Notes

Chapter 1

1. Dave went to graduate school at the University of California at Berkeley. The first time he went to Berkeley, he walked through Sather Gate and onto the main campus. Turning left at the library, he saw the word “Psychology” etched in marble on the side of the Valley Life Sciences building. Turning to his host—a graduate student who was familiar with the campus—he asked whether that was the psychology department, having never seen so nice a building and inspired by the majesty that he would be working in such environs. His host replied “no” with a laugh, and explained that the psychology department used to be housed there, but was moved to Tolman Hall to make room for the other biological sciences etched in the marble around the building. Tolman, a 1960s concrete building, built in a Brutalist architectural style, was something of a letdown. The folk continuum is real even (and maybe especially) inside university structure.

2. Whether those neurological developments are based on biological maturation independent of environmental input is unclear, but seems incredibly unlikely.

3. Because the examples in this section are about psychological knowledge, one might object that these examples are not well-matched to the examples above about physical relations among objects. This argument goes like this: Physics is a science, so understanding the causality of physical relations among objects is clearly “science knowledge.” Psychology is “less” of a science, thus understanding relations among mental states is not science knowledge, but merely knowledge about human behavior. Our response, of course, is that psychology is a science; there is every bit as much causal structure in understanding psychology as there is in understanding physics. The main difference is that most of physics involves deterministic relations, which can be mathematically described. Human behavior, in contrast, is stochastic. But its relations are just as causal, as many philosophers have argued (e.g., Campbell, 2007). Just because psychological relations are stochastic does not stop people from writing formulas that attempt to explain human behavior, which basically summarizes a major goal of cognitive science.

4. Frankly, we doubt that most adults know all the whys behind their behavior, but at least adults can try to articulate them.

Chapter 2

1. Gopnik et al. (1999) describe a similar version of this kind of systematic exploration as the "drop the spoon" game: Whereas younger infants (around 8 months) drop items off their high chair to learn that unsupported objects fall, older infants (around 14 months) might do it to learn how many times they can drop an item before a parent takes it away.

2. We are both still wondering how "carrying clipboards" became a defining property of scientists. That said, on the show *Odd Squad*, the child-scientist character who works in the gadget lab (Oscar) does wear a lab coat and carry a clipboard. So maybe we are missing something.

3. There are certain exceptions to this, such as Bullock et al. (1982) and Shultz (1982), who set the groundwork for the contemporary study of causal reasoning in young children.

4. There is an open question as to whether nonhuman animals have the same kinds of causal reasoning capacities as young children. Although it initially seemed as though the answer to this question was "no," over the past twenty years, work in comparative psychology has uncovered that nonhuman animals can make highly sophisticated causal inferences (e.g., Blaisdell et al., 2006; Schloegl & Fischer, 2017) and can perform similarly to human children in some tasks (e.g., Seed et al., 2012). However, the reasoning mechanisms used by human and nonhuman animals may still be different, even if they sometimes result in similar inferences (see Penn & Povinelli, 2007, for a critical review). We do not wish to take a position on this debate, but we will point out that an important difference between human children and nonhuman animals is the number of trials that members of these two groups often require to make an inference. Children usually require only one, or a very small number, whereas nonhuman animals usually require orders of magnitude more. So while there might be some shared evolutionary foundations, the kind of scientific thinking we describe in this book seems more likely to be unique to human reasoning.

5. Although these simple causal systems may be more common in laboratory experiments than in the real world, there are real-world examples of them as well. After Brown University renovated Dave's building, he discovered that his office lights worked on a touch panel that lacked sensitivity, so simply touching it turned the lights on only about half the time. To make sure that he does not work in the dark, he has to stand there and press the switch really hard, usually for about six seconds. It took a while (and at least one call to the facilities department) to determine that the strength of the button push was the key to activating the lights. In addition, the light switch is probabilistic; it works about 85% of the time. So sometimes Dave just

stands there pushing the button over and over. At this point, the most reasonable conclusion may be that it's operated by a Cartesian demon.

6. At the outset, we would like to say that there are more computationally sophisticated descriptions of this framework (e.g., Glymour, 2002; Gopnik et al., 2004; Pearl, 2000). Our goal here is to describe the framework so that readers can understand that rational constructivism can have a more precise computational formalism. Because of this, we are not presenting many computational details here, and refer the interested reader to these or other references.

7. We'd like to say that what faithfulness implies is the assumption that we are not living in the Matrix, but (a) individuals with philosophical inclinations might argue with this statement and (b) there's a fourth Matrix movie in the works, which might change what that phrase means. So we are not going to make this statement.

