Explaining Influences Children's Reliance on Evidence and Prior Knowledge in Causal Induction

Caren M. Walker, Joseph Jay Williams, Tania Lombrozo, & Alison Gopnik
(caren.walker@berkeley.edu, joseph_williams@berkeley.edu, lombrozo@berkeley.edu, gopnik@berkeley.edu)
Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720

Abstract

In two studies, we examine how prompting 5- and 6-year-olds to explain observed outcomes influences causal learning. In Study 1, children were presented with data consistent with two causal regularities. Explainers outperformed controls in generalizing the regularity that accounted for more observations. In Study 2, this regularity was pitted against an alternative that accounted for fewer observations but was consistent with prior knowledge. Explainers were *less* likely than controls to generalize the regularity that accounted for more observations. These findings suggest that explaining drives children to favor causal regularities that they expect to generalize, where current observations and prior knowledge both provide cues.

Keywords: Explanation; cognitive development; causal reasoning; prior knowledge; generalization; abduction.

Since Piaget, researchers have regarded children's explanations as a window into cognitive development, revealing children's understanding of the world (e.g., Keil, 2006). More recently, however, the very process of seeking, generating, and evaluating explanations has additionally been proposed as a powerful mechanism for learning and generalization, scaffolding knowledge acquisition and contributing to theory change (e.g., Chi, DeLeeuw, Chiu, & LaVancher, 1994; Lombrozo, 2006; Wellman, 2011). Here we investigate the role of explanation in young children's causal learning, focusing on whether and how explaining influences the relative contributions of observed evidence and currently-held theories ("prior knowledge").

Discovering the underlying causal structure in the world is one of the major inductive problems that young learners face during development. By 5 years of age, children have already developed abstract, coherent representations of causal relationships in a variety of domains (e.g., Carey, 1985; Gelman & Wellman, 1991). The acquisition of this causal knowledge is supported by powerful learning mechanisms that allow children (and adults) to infer novel causal relationships from patterns of evidence and prior beliefs (e.g., Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004; Griffiths, Sobel, Tenenbaum, & Gopnik, 2011). For example, 5-year-old children can effectively track covariational evidence to identify a novel cause, but require stronger evidence to endorse a cause that conflicts with their prior beliefs than one consistent with those beliefs (Schulz, Bonawitz, & Griffiths, 2007). These findings reveal that children integrate current observations with prior beliefs in the process of learning and revising their causal knowledge. Here we propose that explaining may play a role in this integration, by influencing the relative contributions of evidence and prior knowledge – that is, the extent to which learners revise their beliefs in light of their observations and prior commitments.

A first possibility, which we call *the evidence hypothesis*, is that explaining a set of observations directs learners to update their beliefs in light of those observations. As a result, explaining will lead learners to change their beliefs *more* than they would have in the absence of explaining. This might be expected, for example, if explanation boosts children's engagement with a task (Siegler, 2002), thereby making them more responsive to feedback and less likely to disregard relevant observations. Consistent with this possibility, a variety of studies have found that presenting young children with feedback is less effective as a means to changing their beliefs than having them explain as well (e.g., Wellman, 2011). More indirect support comes from the fact that explaining can direct children towards anomalous data, which conflicts with their prior beliefs and may therefore signal a need for belief revision (e.g., Legare, 2010; Legare, Gelman, & Wellman, 2010). In sum, the *evidence hypothesis* predicts that explaining should make children more responsive to observed data, leading to greater revision in current beliefs.

A second possibility, which we call the *prior knowledge hypothesis*, is that engaging in explanation invokes learners' currently-held theories, leading them to revise their beliefs *less* than they might have in the absence of explanation. In line with this hypothesis, several researchers have proposed that explaining encourages learners to accommodate what they are trying to explain in the context of what they already know (e.g., Chi et al., 1994; Lombrozo, 2006). Explaining could thus lead children to favor hypotheses consistent with prior knowledge, even when those hypotheses might not have been entertained or preferred on the basis of evidence alone. Kuhn and Katz (2009) further suggest that by encouraging learners to consider *why* something is the case, explaining can lead them to discount observed evidence altogether. From this perspective, learners who explain could be less inclined towards hypotheses that simply match current evidence, either because they weight prior beliefs more heavily or because they underweight current evidence.

