanism knowledge may influence their interpretations. Schulz and Sommerville (2006) used a mechanical device; one might argue that it would be likely to invoke deterministic notions because machines are typically designed to work reliably. However, Kalish (1998) found that four- and five-year-olds expected deterministic cause-effect relationships in disease transmission—if everyone in a classroom played with a sick child, all or none would get sick. Whether this tendency extends to other domains, such as social contexts, should be explored. Further, whether it reflects a lack of familiarity with the underlying causal mechanisms or a more general assumption of determinism also bears further investigation. Research suggests that the familiarity of the task and children's mechanism knowledge may interact with whether they accept probabilistic outcomes (Kuzmak and Gelman 1986; Metz 1998). Mechanism knowledge has played a strong role in children's reasoning in other instances in which they could not rely on co-variation alone (Grotzer and Solis 2015; Sandoval and Cam 2011); researchers have argued that it is a key component in causal reasoning (e.g. Ahn and Kalish 2000).

The extant developmental research studies were conducted in labs and occurred in one session, typically in one context, and without learning supports. This offers insights into what children are capable of with limited opportunity to express or build understanding. However, given the centrality of probabilistic causal reasoning to complex systems understandings, the question of the learnability of probabilistic causal reasoning is paramount. This is a question for the learning sciences and instructional design. For the purposes of future instructional design, it is important to investigate what understandings children reveal with extended opportunities to explore probabilistic causal tasks, both within and across domains, and with support for abstracting the probabilistic causal concepts from the tasks.

Research questions and hypotheses

In an effort to further illuminate how children understand and reason about probabilistic causality, we conducted year-long microgenetic studies of students' reasoning given multiple opportunities and across different contexts with which they might be familiar. Of particular interest was whether they would allow for probabilistic causality or default to deterministic causality and if the later, whether their interpretations shifted over the course of multiple opportunities to reason about probabilistic causality. The study investigated how kindergarten, second, fourth, and sixth grade students reasoned about tasks with probabilistic causal features in open-ended contexts and given scaffolds to make the nature of the causal patterns explicit. The tasks in the study did not focus on complex systems thinking, rather the key feature of the tasks was probabilistic causation without confounding, complexifying features. The study aimed to find out what was possible in students' thinking by offering instructionally promising opportunities (multiple opportunities

over time, a range of domains, and tasks that were likely to be familiar to students, and eventually, analogical scaffolds) rather than to control for specific variables.

The study investigated the following questions:

- (1) How will students across four grade levels respond to tasks characterized by probabilistic co-variation? Will they default to deterministic interpretations or use probabilistic ones?
- (2) Will the pattern of their responses change with repeated opportunities to explore the probabilistic tasks and/or scaffolding?

Based on the extant research, we expected that students would initially default towards primarily deterministic ideas, particularly in domains that were mechanistic in character. However, we hypothesized that with repeated opportunities in familiar contexts such as games and some support in the form of analogical mapping between tasks, students might shift towards a less deterministic stance.

Methods

Design

In depth, microgenetic studies were carried out with sixteen students (4 kindergarteners, 4 second graders, 4 fourth graders and 4 sixth graders) over the course of a school year. Students from public schools in an urban district in the Boston area that is predominantly Black and Latino with lower to middle SES, participated in the study. Across the school year, students were interviewed in depth on tasks with varying levels of stochastic behavior from science and beyond. At least four sessions were conducted with each student. However, in keeping with microgenetic study design (e.g. Siegler and Crowley 1991), the number of sessions was variable with additional sessions added at points during which students appeared to be experiencing change in their reasoning structures or to be productively engaged in figuring out concepts. A total of 108 sessions (minimum of 4, maximum of 10, with a mode of 7) of between 30 and 45 min in length were conducted. Student learning was assessed following each session and this informed the design of the next session. In some cases, the task was repeated to see what students would reveal with more time. In other instances, as elaborated below, the task was modified to reveal new aspects of students' understanding. The modifications were attentive to theory but not driven by theory a priori. So for instance, in the Funny Bunny Game described below, a modified version was designed to strip away some of the game tasks that were extraneous to discerning the probabilistic causality. This was in an attempt to see if more of the kindergarteners would pick up on the probabilistic causal pattern if there was less cognitive load. Probes and tasks were also slightly modified when students appeared to have certain conceptions in mind and those modifications had the potential to reveal more about the students' thinking.

Unless the student generated examples from other contexts, only one context was focused on in the initial sessions in order to avoid adding to the cognitive load. In later sessions, more than one context was discussed to allow for contrasting examples and mapping the analogous deep structures. Interviews proceeded from open-ended to increasingly structured in order to first assess how students framed the concepts and to then assess the accessibility of concepts related to the probabilistic behavior with targeted questions. Support (in the form of analogical reasoning) was only offered in the final session. Most sessions were conducted one on one with the researcher but for some of the sessions with games, students were involved in an

activity with another student. In the final sessions, scaffolds were used; the scaffolds compared the study tasks to familiar examples and compared analogous causal forms in different problem contexts through “mutual alignment” (e.g., Kurtz et al. 2001). Sessions were videotaped for later coding and analysis. This paper focuses on the overall patterns in the data and on whether students would reveal instances of understanding of probabilistic causation and depart from deterministic framing.

Task design

The tasks were from four domains: games, biological, mechanical, and social and included seed planting, hatching chicks, bubble gum machines, videotapes of brief social interactions, and a set of games. Types of tasks that children might be familiar with were intentionally chosen to elicit their existing knowledge. This departs from other studies where the intention is to create causes that children would not have prior expectations about (e.g. Schulz and Sommerville 2006; Sobel and Buchanan 2009). The tasks did not all fall neatly into one domain category or another, for instance, some of the games had mechanical aspects in addition to the game features. Further, mechanical tasks that would potentially invite concepts of probabilistic causation (i.e. a bubble gum machine that did not regularly deliver candy) were chosen instead of a machine characterized by highly reliable causal processes. Given the small number of students, the order of the tasks was not counterbalanced across subjects thus limiting the conclusions that can be drawn.

Tasks with the following features were chosen: (1) cognitive load that was not directly related to the probabilistic causal features of the task was low (or minimized by minor modifications); (2) authenticity (that could be maintained with simple controls, for instance, in growing seeds, the researchers adjusted the number of seedlings “that grew” without impacting the authenticity); and (3) manageability in the classroom. Tasks with complexifying causal features were included if the features were inherent to perceiving probabilistic causality such as in the non-obvious causes involved in knowing what happens to seeds under the ground and the challenge of prediction and finding out over delayed periods of time.³ Some tasks were modified for use with diverse populations, for instance, in a game called, “Don’t Wake Daddy,” a pop-up Caucasian male figure, was covered with a teddy bear form and was renamed, “Don’t Wake the Sleeping Bear.” In some tasks, children could see the causal mechanism if they tried. For instance, in a game called “Funny Bunny” (See Fig. 1.), students were not stopped from turning the game over to examine it.⁴ The classroom experience that students had with the concepts varied; for instance, in the second grade, the teacher hatched eggs. Most students participated in all of the tasks with a few exceptions due to extended absences. All of the fourth and sixth graders and some of the younger students were also given connection-making and analogy tasks to see if it might help them abstract the probabilistic causal schema. Analogy has

³ In authentic biology problems, the probability of particular outcomes cannot be calculated a priori as they can when working with games. Manipulating the number of seeds that sprouted in the seed planting task without the students’ knowledge enabled us to plan a priori which probabilistic outcomes to test (acknowledging that these manipulations depart from authenticity).

⁴ If the study was only looking at children’s grasp of co-variation patterns, we would have obscured this information from students, but reasoning about the broader questions of probabilistic causation can be supported by both mechanism and covariation information. Further, while turning the game over would reveal the mechanism and offer information that one could use to deduce a pattern, it involves transferring that information dynamically to the top of the board and being able to track how two moving plates under the game interact to result in whether and where a hole opens.