8. Some might argue that models like the one shown in figure 2.2 are so simplistic and general (X and Z can stand for just about anything) that the faithfulness assumption does not matter. That is, building these models as a representation does not buy the child anything because these models are not realistic models of the world. We disagree for two reasons. First, children's representations of causal structure are presumably simpler than adults' (i.e., they have less content knowledge), and we have to start somewhere to get a handle on what they might be representing. There are numerous cases in which learners are better off with a "less is more" approach (e.g., Newport, 1990). Second, faithfulness is not really a psychological principle. It's an assumption about translating the representation inherent in a model to an understanding of causality. That said, the initial critique does have merit; the models built using this framework do tend to be (perhaps unrealistically) simple. We address this point in chapter 3 when we discuss nonindependence.

9. Studying children's causal reasoning involves reading a lot of comic books and watching a lot of Marvel movies (at least for us).

10. Alison Gopnik led the team of Jennifer Esterly, Greg Robison, and Dave in creating the blicket detector. Dave and Jen went on to use the detector to study various aspects of children's causal and scientific reasoning. Greg pivoted to studying children's humor; we always suspected that Greg had the better time.

11. If you want to build your own blicket detector, follow the instructions here: https://osf.io/5qt2h/.

12. Schwartz and Reisberg (1991) provide an excellent summary of this theoretical approach and take the reader through many of the same calculations as we present here.

13. This is sometimes separated into two variables, alpha and beta, representing the stimulus and the outcome separately. Because alpha and beta are constants, these numbers can be multiplied together and represented as K.

14. McCormack et al. (2009) used different controls than our studies did—controls that were more faithful to the associative reasoning literature. These researchers found a developmental difference, whereby 4-year-olds did not engage in backward blocking, but 5-year-olds did. We revisit this finding in note 17 below.

15. Xu (2019) also describes a view of cognitive development called *rational constructivism*, which is slightly different from the description we present here. It is beyond the scope of this chapter to detail all the differences between this view and the one we describe, but we want to acknowledge that this term—and this theory—is still evolving.

16. Here, we describe those priors and how they differ in general terms. A formal description of this modeling for the interested reader can be found in Griffiths and Tenenbaum (2007, 2009).

17. There are two other points we want to make here. First, this mechanism also potentially explains the development observed in McCormack et al. (2009), mentioned in note 14. While the 4-year-olds in our studies were able to relate causal properties to insides, it is possible that the children tested in McCormack's study were not. Given this, these children would have interpreted the backward blocking demonstration differently, leading them to respond that the B object was efficacious. Second, it is not clear whether children are explicitly representing each and every one of these hypotheses. In fact, it is unlikely that they are, because the hypothesis space can be infinite or at least very large. Instead, it seems more likely to us that this is an implicit process, so we are not aware that we are engaging in this kind of representation.

18. Perspective really determines if this is a blessing or a curse.

Chapter 3

1. This, in turn, provides a foundation for thinking about other kinds of causal models, since more complex models can be broken down into three-node models of common causes, common effects, and chains.

2. Ahl and colleagues (see e.g., Ahl et al., 2020) posit a different developmental trajectory than the work from our lab. However, they also use more complex stimuli, which might have made it more difficult for children to reason in the same way as they did with our stimuli (see our discussion of contextualization in chapter 6). Nevertheless, we think that these two sets of findings are generally consistent.

3. Carey (2009) clearly disagrees with the rational constructivist framework in other ways, particularly with respect to what other kinds of knowledge might be innate (see also Spelke & Kinzler, 2007) and with respect to the argument that the causal graphical model framework serves as a good representation of knowledge. But regarding the idea that a concept of "cause" is innate, there does seem to be some agreement between these theoretical positions.

4. There is an open question here, which is beyond the scope of this chapter: Do nonhuman animals reason only by associative inference or do they have the same kinds of inferential capacities as human beings? Penn and Povinelli (2007) provide an outstanding review of this issue. They conclude (and we concur) that the answer lies somewhere in the middle. We take this conclusion as evidence that there are some discontinuities between nonhuman animals and human beings, which is one of the bases for our suggesting that causal inference is not innate and starts with associative reasoning mechanisms.

5. Not just for causal concepts, but for conceptual structure in general—it's what Keil (1989) referred to as "original sim."