A final possibility is that explaining does not affect the role of evidence or prior knowledge uniformly across contexts. Instead, explaining could change the criteria employed in evaluating evidence and prior knowledge in the course of learning, resulting in greater belief revision under some conditions and less belief revision under others. The version of this possibility that we explore here is the *generalization hypothesis*, according to which generating explanations leads children to evaluate both evidence and prior knowledge as cues to which patterns in their observations are most likely to *generalize* to new cases. This proposal is motivated by the theory-theory of cognitive development (e.g., Carey, 1985; Gopnik & Wellman, 1994) as well as recent research concerning the role of explanation in adult learning (Williams & Lombrozo, 2010a, 2010b). According to the theory-theory, children construct intuitive theories that support explanation, prediction, and control, where theories are regarded as consisting in "causal-explanatory" knowledge. In order to effectively support prediction, successful theories must account for past observations as well as generalize to future observations. The process of generating explanations could therefore influence learning by directing children to construct the causal-explanatory beliefs that they judge most likely to generalize, and to consult both evidence and prior knowledge in making this assessment.

Insight into how explanation could contribute to this process comes from subsumption and unification theories of explanation from philosophy, according to which successful explanations demonstrate how what is being explained can be subsumed under a broad regularity, such as a natural law (e.g., Friedman, 1974; Kitcher, 1989). If explaining drives learners to understand current observations in terms of broadly-applicable regularities, it should facilitate the discovery and generalization of regularities that subsume the greatest number of cases. This idea has recently been developed as an empirical hypothesis concerning the role of explanation in learning, called the *subsumptive constraints account*, and is supported by studies with adults learning novel categories (Williams & Lombrozo, 2010a, 2010b).

To illustrate, consider Williams and Lombrozo (2010a), in which adults learned novel categories from observations and were prompted to either explain each observation's category membership or to engage in a control task. The categories could be differentiated by a subtle regularity that accounted for all observations, or by a more salient regularity that accounted for only a subset of observations. Compared to participants in other conditions, those prompted to explain were more likely to discover the regularity with greater subsumptive scope (the one that accounted for more cases), and more recent evidence suggests that prior knowledge additionally serves to inform these assessments (Williams & Lombrozo, 2010b, under review). The *generalization hypothesis* proposes that explanation influences children's causal learning in a similar way: by directing young learners to the causal hypotheses with the greatest subsumptive scope, which is assessed on the basis of both evidence and prior knowledge.

These candidate hypotheses are not intended as comprehensive accounts, but rather provide a useful framework for considering the possible mechanisms by which explanation could influence causal induction. In the two studies that follow, we manipulate both the observed evidence and children's prior knowledge in order to test these hypotheses and assess the role of explanation in children's causal learning. Specifically, children learned about a novel causal system in which some objects were causally efficacious and others were not. Half of the participants were prompted to explain the causal outcomes during the training phase of the experiment, and the other half were prompted to describe these outcomes. In Study 1, the observed data suggested two candidate causal regularities that were equally consistent with prior knowledge. However, one of the two causal regularities accounted for more of the observed data. In Study 2, children were again presented with two candidate causal regularities that accounted for all or a subset of the data, but the regularity that accounted for fewer observations was more consistent with prior knowledge. In both studies, children were then asked to generate causal predictions in order to assess which causal regularity was discovered, preferred, and generalized to novel cases.

The three hypotheses outlined above generate different predictions across these two studies. The *evidence hypothesis* predicts that children will generalize according to the regularity that covers more of their observations in both studies. The *prior knowledge hypothesis* predicts that because explaining prompts children to rely upon their current theories, those who explain should respond no differently in Study 1, and be more likely to form generalizations in line with their prior knowledge in Study 2. The *generalization hypothesis* predicts that explainers should consider both the data and their prior knowledge to identify the regularity that is likely to apply most broadly, which could lead to opposite generalizations for Studies 1 and 2. Taken together, the results will thus help characterize the role of explanation in children's causal learning.