Fig. 1 Funny bunny game

been shown to be effective in helping learners abstract the underlying structure of a causal model in complex text (Clement and Yanowitz 2003).

Games

Students were presented with the following games: Funny Bunny, Don't Wake the Sleeping Bear, and Uno Attack. They were given a chance to play them and to reason about the probabilistic causal patterns in each as follows. Students played games either one-on-one with a researcher or with other children in their class. On a few occasions, this included more than one child in the study by design, often because they held similar ideas or because one student held a particular idea that might push another towards an insight (as elaborated in the results section).

Funny Bunny is a commercial game by Ravensburger. Players move their rabbits (each has a set of four) along a path that has two different loops up a hill to the top of a large carrot with the goal of getting a rabbit there first. Cards tell how many steps to move, however, some cards direct players to click the carrot in the middle of the game board causing a hole to open up somewhere along the path most of the time that one's rabbit could fall through. The location of where the hole opens up gives the appearance of being stochastic in the following ways. Initially, there is no indication that the hole in the path moves. Upon the turn of the carrot, the hole moves along the path, alternating between the top row of the path and the bottom row of the path. Periodically, no hole opens at all. Figuring out whether a hole will open and which hole will open involves detecting that some spaces hold the possibility of opening (nine out of 26 are "wiggly" or "soft" whereas others never open and are always safe); detecting that the holes move in a clockwise fashion around the board; that they alternate between the top and bottom rows; and that the hole also disappears at a certain point in the rotation. The game provides evidence that clicking the carrot does not deterministically open a hole along the bunny path. As students played the game, researchers probed their understanding of the relationship between the turning of the carrot and the opening up of a hole. The mathematical relationship between whether the hole opens or not makes predicting the outcome on any given turn highly complex thus it appears random. However, students are able to predict that some spaces will never open up because they are not "soft." Researchers probed students on whether they could predict what would happen before the carrot was turned and why or why not.

For some kindergartners and second graders, the task of keeping all of the information in mind (where their bunnies were, what the cards meant) combined with the chance of where the card told them to go was challenging enough that it was difficult to discern what they understood about when/whether the hole opens up. Therefore, the game was modified to a version called *Last Bunny Standing* that offered repeated opportunities for the child to choose a spot and then turn the carrot. In this version, the player has to figure out where to put a bunny on each turn so that it will be safe when the carrot is clicked. It eliminates some cognitive load related to game strategy (how many rabbits to put on the board and how the randomness of the shuffled cards interacts with outcome) and focuses directly on the goal of figuring out where the hole opens given its seemingly stochastic nature. This enabled the researchers to see how his or her strategy changed over the course of the game.

A second game, *Don't Wake the Sleeping Bear*, was modified from a Hasbro game (Fig. 2). The goal is to get to the finish line without waking up the sleeping bear. However, when spaces are landed upon, the player must push the button on an alarm clock a given number of times and if the bear pops up, must return to start. The relationship between pushing the button and the bear popping up is probabilistic from the perspective of the player. The number of button pushes ranged from 6 to 20 and each of the three physical versions of the games that were used in the study had a different pattern of when the bear would pop up. Further, if the students were not tracking how many pushes others had entered, it could pop up on their first push (presuming five pushes occurred during other turns). The mechanism is internal to the device so there is no visible mechanism information. As students played, the researcher probed the students' predictions about whether the bear would pop up and how s/he knew in addition to whether it was predictable or not. As students played this second game, they were also asked questions about how the game was like the earlier bunny game(s) or not.

The third game was *Uno Attack*, a variation of the game, Uno, in which the player attempts to be the first to get rid of all of his or her cards, but with an automated card dispenser. A player pushes the button on the dispenser and sometimes it dispenses cards (a seemingly random number of them) though most times it does not. There is no discernible regular pattern. There is no visible mechanism to account for what happens. When the dispenser is opened up to add cards, one can see a flywheel, however, it does not work when opened up so it is not possible to test under what conditions it shoots cards or not. As students played, the researcher probed their predictions and ideas about whether or not cards would shoot out and how they knew.



Fig. 2 Don't wake the sleeping bear

Biology tasks

Students were given two authentic biology tasks to reason about. The first involved planting seeds. The student was told that s/he needed to have a certain number of bean plants in a few weeks. S/he was then given a peat pot, soil, and seeds and invited to plant the number of seeds that s/he thought should be planted in order to end up with the necessary plants. Each student engaged in the task two to three times. The researcher took the pots of seeds and brought them back the next week. This enabled the study team to manipulate the number of seeds and growth patterns to present different outcomes the following week as a basis to interview students' interpretations of different levels of stochastic results. Probes sought to reveal students' predictions, whether they thought it was possible to know how many plants would grow and if not, why not. Probes also asked students about prior experiences with growing seeds.

The second task involved making predictions about hatching chicken eggs. The student was asked to predict what the inside of an incubator might look like in 22 days after eggs were set inside it. S/he was told that eggs typically hatch in 21 days, was given a drawing showing eight eggs, and an opportunity to predict and draw the outcome that s/he expected. Following the student's prediction, probes focused on whether the student had ever seen an outcome in which less than the number of eggs in the incubator had hatched and what s/he thought about that.

Social tasks

Two social tasks were used. In the first, the student was shown two brief video clips. In one, a girl is calling her mom for help with her homework. The rate of calling to response varies as follows: (1) girl calls, mom responds, (2) girl calls, calls again, calls again, mom responds; (3) girl calls, calls again, calls again, calls again, and calls again, then mom responds. The student was asked what causes the mom to come and how the versions are different from one another. A second video shows a boy pestering his sister by taking her markers and she responds. The rate of pestering to response varies as follows: (1) boy takes marker, sister responds, (2) boy takes a marker, takes another marker, takes another marker, sister gets mad; (3) boy takes a marker, then takes another and another and another and another, then the girl gets mad. The student was asked, "What makes the girl get mad?" "How do you know? And "Does taking the marker always make the girl mad?" In a second task, the student was asked "If someone cheats on a test or homework, do they always get caught?"

Mechanical task

The student was shown a mechanical candy dispenser and were given coins and invited to make it work. The dispenser dispensed between zero and five candies with each turn of the crank with a mode of five. The actual mechanism for dispensing candies was not visible given the number of candies in the dispenser. Students could detect some information about the mechanism, however, because the crank was less easy to turn on some turns when it would dispense no candies and on others, turning it very slowly yielded higher returns. Probes included asking the student to predict what would happen when the handle was turned, whether it was possible to know, and why or why not. Some students asked to

record the numbers of candies with each turn and were given paper and pencil to record it on.

Scaffolds for discerning the probabilistic causal structure through connection-making and contrasting analogies

The final sessions focused on connection-making across the tasks and other probabilistic causal scenarios as well as to students' own experiences. The student was given probabilistic causal examples (some of which were related to earlier study tasks and some that were not but seemed likely to be within students' experience) to reason about and were asked to think about other examples that might be like them. These examples included volcanoes erupting, an electric pencil sharpener that does not consistently work, and choices to be made in social scenarios. The probes also mapped back to the earlier game tasks, for instance, "Can you think of ways that cheating and getting caught are like the way that any of the games worked? The technique of "mutual alignment" (e.g. Kurtz et al. 2001) was used, in which students map between two analogical cases, each of which is only partially understood, with the goal of enhancing understanding of each. The student was asked to map back and forth between analogical problems making attempts to discern similarities and differences and to use each to further inform understanding of both. The cases were contrasting in their surface features but similar in deep structure and students were guided to compare them (Schwartz et al. 2011).