6. It is not clear whether that acceleration lasts; to our knowledge, no one has given 3-month-olds the sticky mitten experience and then tested them a month later on understanding goal directedness. But it is certainly plausible that this acceleration does last (see Libertus et al., 2016; Wiesen et al., 2016).

7. There is an alternative version of this argument, namely that causality is innate, and what the motor experience does is allow babies to access their existing understanding that event regularity should be processed in terms of causal relations. Although we do not favor this interpretation, as it does not explain the transition between 5- and 8-month-olds that we demonstrated in our previous work (Sobel & Kirkham, 2006, 2007), we suggest that this is an open empirical question.

8. It is tempting to suggest that such accounts of development are more akin to dynamic systems models (e.g., Thelen & Smith, 1996). Indeed, Smith and Thelen (2003) make some suggestions for how sociocultural findings, particularly interaction, might be interpreted in this way. The algorithms that they suggest, however, seem to focus on the kinds of internal, cognitive developmental processes (such as the A-not-B error) that are similar to information-processing algorithms. It is an open question whether such models may be better accounts of the kinds of social learning processes we describe here.

9. After many sociologists and anthropologists disproved of and disavowed Levy-Bruhl's hypotheses, one prominent scholar continued to cite his work: Piaget (see Jahoda, 2000). Piaget compared early childhood to other cultures that showed (according to Levy-Bruhl) "non-logical" thought patterns.

10. These authors would not consider themselves as describing a rational constructivist framework, but there are similarities between these models and the Bayesian algorithms we described in the previous chapter. What is important to note is that we are imagining an extension of these authors' work; Werchan et al. (2016) did not apply their model to cultural differences.

11. One of the ways we tried to consider some of these ideas in Callanan, Legare, Sobel, et al. (2020) was to look at how parents viewed the relation between play and learning and their goals for visiting a children's museum (inspired by work by

Gaskins, 2008). Unfortunately, this part of the study resulted in inconclusive findings. Investigating this issue more systematically is a high priority for our future endeavors.

12. Bronfenbrenner (1979) described developmental psychology as "the science of the strange behavior of children in strange situations with strange adults for the briefest possible periods of time" (p. 513).

13. We assume that some researchers in theory of mind would say that these false belief contrastives are children's initial attempt to verbalize their "implicit" theory of mind capacities, as suggested by looking time studies (e.g., Onishi & Baillargeon, 2005). However, we cannot find a reference in this literature that directly makes this claim. Further, this assumes that such looking time studies are robust; they may not be (see e.g., Sabbagh & Paulus, 2018).

14. There is an open question here as to whether this tendency is culturally universal. Gauvain and colleagues, for example, use cross-cultural investigations to suggest that it is not (e.g., Gauvain et al., 2013, but see Callanan, Solis, et al., 2020, for a different interpretation).

Chapter 4

1. As we write this book, Dave's daughter (age 12) is in seventh grade and is actively writing a report on the control of variables strategy, using a version of the slopes task (Chen & Klahr, 1999) in her science class. This is the first time she has been introduced to the strategy in her public school education.

2. The superpencil detector is actually the heating element of a broken coffee maker, connected via a fake USB port to a computer running a MATLAB script that allows the researcher to control whether "superlead" is detected on a given trial (via a surreptitious mouse click on the part of the experimenter).

3. This study in particular looks at using simpler contexts to ask questions about children's concepts of planets, and finds that simplifying the context facilitates reasoning in elementary-school-age children. We address this issue directly in chapter 6.

4. Indeed, the week they discussed causal and scientific reasoning, Dave polled his advanced seminar on cognitive development. No student thought snake activity was a risk factor for earthquakes a priori. Although this is merely anecdotal evidence, it suggests that even adults have prior knowledge that might influence their reasoning.

5. Critically, these children were all tested before they entered formal schooling environments, so the age-related difference observed here is not the result of only the older children being in school.

6. Unlike the more contemporary studies, which contextualized their findings in terms of the causal structure of a physical system, Reiber (1969) investigated children's hypothesis testing about another's behavior (where they have hidden a candy), which is inherently more probabilistic.

7. At this point, we just want to acknowledge that “little engineers” is a bad description. Most engineers we know (including many we have worked with throughout the years) are thoughtful, scientific in their reasoning, and not purely driven by generating rewarding behavior if it means not understanding causal structures.