Study 1: Regularities in Observed Evidence

In Study 1, we present two groups of children with the task of learning a novel causal relationship from a series of observations, where one group is prompted to explain each observation (the *explain* condition) and the other to describe each observation (the *control* condition). The observed data suggest two causal regularities: a regularity that accounts for 100% of observations (the 100% regularity) and a regularity that accounts for 75% of observations (the 75% regularity). The evidence for candidate regularities is presented using a *blicket detector* paradigm (e.g., Gopnik & Sobel, 2000), in which children learn which objects have causal efficacy in activating a machine.

Participants

Forty-two 5-year-old children (M = 64.2 months; SD = 3.6, range: 59.9 – 72.7; 25 girls) were included, with 21 children randomly assigned to each condition. Children were recruited from local preschools and museums.

Materials

The machine used in the training phase of both studies was a "blicket detector" – a box concealing a wireless doorbell.

When an object "activated" the machine, the doorbell was pressed remotely, producing a melody.

An illustration of the complete set of training blocks appears in Figure 1A. Eight 2" wooden blocks (four "*Go*" blocks and four "*No-Go*" blocks) were used during the training phase. A green plastic lego plate was affixed to the top and front of each block. Attached to each lego plate was a single, small rectangular lego of uniform size. The two causal regularities (100% and 75%) were represented by different-colored legos. For the four blocks that activated the machine (*Go* blocks), the 100% regularity was represented by a green lego and the 75% regularity was represented by a red lego. For the four blocks that did not activate the machine (*No-Go* blocks), the 100% regularity was represented by a yellow lego and the 75% regularity was represented by a white lego. For half of the children, the 100% regularity (the green/yellow lego) appeared on the top of the block and the 75% regularity (the red/white lego) appeared on the front of the block, and for half of the children, these positions were reversed.

For the testing phase, four additional blocks were used. These testing blocks were identical to the training blocks, but included only one of the four lego colors on each. Additionally, a cardboard "hiding box" was constructed, with four cut-out windows that were covered with black felt flaps. These windows were designed so the experimenter could place two blocks inside the hiding box and lift two flaps to show the participant only one of the two legos (on the top or front) of each block.

Procedure

The experimenter introduced the machine, explaining that some things make the machine play music and some things do not. The child was then instructed to sort the blocks into two piles according to whether they activated the machine.

The experimenter placed each of the eight blocks on the machine, one at a time. After children observed each outcome, they were asked for a verbal response. In the *explain* condition, children were asked to explain the outcome: "Why did/didn't this block make my machine play music?" In the *control* condition, children were asked to describe the outcome: "What happened to my machine when I put this block on it? Did it make music?" The order of presentation of the eight blocks was semi-random (see Figure 1A). All blocks remained visible and were grouped on the table throughout the training and test phases of the experiment to eliminate memory demands.

For the test phase, the machine was removed and the experimenter introduced the "hiding box," saying: "This is my hiding box. I am going to put two new blocks into my hiding box, and lift these flaps so you can only see part of each block." The child was told: "One of the blocks I put in my hiding box will make my machine play music, and one of the blocks will not. I want you to tell me which one you think *will/will not* make my machine play music."

Each test question was designed to pit one potential causal regularity against the other. In the first set, the *100% and 75% test items*, the *Go* features were pit against the *No-Go* features for both the 100% regularity (green vs. yellow) and the 75% regularity (red vs. white). In the next set, the experimenter presented *conflict items,* in which the *Go* feature from the 100% regularity (green) was pit against the *Go* feature from the 75% regularity (red). Each time, children were asked to predict which block would make the machine play music. These questions were intended to present a conflict between the two potential causes to examine whether children would privilege one regularity over the other. Each of these test questions was repeated for a second time in which the experimenter asked the child to indicate which block *would not* make the machine play music. There was a total of six test questions in this format: four *100%* and *75% test items* (two green vs. yellow; two red vs. white) and two *conflict items* (two green vs. red).