Scoring and analysis

Each session was transcribed and analyzed using an etic coding process to identify whether students used probabilistic or deterministic reasoning, how they reasoned and whether any shifts in reasoning appeared to be taking place. The data sources include audiotape and videotapes and researcher notes from each session. A coding scheme was developed to identify deterministic (D) and probabilistic (P) statements.

The deterministic thinking code was applied to statements or behaviors by the student suggesting that s/he considered that a cause always leads to an immediate effect (i.e., one-to-one correspondence). S/he used language like, "I'm sure", "definitely", "always". The code was also applied when students sought a underlying consistent pattern or hidden cause to account for the outcome (as in Schulz and Sommerville 2006) even if s/he did not demonstrate knowledge of what it actually was but was convinced that one existed and/or was actively seeking to figure it out. The probabilistic thinking code was applied to statements or behaviors by the student suggesting that s/he considers that the connection between cause and effect is uncertain (i.e., not one-to-one correspondence) and uses language like, "might", "maybe", "probably", "sometimes", "random", "lucky." The code was used when the student was clearly reasoning about cause and effect relationships (for instance, "That makes the bunny fall.") It was applied conservatively in that probabilistic language (e.g., maybe, probably) may be used by the child as a reflexive speech habit rather than an expression of their reasoning, for instance, when a student said it in the course of figuring out what they were thinking such as in, "Yeah, I'd probably say that is true." Examples of deterministic statements include, "A hole IS going to open" as soon as someone starts to turn the carrot in the Bunny Game, "If I want five plants, I need to plant five seeds." "Where is the hole? It is missing, there has to be a hole." Examples of probabilistic statements include, "When you click the button, you never know if it is going to wake up the bear." "When you plant seeds, you can't make them grow, so you don't

know how many plants there will be” “Calling her mom doesn’t cause her to come, well not always.” “It is like a surprise attack, you just never know.”

Transcripts were independently coded by two researchers. One coder scored 100 % of the data while the other scored either 25 % of the pages in each student’s transcript or 25 % of each student’s statements, whichever was greater. This was to make certain that a significant portion of each student’s statements were coded for reliability even if s/he spoke less or participated in fewer sessions. Given the range of ages in the study and the variations in how students talked about the tasks, having the second coder code a proportion of each student’s transcripts was important for determining the reliability of coding across the entire sample. The coding process took into account only statements that both coders independently selected from the transcripts as relevant to deterministic or probabilistic reasoning.⁵ The detection rate between coders was calibrated over two rounds of scoring until the range of detection was 90 % across transcripts. Following the selection of the statements, each statement was coded as deterministic or probabilistic independently by each coder and Cohen’s Kappa was conducted on their level of agreement on each category of statement within grade level (given variations in how students of different ages expressed their understandings). These yielded Kappas all within the “Very Good” range (Kindergarten = .910; 2nd Grade = .811; 4th Grade = .860; and 6th Grade = .915).

The coders also noted emergent themes in the data. Two coders read the transcripts and identified sets of themes. These were discussed and themes that both coders agreed upon are included. ATLAS.TI was used to conduct the emic, grounded coding of transcripts. Narratives were developed describing how the students changed in their explanations of the task and what leverage points could be identified that may be useful in teaching concepts that embed the probabilistic causal concepts.

Results

The quantitative data and qualitative findings revealed support for a number of patterns in students’ reasoning as elaborated below. There were not significant differences in the patterns of reasoning between the grade levels, therefore the results are articulated more generally and examples are drawn from the most illustrative cases rather than balanced across grade levels.

Initial deterministic tendencies

Most students across all four grades levels were initially quite deterministic in their responses. Table 1 illustrates that ten out of the sixteen students approached the initial tasks from a deterministic stance (K: Jeremiah, Tanika, and Tyrone; Grade 2: Rajon, Kaia, and Phillippe; Grade 4: Devon and Elias. Grade 6: Andre and Sai). This was so despite the fact that the first set of tasks focused on games and offered repeating opportunities for students to discover the probabilistic nature of the outcomes. Figure 3 shows the level of deterministic responses given by students within each grade level during the first session, a game task. While the percentages range from 0 % to 100 % of responses, three out of four

⁵ This is important in a study that is designed to assess the extent of particular types of thinking in students’ reasoning. It avoids the situation of having one researcher select the statements and then the reliability carried out only on whether the two coders agree on how the statements are coded, in which case, there is no reliability check on what was left behind in the transcript.

Table 1 Percentage of Deterministic and Probabilistic Responses by Student within Session and Task

Kindergarten											
Task/order ^b	Jeremiah ^{1,B}		Khloe ^{2,W}		Tania ^{2,B}		Tyrone ^{1,B}		Stance- # of students ^d		
	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deter- ministic	Balanced ^a Prob- abilistic	
1- Game (Bunny)	86 % (12)	14 % (2)	29 % (5)	71 % (11)	100 % (7)	0 % (0)	96 % (29)	4 % (1)	3	0	1
2- Game (Bunny)			73 % (11)	27 % (4)	96 % (26)	4 % (1)	93 % (52)	7 % (4)	3	0	0
3- Game (Bear)	86 % (32)	14 % (5)			66 % (16)	33 % (8)			2	0	0
4- Seeds	100 % (4)	0 % (0)	78 % (7)	22 % (2)			100 % (7)	0 % (0)	3	0	0
5- Gumball	53 % (16)	47 % (14)	0 % (0)	100 % (4)	31 % (11)	69 % (25)		0	1	2	
6- Social	72 % (28)	28 % (11)	0 % (0)	100 % (13)	18 % (5)	82 % (23)	78 % (7)	22 % (2)	2	0	2
7- Analogies			52 % (12)	48 % (11)					0	1	0
Second grade											
Task/order	Rajon ^{1,B}		Kendra ^{2,B}		Kaia ^{2,B}		Phillipe ^{1,B}		Stance- # of students		
	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deter- ministic	Balanced ^a Prob- abilistic	
1- Game (Bunny)	82 % (54)	18 % (12)			66 % (2)	33 % (1)	86 % (44)	14 % (7)	3	0	0
2- Game (Bunny)	63 % (19)	37 % (11)							1	0	0
3- Game (Bear)	93 % (26)	7 % (2)			100 % (3)	0 % (0)	81 % (25)	19 % (6)	3	0	0
4- Game (Bear)	82 % (9)	17 % (2)			63 % (20)	37 % (12)	76 % (34)	24 % (11)	3	0	0
5- Game (Uno)			42 % (11)	57 % (15)			61 % (17)	39 % (11)	1	1	0
6- Seeds	55 % (5)	45 % (4)			100 % (2)	0 % (0)	90 % (10)	10 % (1)	2	1	0
7- Game (Uno)					20 % (3)	80 % (12)			0	0	1
8- Incubator	77 % (17)	23 % (5)			71 % (5)	29 % (2)	61 % (13)	39 % (20)	3	0	0
9- Gumball	57 % (4)	43 % (3)			50 % (4)	50 % (4)	40 % (2)	60 % (3)	1	1	1

