Interpreting Probabilistic Causal Outcomes in Science: A Microgenetic Study of Sixth Graders' Patterns of Reasoning

Tina Grotzer, Shane Tutwiler, S. Lynneth Solis, and Leslie Duhaylongsod
Harvard Graduate School of Education

Presented at the National Association of Research in
Science Teaching (NARST) Conference, Orlando, FL, April 4, 2011.

Interpreting Probabilistic Causal Outcomes in Science: A Microgenetic Study of Sixth Graders' Patterns of Reasoning

Tina Grotzer, Shane Tutwiler, S. Lynneth Solis, and Leslie Duhaylongsod
Harvard Graduate School of Education

ABSTRACT

How students deal with probabilistic causality and interpret outcomes that do not have one to one correspondence with their effects is critical to how they learn science and to how they figure out the world around them. Microgenetic studies were carried out over five sessions with four sixth graders to assess their assumptions when dealing with stochastic tasks. Sessions began with open-ended questions to reveal how students structured their expectations and explanations. Attempts to scaffold students' understanding were made reasoning analogically using tasks from four domains: social; games; machines; and biology. Sessions were intensively analyzed to identify whether and when shifts took place and what elicited them. The results are interpreted through the lens of Siegler's "Overlapping Waves Theory" (1996) along five dimensions: path, rate, breadth, source, and variability. All except one of the students responded to the tasks with a deterministic stance, even in the case of biological and social tasks. There was some movement across the sessions towards a reflective stance on probabilistic causation. The results have implications for how students view patterns of evidence in science and whether it supports a given explanatory model as well as for how teachers discuss patterns of evidence in science.

INTRODUCTION

Many causal patterns in science do not give the appearance of being deterministic. The same action or event, for example planting seeds may lead to the growth of plants or not. Further, the patterns of contingency can vary significantly. Sometimes these patterns fall within the boundaries of how we attend to the problem space, so we notice what the contingencies are. However, just as often, they fall beyond our attentional boundaries. Seeds take a long time to grow thus introducing time delays; the intricacies of a social interaction increase in cognitive load until it is difficult to reason well about them; plausible mechanisms become increasingly non-obvious such that detecting them invites a long reductionist investigation.

Probabilistic causation has long been of interest to researchers, in part, for what it can tell us about how we determine the existence of a causal relationship (e.g. Einhorn & Hogarth, 1986; Kahneman, Slovic, & Tversky, 1982) and how sum across multiple occurrences of an event and assess how other events co-vary to suggest the possibility of a causal connection. Extensive research suggests that children use co-variation data in combination with spatial and temporal contiguity (Borton, 1979; Leslie, 1982, 1984; Leslie & Keeble, 1987; Oakes, 1993; Spelke, Phillips & Woodward, 1995; Van de Walle & Spelke, 1993) and information about plausible mechanisms in assessing causality (e.g. Bullock, 1979). Earlier research suggested that children expect reliable cause-effect relationships (e.g. Bullock, 1985; Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982) and that they use consistent covariation to determine whether or not a causal relationship exists. The tendency appeared age-related with the youngest children accepting less than perfect correlation, presumably due to the cognitive load of tracking perfect correlation (Shultz & Mendelson, 1975; Siegler, 1976; Siegler & Liebert, 1974). Bullock and colleagues (1982) argued that determinism is a fundamental, innate causal principle.

Recent research in cognitive development challenges these findings, suggesting that even young children follow Bayesian rules in summing across experiences in their causal reasoning (Gopnik et al., 2004) and ignoring a lack of one to one correspondence. Gopnik and colleagues (2004) argue that young children override imperfect correlation and are able to use different patterns of probability in contiguity to make accurate causal inferences (Kushnir & Gopnik, 2007).