Responses on the *100%* and *75% test items* were scored for accuracy, where accuracy reflected the correct selection of the *Go* block when asked for an item that would make the machine go, and the *No-Go* block when asked for an item that would not. Children received 1 point for selecting the correct response and 0 points otherwise For the *conflict items*, we examined whether responses conformed to the 100% regularity (selecting the green lego over the red lego) or the 75% regularity (selecting the reverse), coding the former as "1" and the latter as "0." [1]

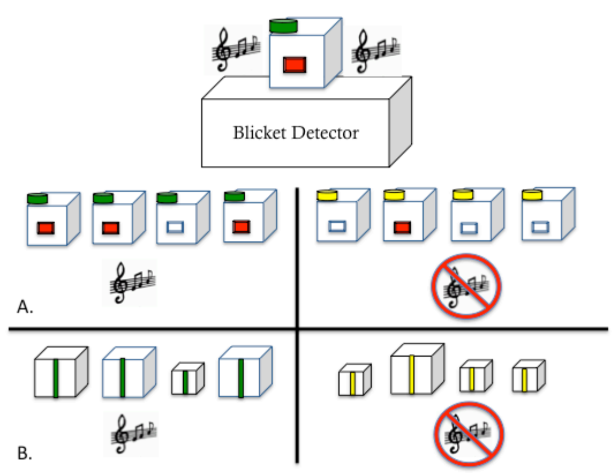

Figure 1: Procedure for Studies 1 (A) and 2 (B).

Results

Whereas children in both conditions were able to learn the 100% and 75% regularities, those who were prompted to explain during the training phase were more likely than controls to privilege the 100% regularity over the 75% regularity when generalizing to novel cases (see Figure 2).

We begin by reporting the results of children's

[1] Both studies additionally included certainty ratings and an explanation selection task with findings that mirror those that we report; we do not report these data in the interest of space.

performance on the *100%* and *75% test items*. A repeated measures ANOVA was conducted with each question type (*100%* and *75% test items*) as the repeated measure and condition (*explain* vs. *control*) as a between-subjects variable. Analysis revealed no difference in children's performance on the two question types, $F(1, 40) = 0.195$, $p = .661$, no difference between conditions, $F(1, 40) = 1.84$, $p = .182$, and no interaction between question type and condition, $F(1, 40) = 0.780$, $p = .382$. Children in both the *explain* condition ($M = 0.87$, $SD = 0.23$) and the *control* condition ($M = 0.77$, $SD = 0.22$) were able to learn the 100% and 75% regularities during the training phase and use this information when generalizing to novel blocks.

To analyze children's performance on the *conflict items,* a one-way ANOVA was conducted with condition (*explain* vs. *control*) as the between-subjects variable. There was a significant difference between conditions, $F(1, 40) = 5.79$, $p < .02$: Children in the *explain* condition were more likely ($M = .667$, $SD = .33$) than those in the *control* condition ($M = .405$, $SD = .38$) to base their generalizations on the 100% regularity. Results support the hypothesis that prompting learners to explain helps them to discover and extend causal regularities that account for more of their observations, consistent with the *evidence* and *generalization* hypotheses.

Study 2: Conflicting Prior Knowledge

Study 1 was designed to assess the role of current evidence in children's causal learning and generalizations. However, because the two candidate causes differed only in color (green vs. red), children had no *a priori* reason to privilege either cause. Study 2 therefore examines the influence of explanation in causal learning when a candidate cause that accounts for more observations is pitted against an alternative that is more consistent with prior knowledge.

Study 2 again presented two sets of causal regularities: a 100% regularity that accounted for all observations (block color), and a 75% regularity that accounted for most observations *and* presented a plausible causal mechanism (block size). Unlike color, relative size provides a plausible cause – larger objects are assumed to be more causally efficacious. This assumption was confirmed during pilot testing. We also included two age groups in Study 2 (5- and 6-year-olds). Because 6-year-olds have more experience with mechanical systems and have begun formal schooling, they could hold stronger prior beliefs in this domain.