Table 1 continued

Second grade											
Task/order	Rajon ^{1, B}		Kendra ^{2, B}		Kaia ^{2, B}		Phillipe ^{1, B}		Stance- # of students		
	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deter- ministic	Balanced ^a abilistic	
10- Social	100 % (6)	0 % (0)	0 % (0)	100 % (1)	58 % (11)	42 % (8)	71 % (5)	29 % (2)	3	0	1
11- Connections	54 % (14)	46 % (12)	0 % (0)	100 % (10)	27 % (6)	73 % (16)			0	1	2
12- Analogies	45 % (5)	55 % (6)	40 % (10)	60 % (15)	66 % (8)	33 % (4)			1	1	1
Fourth grade											
Task/order	Devon ^{1, B}		Ruby ^{2, B}		Iris ^{2, B}		Elias ^{1, B}		Stance- # of students		
	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deter- ministic	Balanced ^a abilistic	
1- Game (Bunny)	74 % (20)	26 % (7)	58 % (11)	42 % (8)	44 % (4)	56 % (5)	62 % (18)	38 % (11)	3	0	1
2- Game (Bear)	65 % (13)	45 % (7)	77 % (7)	23 % (2)	65 % (13)	35 % (7)	68 % (15)	32 % (7)	4	0	0
3- Gumball	56 % (14)	44 % (11)			50 % (7)	50 % (7)	63 % (17)	37 % (10)	2	1	0
4- Seeds	69 % (11)	31 % (5)	86 % (6)	14 % (1)	71 % (5)	29 % (2)	100 % (10)	0 % (0)	4	0	0
5- Incubator	83 % (5)	17 % (1)	25 % (1)	75 % (3)	40 % (2)	60 % (3)	100 % (5)	0 % (0)	3	0	1
6- Social	34 % (6)	66 % (12)	66 % (2)	34 % (1)	19 % (3)	81 % (13)	25 % (2)	75 % (6)	1	0	3
7- Gumball			80 % (4)	20 % (1)					1	0	0
8- Connections	34 % (13)	66 % (25)	41 % (7)	59 % (10)	20 % (3)	80 % (12)	33 % (5)	66 % (10)	0	0	4
Sixth grade											
Task/order	Andre ^{1, B}		Sai ^{1, A}		Elena ^{2, H/L}		Heather ^{2, W}		Stance- # of students		
	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deter- ministic	Prob- abilistic	Deter- ministic	Balanced ^a abilistic	
1- Game (Bear)	92 % (24)	8 % (2)	98 % (47)	2 % (1)	45 % (5)	55 % (6)	0 % (0)	100 % (5)	2	1	1

Table 1 continued

Task/order	Andre ¹ , B		Sai ¹ , A		Elena ² , H/L		Heather ² , W		Stance- # of students		
	Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic	Deter- ministic	Prob- abilistic	Deter- ministic	Balanced ^a Prob- abilistic	
2- Game (Bunny)	100 % (17)	0 % (0)	75 % (6)	25 % (2)	60 % (12)	40 % (8)	46 % (6)	54 % (7)	3	0	1
3- Gumball					60 % (18)	50 % (10)	50 % (10)	0	1	1	
4- Game (Uno)	90 % (28)	10 % (3)	92 % (12)	8 % (1)	60 % (6)	40 % (4)	50 % (3)	50 % (3)	3	1	0
5- Seeds, incubator	58 % (15)	42 % (11)			c	c	c	c	1	0	0
6- Gumball	43 % (12)	57 % (16)	82 % (27)	18 % (6)					2	0	0
7- Seeds, incubator			95 % (19)	5 % (1)					1	0	0
8- Social	56 % (13)	44 % (10)	63 % (19)	37 % (11)	27 % (7)	73 % (19)	51 % (18)	49 % (17)	2	1	1
9- Connections	35 % (14)	65 % (26)	65 % (35)	35 % (19)	44 % (16)	56 % (20)	51 % (18)	49 % (17)	1	1	2

^a Balanced-Cases with a differential of 10 % or less

^b Tables preserve order of tasks. Grayed out sections of table may give the appearance of missing data, however, these blanks reflect the nature of microgenetic studies and the choices made during the study to focus more sessions on those students who appeared to be struggling with concepts or undergoing changes that informed the research questions

^c Missing data due to an audio recorder malfunction

^d All names are pseudonyms

1- Boy; 2-Girl

B Black or African American, *W* White, *A* Asian, *H/L* Hispanic/Latino, As per the Federal Aggregate Reporting Requirements for Reporting Data on Race or Ethnicity for Elementary and Secondary Students (<http://www2.ed.gov/policy/rschstat/guid/raceethnicity/questions.html>)

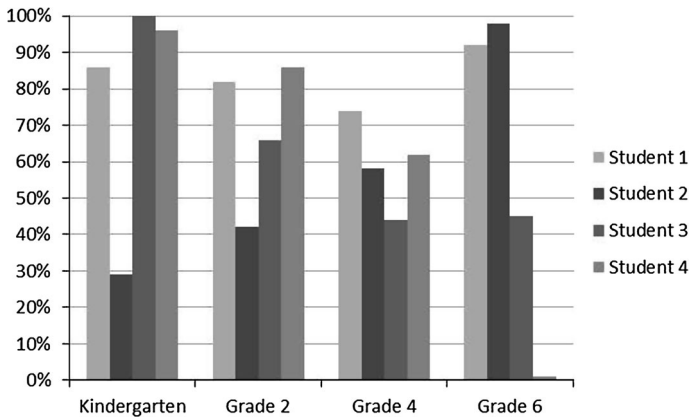


Fig. 3 Levels of deterministic responses by student within grade level on the first task (games)

students in each grade level have a predominantly higher percentage of deterministic than probabilistic responses.

An example of expecting a deterministic outcome is seen in kindergartner Tyrone's first session. He explains that turning the carrot will definitely make a certain hole open up. "When you walk all the way over here, you fall in. And after, you won't make it to the carrot." Then he tells the interviewer that her rabbit will fall.

T: Oh, my god. You're going to fall in there.

I: I don't think so. Why do you think I'm going to fall?

T: Do you see the path going that way? And then, all the way over there? Then after, over here, then fall in...then after you won't make it up there.

I: So if I get here, am I definitely going to fall in?

T: Yeah. If you get one [a turn the carrot card].

Sixth grader, Andre analyzes the game and tells the interviewer where the hole will open. He points to four different places on the board. "Right here.. right here...right here, and right here. Well since all of these are...[pokes at those spaces]—I noticed that all of these yellow spaces—the plastic cover is slightly lower than the rest] It's going to click and then the bunny will fall."

A Few Probabilistic Causal Reasoners from the Outset

Some students used probabilistic interpretations from the beginning and seemed flexible in allowing for the possibility of probabilistic or deterministic responses. As revealed by Table 1 and Fig. 3, there was at least one student at every grade level who started out noticing probabilistic patterns in the games from the first session. A total of five students revealed a tendency towards probabilistic reasoning from the very first session; one kindergartner, second grader, and fourth grader and two-sixth graders. Interestingly, in this small study, all of these students were girls. The responses of these students were more balanced across response types and shifted with predictions in particular domains. Contrast the difference between the percentages of statements in Table 1 made by Tyrone, Phillippe, Rajon, and Sai to those of Kendra, Iris, Ruby, and Elena (See Fig. 4).

In the first session, kindergartner Khloe engages notions of subjective uncertainty, the idea that personally one may not know something, and objective uncertainty, the idea that something can't be known. She attends to the patterns in the game but holds a stance of uncertainty about what will happen. In the Funny Bunny game:

K: *Gets a Carrot.*

I: Khloe, before you turn that, what do you think it's going to make happen?

K: Either one of the other holes are going to open, or one of the rabbits is going to fall down.

I: How do you know which?

K: I don't know which one. I cannot.

In the connection-making tasks, Elena explicitly refers to the uncertainty inherent in the tasks:

I: So does her calling always mean that her mom shows up?

E: No.

I: Um, ok, so how is this like the Uno game?

E: Because in Uno, never know how many cards are gonna come out and you never know how many times you have to push it. It's the same thing like that, cause, she never know how many times she gonna call her mom, before she actually comes.

On cheating....

E: There is a 50/50 chance that you can get caught or you cannot get caught, well they are not really the same, and then, oh yeah, and then there could be a possibility that many times you click on it that many cards may come out and sometimes it could be less or more cards depending on how you click it.

The students who allowed for probabilistic outcomes and were varied in their responses across domains, introduce a different reasoning pattern and perhaps, points of leverage that can be used in service of instruction.