However, this contradicts findings in science education. Fourth graders changed their causal model for lightning when they realized that lightning does not always “strike” in a high place (Grotzer, 2003). Kalish (1998) found that 4- and 5-year-olds expected deterministic cause-effect relationships—if everyone in a classroom played with a sick child, all or none would get sick. One recent developmental study argued that children do hold a deterministic stance and engage in a search for non-obvious causes (Schulz & Sommerville, 2006). However this research poses tasks with constrained causes and outcomes and the directions may have invited children to believe that a deterministic outcome was required on the tasks. Further, little research has investigated how children develop an explicit and reflective awareness of the nature of probabilistic causation and its impact on what we know and think. This interacts with how they reason about evidence in science.

This study investigated how sixth graders reasoned about tasks with probabilistic causal features in open-ended contexts and given scaffolds to make the nature of the causal patterns explicit. It examined the assumptions that they brought to the problems and whether they developed a reflective awareness of probabilistic causal patterns over the course of the sessions. These findings are important to how they analyze evidence in science and whether they notice certain outcomes or not.

DESIGN OR PROCEDURE

Microgenetic studies were carried out with four sixth grade students. Students from with a public school in Cambridge participated in the study. The school has a diverse ethnic population with a predominance of Latino and Black students and mixed SES. Across the school year, students were interviewed in depth (n = 4) on tasks from science and beyond with varying levels of stochastic behavior. Six sessions were conducted with each student. In the initial sessions, unless the student generated examples from other contexts, only one context was focused on to avoid 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 to first assess how students frame the concepts and to then assess the accessibility of concepts related to the stochastic behavior with targeted questions. Scaffolds that made use of familiar examples and compared analogous causal forms in different problem contexts through “mutual alignment” (e.g. Kurtz, Miao, & Gentner, 2001) were incorporated in the form of design studies (Brown, 1992; Collins, 1999). Sessions were videotaped for later coding and analysis.

Task Design

Tasks: The tasks from four domains: games, biological, mechanical, and social. The tasks included seed planting, hatching chicks, bubble gum machines, videotapes of brief social interactions, and a set of games. Tasks that children might be familiar with from their everyday worlds were intentionally chosen to elicit their expectations and existing knowledge. This departs from other studies where the intention is to create causes that children would not have prior expectations about that they might bring to bear on the context (e.g. Schulz & Somerville, 2006; Sobel & Buchanan, 2009). The tasks did not all fall neatly into one category or another. For instance, some of the games had mechanical aspects in addition to their game features. However, the rationale for assessing understanding across domains was to find tasks that could potentially invite probabilistic causal reasoning, not to pit domains against one another to discern fine distinctions in reasoning in different domains. Further, we chose mechanical tasks that would potentially invite concepts of probabilistic causation (i.e. a bubble gum machine that did not regularly deliver candy) instead of a machine that was characterized by highly reliable causal processes.

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, we would be able to adjust the number of seedlings “that grew” without impacting the authenticity); 3) manageability in the classroom; and 4) minimal competition from other forms of causal complexity or competing knowledge (for instance, in Gopnik and Schulz (2004) preschoolers are given a “biology task” wherein a stuffed monkey sneezes or not when he sticks his nose on a flower to determine whether or not he is allergic to each flower. This pits authentic knowledge that children might hold about sneezing due to allergies which typically occurs after a delay against the task goal of assessing acceptance of probabilistic causation.) Tasks with complexifying causal features were included only if these features were an inherent part of the perceived probabilistic causality such as in the non-obvious causes involved in knowing what happens to seeds under the ground. Some tasks were modified for cultural sensitivity. (For instance, a Hasbro game, “Don’t Wake Daddy,” was modified to “Don’t Wake the Sleeping Bear” because most children were not Caucasian and were from single parent homes.) The tasks are detailed below:

<u>Games: Funny Bunny and “Last Bunny Standing”</u>: Funny Bunny is a commercial game created by Ravensburger for 4-8 year olds. None of the students reported seeing the game prior to the experimental sessions. The goal of the game is to be the first one to move your rabbit along a path with two loops up the hill to the top of the big carrot. Cards tell how many steps to move. However, some cards direct players to click the carrot in the middle of the game board. When the carrot is turned, a hole opens up somewhere along the path most of the time and one’s rabbit can 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 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. The cognitive load of figuring out 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.