Chapter 5

1. In the early days of the blicket detector, this enabling switch was wired to the box. Dave tested children in a room with a one-way mirror; a wire led from the detector, down the floor, up the wall next to the one-way mirror, and into a hole in the wall, which led to the observation gallery. It seemed pretty obvious to us that a person was in that gallery, moving the switch; the wire even sometimes moved slightly during the experiment. Because we wanted to be sure that participants were convinced that the detector really worked as we described, we piloted the first blicket detector experiment (described in chapter 2) on a set of adults. None of them figured out that the machine was controlled by a person behind the mirror. Only one child—a five-year-old—figured out how the machine really worked during the year that we tested in this manner. That was in 1996, so that child is 31 upon this book’s publication. We really wonder what they are doing now.

2. For experimental control, we counterbalance the order in which participants see these four events. Half of the participants see the order described here (all four blocks, two combinations of three blocks, the single block), and the other half see the events in the opposite order (the single block, two combinations of three blocks, all four blocks). This variable did not make a difference to any of our results and will not be discussed further.

Chapter 6

1. The real quote uses “magic” instead of “fantasy.”

2. This difference is statistically significant, exact proportion test, $p = .049$.

3. Like the other two versions of this study, half of the children saw the dinosaurs introduced in this order, and the other half saw the dinosaurs introduced in the opposite order. Order never mattered to children’s performance, so we do not discuss this variable further.

Chapter 7

1. Exact proportions tests; all p-values $< .001$

2. Interpretation: $\chi^2(1) = 12.74$, $p < .001$; Preference: $\chi^2(1) = 5.09$, $p = .02$

3. $\chi^2(1) = 0.25$, $p = .62$

4. $\chi^2(1) = 0.18$, $p = .67$

5. $\chi^2(1) = 7.08$, $p = .008$

6. $\chi^2(1) = 67.53$, $p < .001$

7. Paired t-tests, Fact: $t(76) = -4.47$, $p < .001$; Interpretation: $t(77) = -3.66$, $p < .001$; Preference: $t(77) = -1.76$, $p = .08$

8. $\chi^2(1) = 4.96$, $p = .03$

9. $\chi^2(1) = 2.04$, $p = .15$

10. $\chi^2(1) = 10.45$, $p = .001$

11. Because some children received the same task twice, we changed the colors between the testing sessions. We here refer to the colors used in chapter 5, which were presented to these children in 2015; in 2017 we used yellow, brown, green, and gray blocks that made the machine light up either blue or purple.

12. Exact proportions test, $p = .07$

13. $\chi^2(1) = 0.001$, $p = .97$

14. Exact proportions test, $p = .15$

15. Exact proportions test, $p < .001$

16. Exact proportions test, $p < .001$

17. One student did not respond to the main test question on the blicket task, so results for this task involve 111 students rather than the full 112.

18. Binomial test, $p = .04$

19. Exact proportions test, $p = .89$

20. $\chi^2(1) = 9.03$, $p = .003$

21. Binomial test, $p < .001$

22. $\chi^2(1) = 3.95$, $p = .05$

23. Binomial test, $p < .001$

24. $\chi^2(1) = 2.20$, $p = .14$

25. Chi-squared tests, all χ^2 values < 1.63, all p-values $> .20$

26. $t(169) = -0.03$, $p = .98$

27. $t(115) = -2.79$, $p = .006$

28. MAP scores can range from 100 to 350, and this test is designed to measure performance from first grade to high school on the same scale.

29. As with the MAP, scores on the PSSA are designed to measure performance across multiple grades on the same scale (third grade through eighth grade; minimum score = 600).

30. First graders: reading correlation, $r(214) = 0.26$, $p < .001$; math correlation, $r(214) = 0.31$, $p < .001$

31. Third graders: reading correlation, $r(107) = 0.32$, $p < .001$; math correlation, $r(107) = 0.31$, $p = .001$

32. Mean reading score of first graders who responded correctly on the fact trial = 194.82 (SD = 14.85); mean reading score of first graders who responded incorrectly = 186.09 (SD = 13.81); $t(214) = –4.27$, $p < .001$

33. Mean math score of first graders who responded correctly on the fact trial = 195.49 (SD = 10.59); mean math score of first graders who responded incorrectly = 187.13 (SD = 13.79); $t(214) = –4.99$, $p < .001$

34. Mean reading score of first graders who responded correctly on the interpretation trial = 193.20 (SD = 15.33); mean reading score of first graders who responded incorrectly = 189.59 (SD = 14.51); $t(214) = –1.76$, $p = .08$. Mean reading score of third graders who responded correctly on the interpretation trial = 216.17 (SD = 11.00); mean reading score of third graders who responded incorrectly = 208.14 (SD = 15.17); $t(107) = –3.01$, $p = .003$