Shifts towards recognizing probabilistic causality

Some students shifted their stance across the domain content and across the trials (which were confounded in this study). Seven students could be characterized as shifters across the

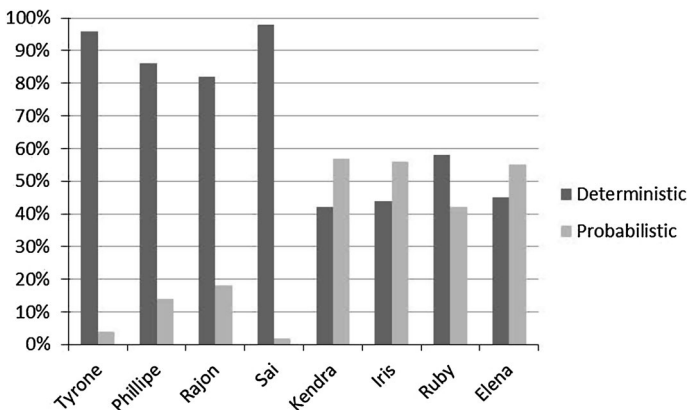


Fig. 4 Deterministic versus probabilistic responses by students on the first task (games)

sessions and domains. Contrasting Fig. 3 which reveals the percentage of deterministic responses on the first task to Fig. 5 which reveals the percentage of deterministic responses on the last task illustrates the levels of change. The percentages of deterministic responses went down across all the grade levels [although unevenly in kindergarten. Notice for instance, Tyrone who experiences little change and Tanika who shows a steady shift in her responses (See Figs. 6, 7)].

Even before the transition to a higher percentage of probabilistic causal or to balanced comments, one can see transitional comments (defined by a differential of less than 20 % between the categories or more than a 25 % change from the previous deterministic remarks) or bouncing back and forth. For instance, kindergartner Tanika shifts from deterministic causal comments of 96 to 66 to 31 % in her third, fourth, and fifth sessions where we see predominantly probabilistic comments (See Fig. 6.) and second grader Kaia goes from 100 % in her fourth session to 20, 71, 50, 58, 27–66 %, respectively, in the next six sessions. After repeated examples of probabilistic outcomes, students started to allow for the possibility that some causes did not reliably lead to the effect. The extent of the shift varied as illustrated by Figs. 8 and 9 where we see that Andre shifted across the sessions from a predominantly deterministic view to a predominantly probabilistic one, yet Sai's responses become less deterministic but are always predominantly so. Students' language shifted to include the possibility that they could not predict the outcome in every case, "just most of the time." They predicted what a best guess would be even if it "would not always be right." These shifts are important because they address the potential learnability of the concept. However, a limitation of the study is that the order of the tasks is confounded with the domain so it is possible that the increases are partly due to how much probabilistic reasoning certain domains elicit.

When they were unable to determine strategies for how to predict, some of the younger students redefined the cause from a mechanistic to an anthropomorphizing one (for instance, Rajon talked about the game trying to trick him). Eventually, they questioned whether it was possible to predict. Whether they thought the particular problem was too hard to figure out or were beginning to recognize the problem as having probabilistic characteristics was not clear.

Andre shifted his language to become more tentative and began to talk about how in some cases, "you couldn't know." As discussed above, Andre initially made deterministic

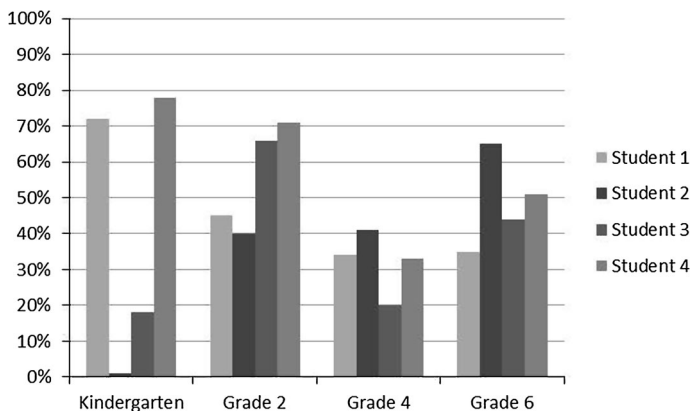


Fig. 5 Levels of deterministic responses by student within grade level on the last task (analogies/connections)

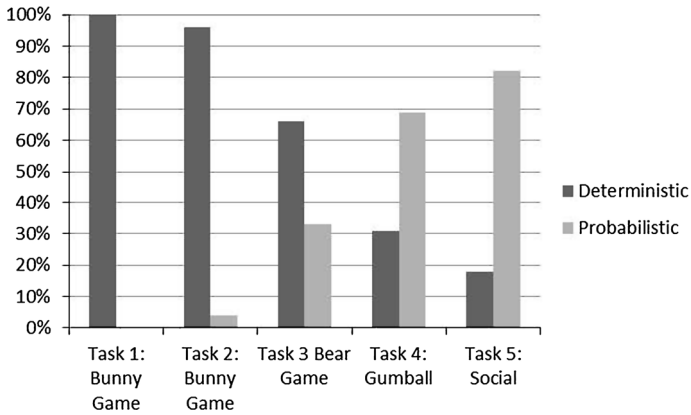


Fig. 6 Percentage of deterministic versus probabilistic responses across tasks by Tanika (kindergarten)

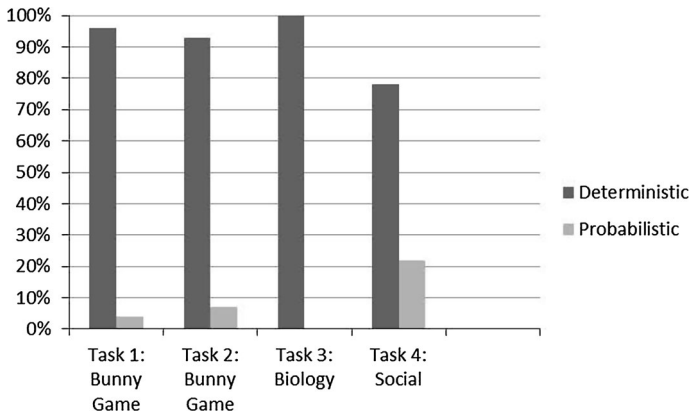


Fig. 7 Percentage of deterministic versus probabilistic responses across tasks by Tyrone (kindergarten)

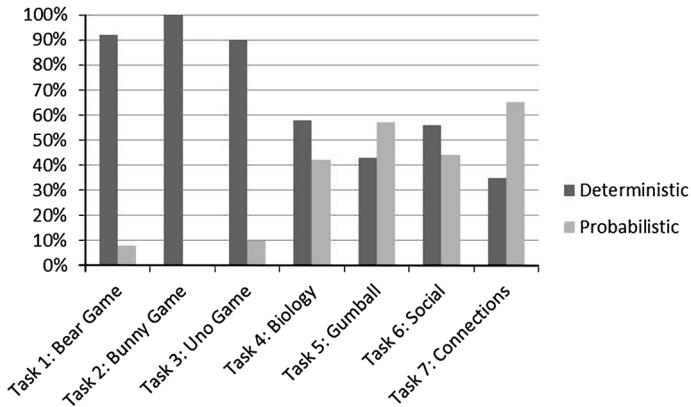


Fig. 8 Percentage of deterministic versus probabilistic responses across tasks by Andre (grade 6)