“Last Bunny Standing” is a version of Funny Bunny where the child 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 of the cognitive load involved in game strategy (how many rabbits to have 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. If one turns the game over, the mechanism becomes visible and offers information that could be used to deduce a pattern. However, 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. Subjects were not stopped from turning the game over to look if they sought to look. Interestingly, while students did pick the game up to look for their rabbits, the only group of students to examine the mechanism were students in an earlier pilot study during task design.

Fig 1. Funny Bunny Game

Fig. 2. Don’t Wake the Sleeping Bear

Games: Don't Wake the Sleeping Bear: This game was modified from a Hasbro game entitled, "Don't Wake Daddy." The game involves getting to the finish line without waking up the sleeping dad. However, when spaces are landed upon, the player must push the button on an alarm clock a given number of times and if the dad pops up, must return to start. A realistic looking stuffed teddy bear was unstuffed and placed over the sleeping dad mechanism because many of the subjects were from single parent homes and from racially diverse schools. The number of alarm clock pushes that caused the bear to pop up ranged from 6 to 20 and each of the three games that were used had a different pattern of when the bear would pop up. However, if the students were not tracking how many pushes others had entered, it could pop up on the first push (presuming five pushes occurred during other turns). There is no visible mechanism to account for what happens.

Games: Uno Attack: This game is a variation of the game, Uno, where the player attempts to be the first to get rid of all of his or her cards. However, it has 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.

Biology: Seed Planting: Inducing causal patterns in many biology concepts presents the challenge of prediction and finding out over delayed periods of time. From a research design stance, there is the additional complication of getting kids to notice the relationship between the number of seeds and how many actually sprout. Open-ended interview questions were conducted first. It became clear that students didn't notice whether all or some of their seeds sprouted. Also, in authentic biology problems, the probability of particular outcomes cannot be calculated a priori as they can when working with games. Therefore, the following task was designed to help them remember their predictions and to help us develop a priori the outcomes (acknowledging that these manipulations area departure from authenticity.) The task engaged students in prediction and we took the seeds away each week to allow us to manipulate the outcomes. Students were told that they needed to have a certain number of bean plants (to give to specific people) in a few weeks. They were then given a peat pot, soil, and seeds and invited to plant the number of seeds that they thought they should plant in order to end up with necessary plants. They engaged in the task two to three times.

Biology: Hatching Eggs: Students were asked to predict what the inside of an incubator might look like in 22 days after eggs were set inside it. They were told that eggs typically hatch in 21 days. They were given a drawing showing eight eggs and were given an opportunity to draw the outcome later. Afterwards, they were probed on what causes the eggs to hatch, their experiences hatching eggs, and whether they had ever seen an outcome where less than the number of eggs hatched. In one of the second grades, the teacher hatched eggs so the study took advantage of this opportunity to interview students about their expectations before and afterwards.

Social Videos: Subjects were shown two brief video clips. These included one in which a girl is calling her mom for help with her homework. The rate of calling to response varies in the following way: 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. Students were 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 calling to response varies in the following way: 1) boy takes marker, sister responds, 2) boy takes a marker, takes another marker, takes another marker, sister responds; 3) boy takes a marker, takes another marker, takes another marker, and another, and another marker, then the sister responds.

Social: Cheating and Getting Caught: Students were asked "If someone cheats on a test or homework, do they always get caught?" "Can you think of ways that cheating and getting caught are like the way that the game worked?

Mechanical: Candy Dispenser: Subjects were shown a candy dispenser that you could put coins in and turn to dispense M&Ms. They were given coins and invited to make it work. The dispenser dispensed between zero and five candies with each turn with a mode of five. The actual mechanism for dispensing candies was not visible given the number of candies in the dispenser. Subjects could detect some information about the mechanism, however, because the handle was less easy to turn on some turns when it would dispense no candies and on others turning very slowly appeared to yield higher returns.