35. Mean math score of first graders who responded correctly on the interpretation trial = 193.93 (SD = 12.76); mean math score of first graders who responded incorrectly = 190.49 (SD = 11.95); $t(214) = –2.02$, $p = .04$. Mean math score of third graders who responded correctly on the interpretation trial = 223.64 (SD = 10.33); mean math score of third graders who responded incorrectly = 217.07 (SD = 14.83); $t(107) = –2.58$, $p = .01$

36. Mean reading score of third graders who responded correctly on the preference trial = 214.94 (SD = 12.03); mean reading score of third graders who responded incorrectly = 204.89 (SD = 12.03); $t(107) = –2.33$, $p = .02$

37. Language correlation, $r(105) = .24$, $p = .01$; math correlation, $r(105) = .25$, $p = .009$; science correlation, $r(104) = .30$, $p = .002$

38. Mean reading score of first graders who responded correctly on the blicket task = 193.63 (SD = 16.27); mean reading score of first graders who responded incorrectly = 192.05 (SD = 15.34); $t(153) = –0.62$, $p = .53$

39. Mean math score of first graders who responded correctly on the blicket task = 191.41 (SD = 13.27); mean math score of first graders who responded incorrectly = 195.32 (SD = 10.94); $t(153) = 2.01$, $p = .05$.

40. Mean reading score of third graders who responded correctly on the blicket task = 216.89 (SD = 11.7); mean reading score of third graders who responded incorrectly = 210.58 (SD = 13.0); $t(107) = –2.66$, $p = .009$

41. Mean math score of third graders who responded correctly on the blicket task = 223.41 (SD = 10.48); mean math score of third graders who responded incorrectly = 220.10 (SD = 13.45); $t(107) = -1.44$, $p = .15$

42. Mean PSSA language score of children who responded correctly on the blicket task = 1137.22 (SD = 107.17); mean PSSA language score of children who responded incorrectly = 1088.40 (SD = 90.77); $t(105) = -2.50$, $p = .01$

43. Mean PSSA math score of children who responded correctly on the blicket task = 1106.77 (SD = 121.77); mean PSSA math score of children who responded incorrectly = 1074.81 (SD = 105.88); $t(105) = -1.42$, $p = .16$

44. Mean PSSA science score of children who responded correctly on the blicket task = 1575.37 (SD = 163.97); mean PSSA science score of children who responded incorrectly = 1506.47 (SD = 144.08); $t(104) = -2.27$, $p = .03$

Chapter 8

1. During the writing of this book, Dave's daughter was in seventh grade (age 12). Her science teacher asked her class to draw a scientist. She reported that the majority of the students in her class drew pictures consistent with the features described here.

2. While the reference to potions might have to do with the prevalence of Harry Potter, responses featuring potions were common in the Draw a Scientist task and other measures that pre-date that book series. So we can't say that Harry Potter alone caused an uptick in children's beliefs that potions are part of science.

3. Chi-squared tests, all χ^2 values > 7.70, all p-values < .006

4. Chi-squared tests, all χ^2 values < 0.70, all p-values > .40

5. $\chi^2(4) = 316.69$, $p < .001$

6. $\chi^2(4) = 122.37$, $p < .001$

7. $r_s(929) = .05$, $p = .13$

8. $r_s(929) = -.04$, $p = .19$

9. $r_s(929) = -.01$, $p = .71$

10. Exact proportions test, $p = .22$

11. Exact proportions test, $p = .27$

12. Exact proportions test, $p = .03$

13. $r_s(929) = .26$, $p < .001$

14. Exact proportions test, $p < .001$

15. $r_s(929) = .22$, $p < .001$

16. Exact proportions test, $p < .001$

17. $\chi^2(1) = 3.85$, $p = .05$

18. $\chi^2(1) = 4.60$, $p = .03$

19. $\chi^2(1) = 3.65$, $p = .06$

20. $\chi^2(1) = 5.32$, $p = .02$

21. $\chi^2(1) = 0.93$, $p = .33$

22. B = –0.02, SE = 0.007, $p = .004$

23. B = 0.30, SE = 0.07, 95% CI = [0.16, 0.43], Wald $\chi^2(1) = 17.90$, $p < .001$

24. B = 0.49, SE = 0.10, 95% CI = [0.28, 0.69], Wald $\chi^2(1) = 22.22$, $p < .001$

25. B = 0.26, SE = 0.16, 95% CI = [–0.05, 0.57], Wald $\chi^2(1) = 2.69$, $p = .11$

26. B = 0.07, SE = 0.13, 95% CI = [–0.20, 0.33], Wald $\chi^2(1) = 0.25$, $p = .62$

27. $t(1051) = -6.02$, $p < .001$

28. $r(938) = .13$, $p < .001$

29. Children: $r_s(938) = .50$, $p < .001$; Adults: $r_s(111) = .35$, $p < .001$

30. Age effect: $t(929) = -7.14$, $p < .001$; Length effect: $t(929) = -3.67$, $p < .001$

31. Binomial regression predicting receiving a Learning code from the interaction of age and response length: B = –0.0001, SE = 0.0002, $p = .60$