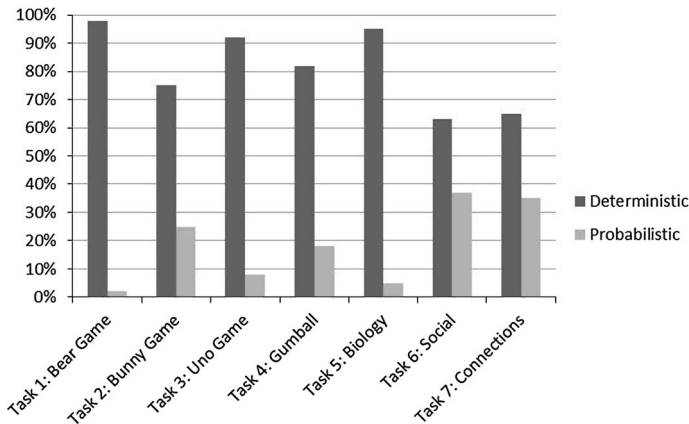


Fig. 9 Percentage of deterministic versus probabilistic responses across tasks by Sai (grade 6)

statements in the first and second sessions. Yet in the fifth session when he is asked to draw conclusions from across the activities, he made more probabilistic statements, for instance, “It was kind of random. It was a random order that –I mean, a random number of times how when you press the button, the card would come out... random does mean unpredictable, can’t predict it, hard to predict, don’t know when it’s going to come.” When asked about connections, he commented about volcanoes “Can’t predict. Can’t always be sure when a volcano’s going to erupt...” This increase in tentative language has been found to support more dialogic interaction in the science classroom (Kirch and Siry 2010) and may offer an opening for instruction that explicitly reflects upon probabilistic causality. When Andre reflected back on his experience after the study, however, he didn’t seem to realize how deterministic his earlier responses had been. He commented that when he and a classmate were looking for patterns in one of the games, “I thought Sai was crazy. I thought, we’ll never be able to figure it out. I didn’t know what he was doing!” However, in the video for that session, he is actively helping Sai gather data and seemed to believe that they could figure out a deterministic pattern. Further, their deterministic stance enabled them to learn a lot about the patterns in the game.

Variation across domains

Students’ reasoning shows considerable variation across domains. As illustrated in Fig. 10, the social task and the gumball machine elicited the most probabilistic causal reasoning. The gumball machine offers immediate cause and effect outcomes. Only three of the 16 students offered deterministic explanations; most students offered probabilistic or balanced ones (cases with a differential of less than 10 % between response types). It was quite compelling for some students, kindergartner Khloe, for example. It is also possible that narratives about gumball machines are affectively powerful as they are often one of the early opportunities that children have to learn that the world is not fair. For a child who is expecting one to one correspondence between the cause (putting the coin in and turning the handle) and the effect (getting a gumball), probabilistic outcomes are a clear violation of expectation that is immediate, within the same attentional frame, and that may become salient enough to anchor future experience. As second grader Kaia expressed indignantly,

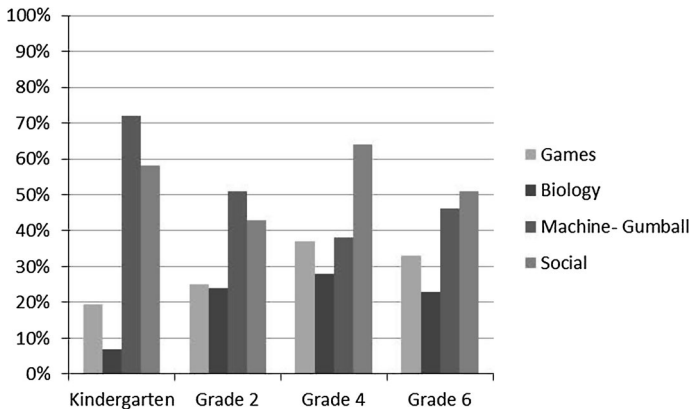


Fig. 10 Percentage of probabilistic responses by domain within grade level

“One time we went to a store. They had a gumball machine. My brother put one quarter in and he got NOTHING.”

Interestingly, seeds did not elicit probabilistic causal reasoning as much as one might have expected. Only two out of twelve kids who participated in this task shifted to a predominantly probabilistic explanation. Reasoning about the growth of seeds as an outcome involves more complex causal features than reasoning about gumball machines. In both cases, the process is non-obvious but the extended time frame for the seeds puts it in a different attentional frame and this makes it difficult to reason about the co-variation relationships (Grotzer and Solis 2015). One of the students who maintained a deterministic stance, Sai, when asked if he had experience with seeds, gave an account of planting seeds with the program “City Sprouts.” However in each case, he talked about planting the seeds but not going back to see them later so he had no way to experience the covariation and its level of regularity. When students related experiences that had a strong affective outcome in a deterministic direction, it seemed to hold a powerful influence on their reasoning. This fits with the concept of availability heuristic in which powerful narratives override statistical probabilities (Kahneman et al. 1982). In another biology example, when predicting that all eight eggs in the incubator would hatch, Sai told a detailed story of visiting his grandfather who lives on a farm in Bangladesh and his grandfather giving him four eggs in a towel to take on the plane with him. He told of all four eggs hatching on the plane and how cute they were, but then the authorities took them when he landed in Boston.

Future research could look within content domains with repeated trials to assess learning that is not confounded with domain content. Within the microgenetic framing of the study here, we carried out a version of this with second grader, Rajon, over four sessions with games as an in-depth analysis of his reasoning, as reported elsewhere (Tutwiler and Grotzer 2014). In this limited case, he persisted with deterministic interpretations. Another second grader, Kaia, did not shift across three sessions with games, but then did shift towards a probabilistic interpretation with new content when analyzing what happens with seeds.

Scaffolding leads to more balanced reasoning

Students demonstrated the most balanced reasoning (with a tendency towards probabilistic causation) in the final session when they were prompted to reason analogically across tasks

and to make connections to their own experiences. Twelve of the sixteen students participated in a scaffolded session and seven of those students demonstrated a tendency towards probabilistic causal statements, four were balanced, and one was still slightly more deterministic. (See Fig. 11.) It is possible that reasoning across a set of cases, all of which involved realizing some level of unreliability between cause and effect, encouraged students to acknowledge their lack of certainty about outcomes. It is expected that reasoning across multiple contrasting cases (Harrison and De Jong 2005) with divergent surface features and similar deep structure facilitated students' ability to abstract the deep structure (Goldstone and Sakamoto 2003). In this session, the divergent surfaces features would have highlighted the deep structure of probabilistic causation.

It is important to note that many of these students had offered a higher percentage of probabilistic causal responses to other tasks as well, but some students who tended towards deterministic interpretations were able to offer probabilistic responses to the scaffolded tasks. Rajon who was highly deterministic throughout the sessions with the exception of the seeds task and the gumball task, offered 54 % (D) and 46 % (P) and 45 % (D) and 55 % (P) on connections and analogies respectively. Even Sai whose responses were all 75 % and above for deterministic explanations drops down to 63 % (D) 37 % (P) and 65 % (D) and 35 % (P) on connections and analogies, respectively.

When asked to make connections to their own experiences, even the younger students offered examples from their own experiences related to probabilistic causality. Second grader Kendra said that her shower worked like the games. She said "...because sometimes when you turn the thing nothing comes out. And I'm like, mom, the shower stopped working! And she just tells me to get in the shower, so I go in the shower and all of a sudden water starts popping out and it's cold and sometimes it will be piping hot."

We are not claiming that the analogical scaffolding alone accounts for the shift in students' reasoning but that the shift resulted from the sum of the students' experiences in the sessions facilitated by the scaffolding. Our goal was to illuminate the learnability of the concepts. From an instructional stance, further investigation into the most effective instructional techniques is warranted.