The initial sessions focused on how children reason about games. Games were chosen for the initial domain because they offer many repetitive opportunities to make predictions. Repeatedly in the course of one session, children can predict what will happen and why. Having numerous opportunities in one session cuts down on issues of cognitive load in recalling and reasoning about what happened across sessions. Games also allow for demonstrating competence on many levels. Even if the students are not able to tell us what they predict will happen or, at a more advanced level, offer explanations for why, however, implicit strategies through shifts in the playing choice in the games can reveal changes in thinking. The games also offered a platform for the students to become acquainted with and comfortable with the researchers.

Following each session, we made adjustments to the game apparatus, rules, or goals structures in the game in an attempt to eliminate certain types of cognitive load and to test certain ideas that the students may hold. For instance, to eliminate some of the cognitive load and to focus on one aspect of the probabilistic causation in playing the game "Funny Bunny," we had students focus on a version called "Last Bunny Standing" so that their goal was merely to strategize about having the last rabbit on the board, not trying to get to the end of the board. We have also manipulated the probability of certain events in the game by changing the number of rabbits and the order of the cards to see how students handled the changes.

Each task was analyzed in terms of the way that it presented as a probabilistic causal problem space. Some tasks appear probabilistic because our perception is imperfect or limited and we are unlikely to take in all the information. For instance, a game that requires students to hold information about the nature of the causes in their heads might appear probabilistic even if from the stance of a perfectly unlimited reasoner (such as a powerful computer), it is deterministic. This would also be the case with information that is gained over time. For instance, the game of Funny Bunny holds three linked causes that lead deterministically to an outcome. However, but unless one can figure out how the patterns work together, the game gives the appearance of a probabilistic causal mechanism. First of all, the large carrot in the middle of the board turns and when it is clicked (in response to a game card). It moves (unseen) a mechanism under the game that has holes in particular places. Secondly, there are only some spots on the board where the holes can open up. These spots are "soft' or wiggle when you push on them. Thirdly, the mechanism has two plates that move and this causes the opening to shift from the bottom path towards the carrot to the top path towards the carrot. Finally, to further the appearance of probabilistic causation, there is a place on the turning mechanism below where no hole will open up when the carrot is turned making the relationship between turning the carrot and producing a hole probabilistic. So what causes a rabbit to fall? The carrot has to be turned. The rabbit has to be on a soft spot. A spot has to open. And the rabbit has to be on the particular spot that opens. The interaction between these factors makes it appear stochastic.

Other tasks might be viewed as deterministic if you could follow a long reductive trail and have answers about all the variables along the way. For instance, when hatching eggs, even if you control for temperature, moisture, type of care, and the physical features of the eggs, the eggs will typically not all hatch. If you could know detailed specifics of each egg, perhaps you could know why it didn't hatch, but this information is unavailable for most observers. This is also the case with mechanisms that appear to function randomly. If you had detailed information about tension on the spring, you might be able to tell what turn will cause it to snap, but the information is not available. It involves thinking about what the knower knows versus what is knowable (and with how much effort and with what outcomes.)

Design Study Aspects: Leveraging Students' Understanding:

One of the aims was to look for affordances in students' thinking and to use these affordances in service of learning about probabilistic causation in other topics. However, for three of the students, their reasoning did not differ much between domains. Therefore, we engaged the students in comparing analogous causal forms in different problem contexts through "mutual alignment" (e.g. Kurtz, Miao, & Gentner, 2001). Mutual alignment involves mapping back and forth between analogical problems making attempts to discern similarities and differences and to use each to further inform understanding of both.

SCORING AND ANALYSIS

Sessions were intensively analyzed to identify when shifts appeared to be taking place and why. The data sources include audiotapes and transcripts from each session; the series of students' predictions for each task (including mathematical analyses of moves that support probabilistic or deterministic assumptions) and students' reflections upon previous sessions in later ones. Our analysis included using ATLAS.TI to consider 1) transcripts and videos of students responses using etic, top-down questions to discern what students perceive about the nature of the causality; and 2) Emergent analysis of patterns in students' reasoning where we identified themes in the students' reasoning. Independent coders coded the interpretive aspects and their levels of agreement were assessed.