If explaining principally drives learners to favor regularities from the data that they observe, then we would expect to replicate the results from Study 1, with explainers more likely to generalize according to the 100% regularity. If, however, explaining prompts learners to privilege causes that are more plausible in light of prior beliefs, then children prompted to explain should be less likely to generalize according to the 100% regularity, instead favoring the prior-knowledge consistent 75% regularity.

Participants

Seventy-two children were included in Study 2, including 36 5-year-olds ($M = 64.4$; $SD = 3.8$; range = 60.1 – 71.7; 20 girls) and 36 6-year-olds ($M = 78.3$; SD = 4.1; range = 72.7 – 83.7; 18 girls). Eighteen 5-year-olds and 18 6-year-olds were randomly assigned to each condition. Children were recruited from preschools and museums.

Materials

Study 2 used the same machine as in Study 1. Training stimuli consisted of four large, 3" wooden blocks and four small, 1" wooden blocks. A strip of colored electrical tape was affixed to each of the eight blocks. The four *Go* blocks had a green strip and the four *No-Go* blocks had a yellow strip. An illustration of the complete set of training blocks appears in Figure 1B. In place of the hiding box, test blocks were hidden in an opaque bag.

Procedure

While the procedure for the training phase and test questions was similar to Study 1, there were some changes to the procedure for the testing phase. Rather than placing the test objects in the hiding box, the experimenter looked inside the opaque bag and described one feature for each of two new blocks, saying, for example, "I see a green one and I see a yellow one. Which one will make my machine play music, the green one or the yellow one?"

Results

As in Study 1, children in both conditions were able to learn the 100% and 75% regularities. However, unlike in Study 1, children who were prompted to explain were significantly *less* likely than controls to privilege the 100% regularity over the 75% regularity when forming generalizations about novel cases, instead favoring the regularity consistent with prior knowledge (see Figure 2).

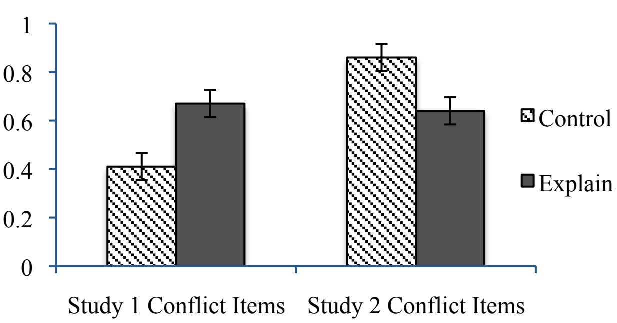

Figure 2: Mean proportion of responses consistent with the 100% regularity on the conflict items in Studies 1 and 2.

We first report results of children's performance on the *100%* and *75% test items*. A repeated measures ANOVA was conducted with question type (*100%* and *75% test items*) as the repeated measure and both age (5- vs. 6-year-olds) and condition (*explain* vs. *control*) as the between-subjects variables. Analysis revealed a significant interaction between age and performance on the two question types, $F(1, 68) = 4.46$, $p < .05$, with 6-year-olds showing higher accuracy (M = .96 SD = 0.18) than 5-year-olds (M = .79, SD = 0.33) on the *75% test items*. However, there was no main effect of condition, F (1, 68) = .73, p = .40, and no interaction between question type and condition, F(1, 68) = 0.04, p = .85. As in Study 1, children in both the *explain* condition (M = .93, SD = .24) and the *control* condition (M = .89, SD = .30) were able to learn the 100% and 75% regularities during the training phase, and use this information when generalizing to novel blocks.