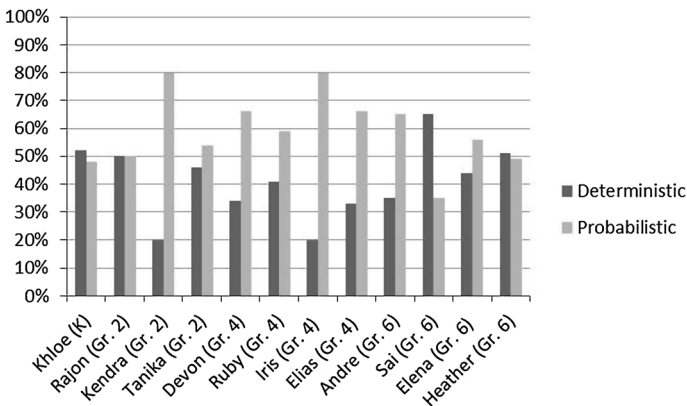


Fig. 11 Percentage of deterministic versus probabilistic responses on the scaffolded analogies by student

Modes of engagement that supported learning about probabilistic causation

Beyond the instructional scaffolds, there were particular modes of engagement on behalf of the students that appeared to serve the learning process. (1) The careful tracking of patterns, outcomes, and evidence, often the result of a persistently deterministic stance, often fueled deep exploration of the games and the drive to explain difficulties of prediction. (2) A focus on mechanism also supported students' realization that uncertainty and probability were inherent features of the tasks.

Tracking patterns, outcomes, and evidence

A number of students who initially held a strongly deterministic stance, kept very careful track of the outcomes and discussed patterns in the evidence. This approach seemed to support the eventual realization of the probabilistic nature of the outcomes. Fourth grader Elias clearly believed that the relationship between the cause (clicking the alarm) and effect (the Bear waking) in the Bear game was deterministic; he kept track of the evidence, specifically in instances when his predictions were incorrect. This seemed to feed into some subjective uncertainty. He was convinced that there was a certain factor (e.g., the number of presses) that determined when the bear woke up, but he also acknowledged that he, himself, had not discovered this factor. In earlier sessions, he was aware of his errors in prediction as he reasoned about Funny Bunny and kept these in mind as he thought about the task. An interesting shift occurred in the latter part of the Bear Game session, when the interviewer asked him to summarize his understanding of the Sleeping Bear game and to compare the causal nature of this game with Funny Bunny. He took a marked probabilistic stance for the first time in discussing the games. He explained that you "never know" when the cause would lead to the effect in these games. Alternatively, what seems to be a probabilistic shift could actually be his sharing of a subjective uncertainty stance about the games, while still believing that there is an 'ultimate' pattern to the games that is difficult to grasp. The data does not clearly disambiguate these two possibilities. We reconsider Elias's stance in the discussion because this combination of conceiving of an "ultimate pattern," even if it is hard to grasp, while pursuing evidence seems particularly powerful from a learning stance.

Sixth grader Andre focused heavily on patterns when playing the Uno Attack game. He remarked, "I'm going with the pattern, plus 2, divided... times 2...that was 6...and opposite of division is probably multiplication... I know it's not random." In a different game, he says:

A: Gotta pay attention to that button. I'm paying attention to that button. [Pause] It's like the same from the bear.

I: What do you mean that it's the same as the bear?

A: Probably maybe the same as the bear. It's a button... [He points to the button. S pushes button]...and you press it and cards spit out, like if you press it a certain amount of times, like when the card says press something times, 16 I think...[mimes pressing button in time with counting] 1, 2, 3, 4, 5 [raises hands and mimes an explosion/cards flying out.]

Andre's comments suggest that he recognizes that there could be different contingency patterns and that he expects them to be replicating and predictable. This is a reasonable assumption to make without information that would suggest that the mechanism behaves randomly. Andre also refers to "finding evidence in science" and that he had learned that if

you look for the pattern, you'll find evidence. A conversation with the classroom teacher revealed that they had talked about looking for evidence in science as part of what scientists do, had learned about the scientific method, and had learned about how scientists do tests and expect to get the same results each time. Andre seemed to have connected the tasks he was analyzing to what he was learning in science about the nature of evidence.

A focus on mechanism

Some students sought underlying causes for the patterns and discussed possible mechanisms. This led to learning about how the relationships could be probabilistic. In their first exploration of *Funny Bunny*, Andre and Sai, two-sixth graders, consider information that may illuminate aspects of the mechanism in trying to figure out how the game works:

A: [Points to four different places on the board.] Right here...right here...right here, and right here. Well since all of these are...[pokes at these spaces] I noticed that all of these yellow spaces—the plastic cover is slightly lower than the rest....

S: Yes, those are loose.

A: It's going to click and then the bunny will fall. [Flips card, twists carrot once, and his blue bunny fell through the board.] I knew it.

S: Ha ha!

Fourth grader Elias proposes a possible mechanism, “maybe that there’s like a sheet of paper at the bottom... And that when you turn it, it has holes in it, in different spots, so that it, it moves around to different places.” Even though these games had mechanisms that were not fully accessible and involved enough complexity that the data available to students for predicting the outcomes appeared highly stochastic, mechanism knowledge can be a powerful way to assess a system’s behavior and to form expectations about its reliability.

Discussion

The findings here confirm that most of the students held a generally deterministic stance despite their ages. However, shifts towards more probabilistic causal reasoning were seen across the sessions and there was variation across the domains. The fact that there were students at each grade level who discerned and reasoned about the probabilistic causal aspects from the start strongly suggests that probabilistic causal reasoning is not beyond their developmental reach even as kindergartners. The ability of students to respond to the parameters of the problems in different domains suggests that they are sensitive to features that invite probabilistic causal framing. This research is consistent with the Causal Bayesian findings that children, even as preschoolers, can sum across instances of covariation as an implicit means of causal induction. It is also consistent with the findings of Schulz and Somerville (2006) and Kalish (1998) in offering evidence that most children hold assumptions of determinism when reasoning about particular causal instances. It extends this earlier research in finding that most of the students shifted towards more probabilistic responses across the repeated sessions and across domains. This suggests the learnability of probabilistic causal reasoning. Repeated opportunities to engage with probabilistic causal tasks and the opportunity to reason analogically across cases enabled some students to realize the probabilistic causal schema. Further, the finding that

some students reasoned probabilistically at the outset may suggest that certain kinds of experiences and perspectives may help students learn to perceive when instances of probabilistic causation are relevant to a phenomenon. These outliers offer a sense of the promise that instruction might realize. It is important to note that the instructional aim should be one of helping students map the appropriate interpretation to the appropriate problem features as these students did. It is not the intent to argue that probabilistic causal reasoning is preferential in a vacuum, rather that it is important to perceive when each type of pattern is relevant.

What do these findings suggest for instruction aimed at helping students to learn about the role of probabilistic causality in complex systems? Classroom instruction in science often underscores a deterministic search for reliability and evidence. This is a necessary and important part of learning about the nature of science. However, this goal will need to be achieved in balance with the Performance Expectation in the NGSS that “science seeks patterns and evidence in terms of process and yet evidence may appear probabilistic at times” if we want students to understand the nature of causality in complex systems. Students will need opportunities to learn that causes may behave probabilistically and that the features of complex systems often make it challenging to discern underlying causal patterns, particularly when they are non-deterministic. When we are not aware that a possible relationship exists, statistical relationships are easy enough to miss given their lack of reliable outcomes. Holding an assumption of determinism means that if we notice them, we might also discount them. Teachers need to be aware of the implicit causal assumptions that students might bring to their learning. Sensitivity to probabilistic causal schemas in complex systems is particularly at issue because probabilistic causal patterns characterize many of the most recalcitrant, imminent, and systemic problems of our time, such as climate change, ecosystems decline, and global disease transmission. Being alert to the possibility that a covariation relationship is possible even when it is not reliably discernible is critical to understanding causality in complex systems.