The coding scheme can be found in Appendix A. It involved developing clear distinctions between what deterministic and probabilistic statements sounded like. Transcripts were independently coded by two researchers for probabilistic or deterministic statements. Only statements that both coders independently detected were included the subsequent analysis. The detection rate between coders was calibrated over two rounds of scoring until the range of detection was 89.98% across transcripts (ranging in misses from 1-7 with a mean of 3 and mode of 2) After two rounds of coding refinements, they agreed on detected statements (91.46%) with a high level of agreement on type of statement (97.96%). Emergent codes independently generated by two of more coders are reported here. Narratives were developed of how the students changed in their explanations of the task, accompanied by detailed task analyses and consideration of leverage points that may be useful in teaching concepts that embed the causal concepts.

The analysis was conducted using the five criteria set forth by Siegler in his 1996 "overlapping waves theory" for microgenetic analysis to reflect how children's understanding changes along five dimensions: path, rate, breadth, source, and variability. The narratives consider how children's understanding develops along five dimensions: path, rate, breadth, source, and variability. Path describes how children sequenced their behaviors to get to the change. Rate describes how quickly and with what supports the child moved from the realization of the new concept to the consistent application of that concept. What kinds of time and experience were needed to make the shift? Was it abrupt, gradual, and what kinds of investments may have led to that shift? The breadth of the change refers to how narrowly or broadly the child gained the concept. Did it apply just to the particular game, to other games with the same inherent features, or to other contexts? The source addresses what the child did to make the change. What changes in behavior appear to have led to the change? What realizations seemed to have helped the child to see the new feature? The variability refers to the difference between students on the dimensions above. Are there similarities in the patterns of change that would suggest instructional approaches beyond this set of students? Are there idiosyncratic patterns that are important for teachers to be aware of?

RESULTS

The following patterns emerged from and were supported by the subsequent data analysis. We are continuing to mine the data for further patterns and to help us to further understand those that we have found.

Most students were initially quite deterministic across domains. Three of the four sixth graders approached the tasks from a deterministic stance across the domains tested (expected for mechanical devices, but not for biology and/or social tasks.) Even prior experiences such as planting seeds were interpreted deterministically and explained in a reductionist manner. For example, statements for Andre were all in the 90% range. The two male students were especially inclined to treat the tasks as puzzles to be solved in terms of the patterns in how they worked.

Even in the context of biology, three of the students approached the tasks in a deterministic manner. When asked if they had experience with seeds, Etu and Andre both gave accounts of planting seeds with the program “City Sprouts.” However in each case, they talked about planting the seeds but not going back to see them later. When predicting that all eight eggs in the incubator would hatch, Etu 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 and how cute they were, but then the authorities took them when he landed in Boston. The whole story was surprising to the interviewer, except, perhaps the ending.

When seeking patterns, Andre referred to “finding evidence in science” and 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 and how they find out what happened.

In unsupported contexts, there were small shifts in language towards recognizing the possibility of probabilistic causality, particularly when contrasting across examples. However, the shift appeared to pull against ideas of how one seeks evidence in science. 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.” However, they emphasized the role of evidence in being able to determine causal outcomes and that if you persisted in following the evidence, you would get there.

Some students who held a deterministic stance at the outset began to make a shift towards 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.” When they were unable to determine strategies for how to predict, some students redefined the cause (from a mechanistic to an anthropomorphizing one, for instance.) 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.

One of the students approached the tasks in a more open manner—allowing for the possibility of probabilistic or deterministic responses. Her responses were significantly more balanced across response types and shifted with predictions in particular domains. Elena presented with a different stance than her classmates. She allowed for probabilistic outcomes and was varied in her responses across domains. She introduces a different growth pattern and perhaps, points of leverage that can be used in service of instruction.