32. $t(92) = -1.76$, $p = .08$

33. $t(107) = -2.76$, $p = .007$

34. $t(107) = 2.74$, $p = .007$

35. Learning: $t(105) = -1.78$, $p = .08$; Other Process: $t(105) = 1.72$, $p = .09$

36. $t(105) = -2.00$, $p = .05$

37. $t(105) = -1.78$, $p = .08$

38. For these analyses we constructed 2 x 2 tables of children whose responses fit both coding categories under consideration, only one or the other category, or neither. We conducted a series of chi-squared tests on these tables to test for relations; all p-values > .06.

39. $\chi^2(1) = 46.47$, $p < .001$

40. $\chi^2(1) = 7.49$, $p = .006$

41. Learning: $\chi^2(1) = 3.97$, $p = .05$, Other Process: $\chi^2(1) = 11.02$, $p < .001$

42. $r_s(70) = .39$, $p = .001$

43. $r_s(69) = .37$, $p = .001$

44. $r_s(69) = .07$, $p = .08$

45. Exact proportion test, $p = .03$

46. These data were collected by Elena Schiavone as part of her undergraduate honors thesis project.

47. "Puzzle" framing: 48% correct, $SD = 51\%$; "Science" framing: 49% correct, $SD = 50\%$; $t(87) = -0.04$, $p = .97$)

48. Callanan et al. (2020) used a slightly different coding system than what we used here. However, the change in coding system does not change the results reported in that paper.

49. All $r_s(301)$-values $> |.15|$, all p-values $< .008$.

50. B = –13.19, SE = 4.10, 95% CI [–21.23, –5.16], Wald $\chi^2(1) = 10.36$, $p = .001$. This is still a negative relation, so generating a definition of science as a specific activity still related to *less* systematic exploration when age is factored into the analysis.

51. B = 14.22, SE = 4.90, 95% CI [4.62, 23.82], Wald $\chi^2(1) = 8.43$, $p = .004$.

52. Engagement: $r_s(107) = -.20$, $p = .04$; Performance $r_s(107) = -.22$, $p = .02$. Note that these are negative correlations, which means that if children generated a definition of science that included a specific activity, they were less engaged by the challenges and less likely to complete challenges on their own.

53. Engagement: $r_s(106) = -.12$, $p = .21$; Performance r_s $(106) = .14$, $p = .14$

54. Correlations with age: Engagement: r_s $(109) = .49$, $p < .001$; Performance: $r_s(109) = .47$, $p < .001$; Generating a definition of science that was specific: $r_s(107) = -.20$, $p = .04$.

55. We analyzed this via a set of Generalized Estimating Equations, controlling for children's age and the type of circuit that they constructed during free play (because not every parent-child dyad built every type of circuit we considered during free play), looking at child and parent action as a count variable (Poisson distribution). Child action: B = 0.28, SE = 0.12, 95% CI [0.05, 0.50], Wald $\chi^2(1) = 5.65$, $p = .02$; Parent action: B = –1.03, SE = 0.45, 95% CI [–1.91, –0.16], Wald $\chi^2(1) = 5.35$, $p = .02$

56. Chi-squared tests, all χ^2 values < 6.60, all p-values $> .16$

57. $\chi^2(4) = 6.49$, $p = .17$

58. Chi-squared tests, all χ^2 values > 10.58, all p-values $< .04$

59. Chi-squared tests, all χ^2 values > 7.80, all p-values $< .10$

60. $r(86) = .52$, $p < .001$

61. This was confirmed by a univariate ANOVA on these scores with age group as a fixed factor, which revealed a significant effect of age group: $F(3, 84) = 8.83$, $p < .001$, $\eta^2 = 0.24$

62. Post-hoc test with Tukey correction, $p = .99$

63. Post-hoc test with Tukey correction, $p = .97$

64. Post-hoc test with Tukey correction, $p = .005$

65. $t(84) = -2.20$, $p = .04$; confirmed with ordinal regression, odds ratio = 3.20, $p = .02$