The work here also suggests that students would benefit from opportunities to grapple with uncertainty and probabilistic causality in reasoning about complexity. An initially deterministic stance might support that learning if there is sufficient feedback for the learner to discern the lack of one to one correspondence between causes and effects. Across the grades and domains studied, some students took the stance that there is a pattern and that they can find it despite the seemingly random nature of the outcomes. This assumption compelled investigation and inquiry on behalf of the students as exemplified in the discussion of Elias above. The assumption of discernible causes and regular, reliable patterns is central to scientific inquiry and to the sense that we can reach a deeper understanding of our world. In cases where we expect that a relationship exists, this stance may compel deeper exploration. This dynamic bears further investigation.

It is not surprising that analyzing aspects of the possible mechanisms helped students to advance in their understanding of the causal dynamics. A reliance on mechanism has been seen in other instances when students found that they could not rely on co-variation data (Grotzer and Solis, 2015; Sandoval and Cam 2011). Schlottman (1999) has argued that mechanism knowledge is what enables us to distinguish between correlation and causation and Subbotsky (2001) has argued that it is what allows one to discount magic. This suggests a focus on mechanism as a means to advance students' understanding of complex probabilistic systems. For instance, if we know how carbon behaves in the atmosphere, we can make predictions about the outcomes of particular behaviors related to the production of carbon even if we fail to detect the covariation relationships between the behaviors and

the outcomes. Understanding the behavior and roles of mechanisms certainly fits with how scientists seek to understand complex phenomena.

The study here has a number of limitations that impact the interpretation of the findings. It focused on a small number of students and should be replicated with larger sample sizes. In an attempt to demonstrate the feasibility of students' reasoning about probabilistic causation, it confounded task domain with the number of sessions. A larger study should counterbalance the order of the exposure to tasks to disambiguate the influence of each variable. It also looked at probabilistic tasks that were not embedded in larger complex systems. Future studies should consider how students reason about such embedded probabilistic causal tasks and the best ways and points in development to help students learn about probabilistic causality. Some of the tasks, such as the games, offered immediate feedback about the students' predictions, such as the games. This was a useful feature in enabling us to study students' reasoning. However, it departs considerably from the information typically available to reasoners when they are grappling with complex systems.

However, despite these limitations, the results suggest the promise that we can help students to attain a stance that is open to both deterministic and probabilistic causation. Both the shifts in students who were highly deterministic at the outset and the students who reasoned in a balanced manner throughout offer evidence for the learnability of the concept. Developing the instructional capacity to teach students the features of probabilistic causality is an important goal in helping students to reason about this aspect of complex systems.

Acknowledgments The authors wish to thank Leslie Duhaylongsod, Molly Levitt, Erika Spangler, Adi Flesher, and Samantha Marengell for their assistance collecting and analyzing these data. We also wish to thank Suzannah Bukenya, Maureen Danahy, Elise Morgan, Christine Fetter, Lauren Mueller, and Mark Breneman for enabling us to work with their classes. We also appreciate the helpful editorial suggestions of Cindy Hmelo-Silver, Susan Yoon, and two anonymous reviewers. This work is supported by National Science Foundation, Grant No. NSF#0845632 to Tina Grotzer. All opinions, findings, conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation. Portions of this paper and research were presented earlier at the American Educational Research Association Conference.

References

- Achieve. (2013). *Next generation science standards*. Washington DC: Achieve.
- Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–227). Cambridge, MA: MIT.
- Atran, S. (1995). Causal constraints on categories and categorical constraints on biological reasoning across cultures. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 205–233). Oxford: Clarendon Press.
- Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, 95(4), 639–669. doi:10.1002/sc.20449.
- Bullock, M. (1985). Causal reasoning and developmental change over the preschool years. *Human Development*, 28, 169–191.
- 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.
- Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for a Bayesian cognitive science*. New York: Oxford University Press.
- Clement, C. A., & Yanowitz, K. L. (2003). Using an analogy to model causal mechanisms in a complex text. *Instructional Science*, 31(195–225), 2003.
- Eerola, M. (1994). *Probabilistic causality in longitudinal studies*. New York: Springer-Verlag.

- Einhorn, H., & Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin*, *99*(1), 3–19.
- Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*(1), 3–32.
- Gopnik, A., & Schulz, L. (2007). *Causal learning: Psychology, philosophy, and computation*. New York: Oxford University Press.
- Grotzer, T. A. (2012). *Learning causality in a complex world: Understandings of consequence*. Lanham, MD: Rowman & Littlefield.
- Grotzer, T. A., & Solis, S. L. (2015). Action at an attentional distance: A study of children's reasoning about causes and effects involving spatial and attentional discontinuity. *Journal for Research in Science Teaching*, *52*(7), 1003–1030.
- Grotzer, T. A., & Tutwiler, M. S. (2014). Simplifying causal complexity: How interactions between modes of causal induction and information availability lead to heuristic driven reasoning. *Mind, Brain, and Education*, *8*(3), 97–114.
- Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Harvard University Press.
- Harrison, A. G., & De Jong, O. (2005). Exploring the use of multiple analogical models when teaching and learning chemical equilibrium. *Journal of Research in Science Teaching*, *42*(10), 1135–1159.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss and Giroux.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Kalish, C. W. (1998). Young children's predictions of illness: Failure to recognize probabilistic causation. *Developmental Psychology*, *34*(5), 1046–1058.
- Keil, F. C. (1994). The birth and nurturance of concepts by domains. In L. A. Hirschfield & S. A. Gelman (Eds.), *Domain specificity in cognition and culture* (pp. 234–254). New York: Cambridge University Press.
- Kirch, S. A., & Siry, C. A. (2010). Maybe the Algae was from the Filter": Maybe and Similar Modifiers as Mediation Tools and Indicators of Uncertainty and Possibility in Children's Science Talk. *Research in Science Education*, doi:10.1007/s11165-010-9197-y.
- Konold, C., Madden, S., Pollatsek, A., Pfannkuck, M., Wild, C., Ziedins, I., et al. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, *13*, 68–86.
- Kurtz, K. J., Miao, C. H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences*, *10*(4), 417–446.
- 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*(1), 186–196.
- Kuzmak, S. D., & Gelman, R. (1986). Young children's understanding of random phenomena. *Child Development*, *57*, 559–566.
- Metz, K. E. (1998). Emergent understanding and attribution of randomness: comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction*, *16*(3), 285–365.
- Resnick, M. (1996). Beyond the centralized mindset. *Journal of the Learning Sciences*, *5*(1), 1–22.
- Sandoval, W., & Cam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education*, *95*, 383–406.
- Schlottmann, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology*, *35*(5), 303–317.
- Schulz, L., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*(2), 427–442.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). Reston, VA: National Council of Teachers of Mathematics.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*(1, 194), 1–51.
- Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, *46*, 394–399.

- Siegler, R. S. (1976). The effects of simple necessity and sufficiency relationships on children's causal inferences. *Child Development*, *47*, 1058–1063.
- Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, *46*, 606–620.
- Siegler, R., & Liebert, R. (1974). Effects of contiguity, regularity, and age on children's causal inferences. *Developmental Psychology*, *10*(4), 574–579.
- Sobel, D. M., & Buchanan, D. W. (2009). Bridging the gap: Causality-at-a-distance in children's categorization and inferences about internal properties. *Cognitive Development*, *24*, 274–283.
- Subbotsky, E. (2001). Causal explanations of events by children and adults: Can alternative modes coexist in one mind?. *British Journal of Developmental Psychology*, *19*, 23–46.
- Tutwiler, M. S., & Grotzer, T. A. (2014). *Comparing paths: Bayesian sequence analyses in a microgenetic study of student learning*. Pittsburgh, PA: National Association for Research in Science Teaching (NARST).
- Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling. Evanston, IL: Northwestern University <http://ccl.northwestern.edu/netlogo/>.