Contrast the difference between the percentages of statements made by Elena and Andre in Table 1. Andre presents as very deterministic while Elena entertains the possibility of probabilistic statements from the outset.

Table 1. Comparing Response Patterns: Elena and Andre

		Elena		Andre	
Session	Task Domain	Deterministic	Probabilistic	Deterministic	Probabilistic
1	Mechanical/Game (Funny Bunny)	63%	36%	94%	5%
3	Mechanical/Game (Uno Attack)	60%	40%	90%	10%
5	Connections Across Domains	43%	57%	57%	43%

Elena discusses the probabilistic nature of 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. Its 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.

Elena also makes more varied statements within different domains. Notice the shift towards more probabilistic statements in social reasoning.

Table 2. Percentage of Response Types Across Domains: Elena

Session	Task Domain	Deterministic	Probabilistic
1	Mechanical/Game (Funny Bunny)	63%	36%
2	Mechanical (Bubble Gum Machine)	38%	61%
3	Mechanical/Game (Uno Attack)	60%	40%
4	Social (cheating; calling mom, pestering sister)	27%	73%
5	Connections Across Domains	43%	57%

When contrasting domains, students were more likely to offer both kinds of statements. Shifts could be seen in the balance of the types of statements that students made. For instance, Andre shifts from 90% deterministic statements to 57% deterministic and 43% probabilistic. He began to talk about how in some cases, “you couldn’t know.”

When reflecting back on his experience, however, he doesn’t seem to realize how deterministic his responses have been. He comments that when he and a classmate were looking for patterns in one of the games, "I thought Etu 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 Etu gather data and seemed to believe that they could figure out the pattern.”

Andre makes deterministic statements in Session 1 (“I’m going with the pattern - plus 2, divided… times 2…that was 6…and opposite of division is probably multiplication.”) and again in Session 2 (“I know its not random.”) Yet in Session 5 where he is asked to draw conclusions from across the activities, he makes 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, and yeah. I can’t think of anything else.” When asked about connections, he comments about volcanoes “Can’t predict. Can’t always be sure when a volcano’s going to erupt…“ Even with this subtle shift, the grasp of the concept of probabilistic causality seems local and fragile, specific to examples or scenarios. Students had a hard time transferring their newfound reasoning to a new context (whether or not scaffolding is present). They seem to “get it” in one context, but if the context changes, their conception reverts to deterministic thinking.

DISCUSSION

The findings here confirm children’s generally deterministic stance while also resulting in small shifts in their reasoning when engaged in scaffolded reflective comparisons between instances. It suggests that the students might see a deterministic search for evidence as part of the epistemic nature of science. However, the small shifts in students’ patterns when comparing across problem types suggests that it may be possible to teach students to reflectively reason about patterns that appear less than probabilistic and how they impact our thinking in science and the everyday world. The different stance adopted by Elena signals that reasoning about probabilistic phenomena is certainly not beyond a sixth graders developmental grasp.

We believe that a reflective stance is important to the broader problem of whether we detect causality with probabilistic characteristics in our world. Our reflective ability to detect stochastic event structures matters because in a complex world, it is easy to lose sight of the impact of particular choices, individual and societal, when there is no observable effect following an action. These patterns characterize many of the most recalcitrant and imminent problems of our time, such as climate change, ecosystems decline, and global disease transmission. Of particular interest is whether it is possible to shift students’ attentional boundaries to some extent. While this research focuses on only four students in a microgenetic analysis, it does illuminate some of the challenges of helping students attain such a stance as well as some promise that it might be possible. It leaves open many questions about the extent that such a stance can be encouraged and the best ways and points in development to do so. Future research with larger groups of students could help to inform answers to these questions.

ACKNOWLEDGEMENTS

The authors wish to thank Molly Levitt, Erika Spangler, Adi Flesher, Samantha Marengell, Evelyn Chen, and Reuben Posner for their assistance in collecting and analyzing this data. We also wish to thank Elise Morgan for enabling us to work with her class.