To analyze children's performance on the *conflict items*, a univariate ANOVA was conducted with both age and condition as the between-subjects variables. As in Study 1, there was a significant difference between conditions on the *conflict items*, F(1, 68) = 6.46, p < .02. However, results reveal a reversal of the findings. While children in both conditions responded in line with the 100% regularity more often than chance, p < .05, children in the *explain* condition were significantly *less* likely to do so (M = .64, SD = .37) than children in the *control* condition (M = .83, SD = .27). There was no difference in performance between 5- and 6-year-olds, F(1, 68) = .13, p = .72, and no interaction between age and condition, F(1, 68) = 1.18, p = .28.[2]

General Discussion

In two studies, young children's attempts to generate explanations influenced which candidate cause they privileged when generalizing to novel cases. In Study 1, children were presented with data consistent with two causal regularities, where one accounted for more observations. Children who were prompted to explain were more likely than controls to generalize according to the regularity that accounted for more of the observed data. In Study 2, a regularity that accounted for all observations was pitted against an alternative that accounted for fewer observations but was consistent with children's prior knowledge about plausible causes. When presented with this conflict, children who were prompted to explain their observations were now *less* likely than controls to generalize the regularity that accounted for more of their observations, instead making judgments in line with prior knowledge.

Taken together, these studies shed light on the mechanisms by which explanation informs and constrains causal learning. Our findings challenge the *evidence hypothesis*, that explaining (always) makes children more responsive to evidence. If this were the case, explaining should have led children to base their judgments on the regularity that was most consistent with their current observations in both Studies 1 and 2. Our findings also challenge the *prior knowledge hypothesis,* according to which the (primary) role of explanation is to align new information with current theories. While children prompted to explain were more likely to form a generalization in line with their prior knowledge in Study 2, children were also more likely to generalize according to their observations in Study 1, indicating that the *prior knowledge hypothesis* is limited at best, and false at worse. Instead, these findings best support the *generalization hypothesis*, that explanation prompts children to privilege observed regularities that they expect to generalize most broadly, with both observations and prior knowledge serving as cues to generalization.

The *generalization hypothesis* is broadly consistent with the "theory-theory" approach to cognitive development, and additionally provides some insight concerning how the process of theory construction and revision occurs. Within cognitive development, researchers have suggested that children's theory-like conceptual development is largely motivated by explanation, which acts as a mechanism for building abstract causal theories. In particular, Wellman and Liu (2007) propose that explanations make a particular occurrence understandable by placing it within the context of a larger, coherent framework. In so doing, explanation must accommodate both observations and prior beliefs. Relatedly, explanations have been shown to encourage adult learners to *subsume* the observation being explained under a general pattern that is expected to generalize to novel cases (Williams & Lombrozo, 2010a, 2010b). Our current findings not only provide additional support for these ideas by demonstrating that young children are also sensitive to such constraints, but additionally shed light on how explanations contribute to the formation of causal theories: Explaining helps negotiate the critical balance between prior beliefs and novel evidence, directing learners to regularities that will generalize to new cases.

Important questions for future research include precisely *how* explanation influences the evaluation of evidence and prior beliefs, and whether this influence results in judgments that are more or less closely aligned with the predictions of normative (Bayesian) models (e.g., Griffiths et al., 2011). For example, computational approaches to cognitive development have proposed that the formation of multiple levels of generalizations, or *overhypotheses*, enables learners to make principled abstractions from a class of observations, which then serves to inform future inferences about novel cases (e.g., Kemp, Perfors, & Tenenbaum, 2007). By prompting children to evaluate which candidate regularity generalizes most broadly, explanation could play a role in pushing children to consider higher-order inductive generalizations that support abstract knowledge.

The interpretation outlined thus far has focused primarily on the impact of prompts to explain on children's causal judgments. However, it is also worthwhile to consider the performance of children who were not prompted to explain. In particular, why didn't the control children in Study 2 spontaneously consult their prior knowledge in forming causal judgments? Williams and Lombrozo (under review) report similar findings in an adult population: Only learners who were prompted to explain during learning utilized informative labels in guiding their discovery of subtle

[2] Due to limited space, we do not report the results of qualitative analyses examining the content of children's explanations in Studies 1 and 2. These results are consistent with the quantitative data and provide additional support for our conclusions in the General Discussion.

patterns underlying novel categories. The authors suggest that explanation can guide learners to consult prior knowledge that would otherwise remain under-utilized.