66. $t(84) = -2.88$, $p = .005$; confirmed with ordinal regression, odds ratio = 3.34, $p = .007$

67. Exact proportions tests, all p-values $< .001$

68. $\chi^2(1) = 14.43$, $p < .001$

69. Exact proportion tests, all p-values $< .04$

70. Exact proportion tests, all p-values $> .07$

71. $\chi^2(1) = 2.67$, $p = .10$

72. Psychology: $\chi^2(1) = 19.67$, $p < .001$; chemistry: $\chi^2(1) = 13.65$, $p < .001$

73. Exact proportion tests, all p-values $< .015$

74. Exact proportion test, $p = .14$

75. $\chi^2(1) = 5.68$, $p = .02$

76. Biology: $\chi^2(1) = 2.85$, $p = .09$; chemistry: $\chi^2(1) = 2.85$, $p = .09$

77. Exact proportion tests, all p-values $< .04$

78. Exact proportion tests, chemistry: $p = .05$, psychology: $p = .005$, biology: $p = .19$

79. Exact proportions tests, all p-values $< .08$

80. Exact proportions tests, all p-values $< .04$

81. Exact proportions tests, all p-values $< .04$

82. Exact proportions test, $p = .28$

83. Exact proportions test, $p < .001$

84. Exact proportions test, $p < .001$

85. Exact proportions test, $p = .23$

86. Exact proportions test, $p = .23$

87. Exact proportions test, $p = .43$

88. Exact proportions test, $p = .005$

89. Chemistry appropriate method, $t(97) = -2.50$, $p = .01$; chemistry inappropriate method, $t(97) = -2.34$, $p = .02$

90. Biology appropriate method, $t(99) = -3.63$, $p < .001$; biology inappropriate method, $t(99) = -0.56$, $p = .58$

91. Psychology appropriate method, $t(99) = 0.77$, $p = .44$; psychology inappropriate method, $t(99) = -2.41$, $p = .02$

92. The semantics of the word "just" are complicated (Lee, 1987). On Lee's analysis, we seem to be following a "restrictive" account, but sometimes children used words like "only," and we suspect that the intention of many of their utterances was to be exclusive (i.e., the character is not doing science). When words like "just" or "only" were applied to cases where the child said the character was doing science, then the word might be more elaborative (Warstadt, 2019).

93. $\chi^2(2) = 3.74$, $p = .15$

94. $\chi^2(2) = 0.51$, $p = .77$

95. $\chi^2(2) = 14.77$, $p < .001$

96. Exact proportion tests, chemistry $p < .001$; biology $p < .001$; psychology $p = .008$

97. Exact proportion tests, chemistry $p = .04$; biology $p = .64$; psychology $p = .18$

Chapter 9

1. Wald $\chi^2(4) = 9.38$, $p = 05$. Specifically, we build a Generalized Estimating Equation assuming a binomial logistic response, with the response to the test question (yes or no) as the dependent measure and with Question Type (Fact vs. Skill), Intention, Outcome, and age (in months) as independent factors, testing for these main effects and this interaction.

2. This work was done collaboratively with Susan Letourneau.

3. McNemar $\chi^2(1) = 1.16$, $p = .28$

4. $r_s(68) = -.24$, $p = .04$

5. $r_s(68) = .28$, $p = .02$

6. $r_s(68) = -.14$, $p = .22$

7. Fisher Exact Test, $p = .02$

8. Factoring out children's age and the mean length of utterance of the definition (a gross measure of language capacity, which we have used throughout these studies), all $r(66)$-values $< |.16|$, all p-values $> .20$.

Chapter 10

1. Much like science, learning, teaching, play, and (as we discuss below) pretending, we have also asked children to define "creativity." This work focuses mostly on 5- to 10-year-olds' understanding that creativity involves novelty and innovation. Understanding the former emerges earlier in development (around age 8) than the latter (sometime in adolescence; see Stricker & Sobel, 2020).

2. It is conceivable that longer-looking time in studies that use the violation of expectation method also involve a certain kind of imagination, given that infants are reacting to what they actually observe based on their expectations of how the world should be, and expectations are imagined representations. We are agnostic as to whether differences in looking-time behavior in this paradigm indicate the same kind of imaginative capacities we discuss here (see e.g., Wellman & Liu, 2007).