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.

REFERENCES

Borton, R.W. (1979, March). *The perception of causality in infants.* Paper presented at the meeting of the Society of Research in Child Development (SRCD), San Francisco.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences, 2* (2), 141-178.

Bullock, M. (1979). *Aspects of the young child's theory of causation*. Unpublished doctoral dissertation. University of Pennsylvania.

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.

Carey, S. & Spelke, L. (1994). Domain-specific knowledge and conceptual change. In L.A. Hirschfeld & S.A. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture* (pp 169-200). New York: Cambridge University Press.

Collins, A. (1999).The changing infrastructure of education research. In E. Lagemann & L. Shulman (Eds.) *Issues in education research* (pp 289-298). San Francisco: Jossey-Bass

Einhorn, H. & Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin, 99(1)*, 3-19.

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. (2003). Learning to understand the forms of causality implicit in scientific explanations. *Studies in Science Education, 39,* 1-74.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, MA: 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. Hirschfeld & S.A. Gelman (Eds.), *Domain specificity in cognition and culture* (pp 234-254). New York: Cambridge University Press.

Keil, F.C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A.J. Premack (Eds.), *Causal cognition: A multidisciplinary debate.* (pp 234-262). Oxford: Clarendon Press.

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.

Leslie, A.M. (1982). The perception of causality in infants. *Perception, 11*, 173-86.

Leslie, A.M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception, 13*, 287-305.

Leslie, A.M., & Keeble, S. (1987). Do sixth month old infants perceive causality? *Cognition, 25*, 265-288.

Oakes, L.M. (1993, March). *The perception of causality by 7- and 10-month-old infants.* Meeting of the Society for Research in Child Development (SRCD), New Orleans, LA.

Opfer, J. E. & Siegler, R.S. (2004). Revisiting preschoolers' *living things* concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*, 301-332.

Schulz, L & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, 40(2), 162-176.

Schulz, L. & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, 77(2), 427-442.

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., & 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.

Spelke, E.S., Phillips, A., & Woodward, A.L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A.J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp 44-78). Clarendon Press: Oxford.

Van de Walle, G., & Spelke, E.S. (1993, March). *Integrating information over time: Infant perception of partly occluded objects.* Biennial meeting of the Society for Research in Child Development (SRCD), New Orleans.

Appendix A: Codes for
Deterministic/Probabilistic Causality

Main Code: Deterministic Thinking [*Atlas.ti code name: Deterministic thinking*]

Main Code Explanation: Statements or behaviors by the child that show that he/she considers that a cause always leads to an immediate effect (i.e., one-to-one correspondence). He/she uses language like, "I'm sure, definitely, always, pattern…"

Decision rule *This code still applies even when the child does not demonstrate knowledge of what the pattern actually is but is convinced that one exists and/or is actively seeking to figure it out.

*This code can be double coded with Certainty or Non-Committal Language [see below for explanations of these codes]. For example, "I can know the outcome," would be coded as Deterministic Thinking and Certainty and "It will open up here or here," would be coded as Deterministic Thinking and Non-Committal Language.

Main Code: Probabilistic Thinking [*Atlas.ti code name: Probabilistic thinking*]

Main Code Explanation: Statements or behaviors by the child that show that he/she considers that the connection between cause and effect is uncertain (i.e., not one-to-one correspondence). Uses language like, "Might, maybe, probably, sometimes, random, lucky."

Decision rule *Probabilistic language (e.g., maybe, probably) may be used by the child as a reflexive speech habit rather than an expression of their reasoning. It is important that the Probabilistic Thinking code be used only when the child is clearly reasoning about the co-variation between cause and effect. Beware of using this code simply because these words are used.

*This code is sometimes double coded with Certainty or Non-Committal language[see below for explanations of these codes].