The current work also suggests a number of future directions. For example, while the results reported here demonstrate the presence of an early effect of explanation on causal learning, questions remain regarding the development of this mechanism from infancy to adulthood. Additionally, future research should consider how these findings can be productively combined with previous research on the self-explanation effect (e.g., Chi, et al., 1994; Lombrozo, 2006; Siegler, 2002), as well as examining educational implications of the present findings.

In sum, these two studies provide evidence for the mechanisms by which explaining scaffolds causal learning in early childhood and serve to inform prevailing theories of learning. Learning by explaining challenges a simple data-driven view of knowledge acquisition in which children's learning is simply a function of observation and testimony. Instead, these findings provide evidence for more complex theories of learning in which processes such as explaining to oneself influence how data and current theories inform judgments. Understanding how engaging in explanation influences learning therefore contributes to a more complete understanding of how knowledge is acquired and revised.

Acknowledgments

Thanks to Anna Akullien, Brynna Ledford, Sierra Eisen, and Rosie Aboody, the UC Berkeley Early Childhood Centers, Lawrence Hall of Science, and Habitot. Research supported by a McDonnell Foundation grant to A. Gopnik and NSF grant DRL-1056712 to T. Lombrozo.

References

Ahn, W.K., & Kalish, C. (2000). The role of mechanism beliefs in causal reasoning. In F. Keil & R.A. Wilson (Eds.), *Explanation and cognition.* (pp. 199-226). Cambridge, MA: MIT Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press/Bradford Books.

Chi, M.T.H., de Leeuw, N., Chiu, M.H., LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science,18,* 439-477.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, *71*, 5-19.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition, 38,* 213–244.

Gopnik, A. & Wellman, H.M. (1994). The "theory theory". In L. Hirschfield and S. Gelman (Eds.) *Domain specificity in culture and cognition*. New York: Cambridge University Press.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 1–31.

Gopnik, A. & Sobel, D. (2000). Detecting blickets: How Young children use information about novel causal powers in categorization and induction. *Child Development*, 71: 1205-1222.

Griffiths, T. L., Sobel, D., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science, 35*(8): 1407-1455.

Keil, F. (2006). Explanation and understanding. *Annual Review of Psychology*, 57: 227-254.

Kemp, C., Perfors, A., & Tenenbaum, J.B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307-321.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher and W. Salmon (Eds.) *Scientific Explanation* (pp. 410-505). Minneapolis: University of Minnesota Press.

Kuhn, D. & Katz, J. (2009). Are self-explanations always beneficial? *Journal of Exp. Child Psych*, *103*, 386-394.

Legare, C.H. (2011). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Dev., 81*, 929-944.

Legare, C.H., Gelman, S.A., & Wellman, H.M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Dev., 81*, 929-944.

Lombrozo, T.L. (2006). The structure and function of explanations. *Trends in Cog. Sci.*, 10: 464-470.

Schulz, L.E., Bonawitz, E.B., & Griffiths, T.L. (2007). Can being scared make your tummy ache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology, 43,* 1124-1139.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), Microdevelopment: Transition processes in development and learning (pp. 31-58). New York: Cambridge Univ.

Wellman, H.M. (2011). Reinvigorating explanations for the study of early cognitive development. *Child Development Perspectives*, 5(1): 33-38.

Wellman, H.M. & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In *Causal Learning: Psych, Philosophy, & Computation* (Eds.) L. Schulz & A. Gopnik, pp. 261-279.

Williams, J. J., & Lombrozo, T. (2010a). The role of explanation in discovery and generalization: evidence from category learning. *Cognitive Science*, 34: 776-806.

Williams, J. J., & Lombrozo, T. (2010b). Explanation constrains learning, and prior knowledge constrains explanation. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Austin, TX.

Williams, J. J., Lombrozo, T., & Rehder, B. (2010). Why does explaining help learning? Insight from an explanation impairment effect. Proceedings of 32nd Annual Conference of the Cognitive Science Society.

Williams, J. J., & Lombrozo, T. (under review). Explanation constrains learning and prior knowledge constrains explanation.