3. Since Dave started working on children's understanding of pretense in 1996, he has revised his belief about just about everything related the topic, except for the fact that these studies are just plain strange. He loves them. But they are strange.

4. Overall, the difference among the three codes was significant, Cochran's $Q(2, N=66)=19.60$, $p<.001$. Regression models suggest that the best model that captures generating a definition with agency and age in months is defined by the equation $y=-6.13+0.08x$. For the Not Real code, that equation is $y=-10.29+0.13x$, and for the Mental State code, that equation is $y=-13.47+0.15x$.

5. In *The Lion Guard*, it's revealed (retconned, really) that the "Roar of the Guard" skips a generation. Simba does not have this ability, but it is passed on to his son, Kion.

6. And that trolls are not just limited to studies of pretend play.

7. It's worth noting that it was the use of the tractor beam that kept the *Enterprise* inside a time loop in the episode "Cause and Effect" (S5, E18), so it's possible that the crew was a bit shy of deploying it after that.

8. Bathrooms are never shown on the *Enterprise* in the original *Star Trek* series, but there is a bathroom on a shuttlecraft in *Star Trek: The Motion Picture*. There is also a reference to characters being in a bathroom on the *Enterprise-D* in the *Star Trek: The Next Generation* episode "Home Soil" (S1, E17).

9. That said, *The Love Boat*, *Charlie's Angels*, and *Fantasy Island* all seemed to exist in the same shared universe, so the ontological structure of that show might have been more complex than meets the eye.

10. The original blicket detector did not work this way. It was based on a pressure plate, which was designed by the engineers in the UC Berkeley Psychology Department. They did good work, because no one could replicate their design. That original

detector was built in 1995 and functioned until 2017 (twenty-two years!). We moved to this newer design only when we conceptualized some experiments in which we wanted the detector to activate without any objects on it. No box that we have built subsequently has yet to last more than ten years.

11. Alternatively, a change was made to the fourth picture, and children were asked to fill in a new third picture, which represents a different kind of counterfactual inference (see also Guarjado & Turley-Ames, 2004).

12. It is worth noting, however, that if looking-time studies using violation of expectation provide evidence for counterfactual reasoning abilities (see note 2), then children might have this domain-general capacity much earlier in development.

Chapter 11

1. This is, by the way, the nature of science, and it highlights the importance of reproducibility.

2. This was analyzed via a multinomial logistic regression, treating response as a nominal response, and age and location as independent variables. The overall model was significant, Likelihood Ratio $\chi^2(4)=9.28$, $p=.05$. The difference between the “ask a grown-up” response compared with the “play with it myself” response was significant when location was contrasted, $B=0.91$, $SE=0.44$, Wald $\chi^2(1)=4.16$, $p=.04$, Odds Ratio=2.47, 95% CI [1.04, 5.90].

3. In a parallel sample of 51 children interviewed at the Academy (22 girls and 29 boys; mean age=95.97 months, age range=72–122 months), 53% of children said that they would play with the toy, 22% said that they would ask a grown-up, and 25% said that they would watch another child. This distribution is similar to what we obtained at Providence Children’s Museum, but data collection for this sample was interrupted by COVID. Further investigations should consider whether children’s exploration differs between the two museums and whether children’s reflections on what they learn from their exploration differs. We are currently working on these projects.

4. We also gave these children the same “What is learning?” interview that was described in chapter 9. Their responses on that interview did not differ between the two settings, only their beliefs about how they would want to go about learning something new. Our next step is to reproduce this finding by working with children outside of the museum, giving them the two interviews either immediately before or immediately after their visit.

5. Incidentally, in support of this interpretation, the three museums where we tested the relation between children’s definitions of “science” and their exploratory behaviors were Providence Children’s Museum in Providence, RI, Children’s Discovery Museum in San Jose, CA, and Thinkery in Austin, TX. Although Thinkery does not have the term

“science” in its name, it was rebranded to Thinkery from Austin Children’s Museum to reflect more of a focus on STEAM content (Science, Technology, Engineering, Arts, and Math) and on a more inquiry-rich, play-based learning experience. It is, frankly, much more of a children’s science museum than the other two environments. When the data from the three museums were considered separately, the relation between children’s exploration and their definitions of “science” held in the Providence sample and in the sample from Children’s Discovery Museum individually, but not in the sample from Thinkery.

Constructing Science

Connecting Causal Reasoning to Scientific Thinking in Young Children

By: Deena Skolnick Weisberg, David M. Sobel

Citation:

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

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

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

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

Library of Congress Cataloging-in-Publication Data

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