Main Code: Certainty [*Atlas.ti code name: Certainty*]

Main Code Explanation: Interactions in which either the experimenter asks or the child responds as to whether it is possible to know the outcome of an action in the task at hand. Examples may be, "Can you know? How sure are you?" [experimenter], "I can know…" "I know…" [child].

Decision rule *Use this code to indicate instances when an interviewer asks a child to think about and/or a child reflects upon whether it is possible to figure out how the pattern/game/activity works, even if he/she does not know how it works right now.

*This code is oftentimes double coded with Interviewer Support codes, Deterministic Thinking, and/or Probabilistic Thinking.

Main Code: Cause & Effect [*Atlas.ti code name: Cause & Effect*]

Main Code Explanation: Statements the child makes about how the task works, specifically identifying the cause and/or effect of the event, that do not contain information about whether the child was thinking deterministically or probabilistically.

Decision rule *This code is often used when the child indicates a mechanism of the task, without explaining how the cause and effect co-vary. Students often use the phrase, "Depends on…" to begin talking about cause and effect. However, when the child simply says, "It depends," without explaining on what mechanism it depends, then this may likely be a probabilistic statement.

Main Code: Contradictory Statement [*Atlas.ti code name: Contraction Statement*]

Main Code Explanation: Statements in which the child seems to switch back and forth between deterministic and probabilistic language or the child demonstrates local backward reasoning. For example, "That happens every time, sometimes" or "It might happen. [The hole opens] I told you it would happen!"

Main Code: Interviewer Support

Subcodes: Interviewer Neutral [*Int. Neutral*] Interventions/probes/questions on the part of the interviewer that do not seem to have an intention of leading the student to probabilistic or deterministic thinking, but still seem important in helping the child think about the problem space.

Leading toward Deterministic [*Atlas.ti code name: Leading Det*.] Statements or actions by the interviewer that encourage (lead) the child to reason/respond in a deterministic fashion.

Leading toward Probabilistic [*Atlas.ti code name: Leading Prob.*] Statements or actions by the interviewer that encourage (lead) the child to reason/respond in a probabilistic fashion.

Decision rule *Interviewer Support codes may be double coded with Certainty.

Main Code: Non-committal Language [*Atlas.ti code name: Non-committal*]

Main Code Explanation: Statements in which the child's reasoning and response wording mismatch. For example, the child talks deterministically (about a

	pattern or one-to-one correspondence) but uses noncommittal (probabilistic) language, “The next hole to open is *probably* going to be #12.”
Decision rule	*Non-committal language is almost always double coded with either Deterministic Thinking or Probabilistic Thinking. For example, “I think it will open here or here,” will be coded both as Deterministic Thinking and Non-Committal Language.

Main Code: Source [*Atlas.ti code name: Source*]

Main Code Explanation:	Instances when the student speaks about an outside experience or an observation in the present task that provides evidence for a probabilistic explanation.

Miscellaneous Decision Rules

- “I think”: Does not indicate one style of thinking one way or another, but it should be considered in the context of a whole phrase or interaction.

- Indication of worry/fear/hesitation about the task: An explicit or non-verbal expression of worry/fear/hesitation does not indicate one style of thinking one way or another, but it should be considered in the context of a whole phrase or interaction.

- Anthropomorphizing the game: For example, “The game doesn’t like me. It’s mad at me,” can be coded as deterministic thinking, because it indicates that the child thinks there’s a [non-obvious] reason for the outcome of the game. However, if you can’t tell whether or not the child is serious about their statement, don’t code.

- Other non-obvious causes: Whether the child talks about obvious or non-obvious causes, if they refer to a one-to-one correspondence between the cause and effect, code as “Deterministic.” For example, “This happened, because the teacher doesn’t like me” or “It knows I’m there,” would be coded as deterministic.

- Silly responses: Sometimes students get silly and start throwing out wild ideas; if the student is obviously kidding, don’t code.

- No chunking: As much as possible, code each line separately. Chunking whole sections of child-interviewer interactions makes it harder to figure out reliability.