Restructuring Causal Concepts

Eric G. Taylor (etaylor4@illinois.edu)
Department of Psychology, 603 East Daniel St.
Champaign, IL 61820 USA

Abstract

Typical studies of concept learning in adults address the learning of novel concepts, but much of learning involves the updating and restructuring of familiar conceptual domains. Research on conceptual change explores this issue directly but differs greatly from the formal approach of the adult learning studies. This paper bridges these two areas to advance our knowledge of the mechanisms underlying concept restructuring. The main idea behind this approach is that concepts are structured by causal-explanatory knowledge, and hence, models of causal induction may help to clarify the mechanisms of the restructuring process. A new learning paradigm is presented to study the learning *and revising* of causal networks. Results show that some behaviors indicative of conceptual change arise from basic causal learning mechanisms. Results also support models of causal induction that assume inhibition between competing causes.

Keywords: knowledge restructuring, conceptual change, belief revision, causal induction, concept learning.

Concept learning is an incremental process. We learn a concept for the first time only once, and often our initial understanding is flawed. The remainder of learning involves the updating, revising, and restructuring of previous conceptual knowledge. The critical implication—that most concept learning is actually the refinement of familiar concepts—runs counter to the traditional approach in the study of concept learning in adults, which has focused on the learning of entirely novel concepts (Murphy, 2002). Many open questions remain on the nature of concept restructuring.

The goal of this work is to better understand the basic mechanisms of concept restructuring by forging a connection between traditional work on concept learning and the literature on *conceptual change*. Although these two areas differ greatly (in everything from goals to dependent measures), this paper builds on recent work that highlights their commonalities.

Studies of conceptual change typically outline the process of knowledge restructuring in broad strokes: e.g., by showing that it often occurs abruptly (Kuhn, 1962), that people are highly resistant to giving up their prior beliefs (Chinn & Brewer, 1993), and that novice concepts appear to "differentiate" and "coalesce" over the course of development (Carey, 1985). To support these claims, authors have focused on specific real world domains and the shifts in knowledge therein, such as children's learning of biological concepts (Carey, 1985) and young adults' learning of physics (diSessa & Sherin, 1998).

These studies differ dramatically from the traditional research on concept learning in adults, despite great overlap in interests. The adult work has primarily used domain-general laboratory paradigms and formal models to assess the specific representations and processes underlying basic conceptual tasks like classification, inference, and category-based induction (Murphy, 2002).

A complete understanding of concept learning and restructuring requires explanations from both levels of analysis. This paper suggests that recent work developing the *theory view* of concept representation (Gopnik et al., 2004; Murphy & Medin, 1985; Wellman & Gelman, 1992) serves as a linkage between these levels. The theory view states that concepts are built upon networks of causal-explanatory knowledge. This knowledge affects performance in laboratory-based learning tasks (Murphy, 2002) and plays a role in the learning and development of real world concepts where conceptual change effects are typically demonstrated (Vosniadou, 2008). Assuming that concept learning amounts, in large part, to the learning of causal relations, then models of causal reasoning (Kim & Ahn, 2002; Rehder, 2003) provide the requisite theoretical tools for understanding the basic mechanisms of concept learning and potentially also conceptual change.

Few previous studies address this linkage to concept restructuring, however. Murphy's work (e.g., Kaplan & Murphy, 2000) has examined cases where prior causal knowledge is invoked when learning later concepts, but in these studies the prior concepts are not revised. Work on order effects in causal induction suggests that what is learned from the first half of a set of contingency data may be overwritten by later contingencies (Ahn & Marsh, 2006), but the initial learning (and hence, what is restructured) is not typically evaluated. A developmental study by Schulz, Bonawitz, and Griffiths (2007) showed that 4 to 5-year old children inferred causal relations from evidence that ran contrary to their prior beliefs. However, their evidence for belief revision, as measured by transfer performance, was mixed. This study is perhaps the strongest empirical evidence linking studies of causal induction to concept restructuring.

Other findings bearing directly on concept restructuring are less tied to the formal approach. Chinn & Brewer (1993) documented the many ways that people react to anomalous data, only one of which (the least common) was genuine concept revision. Chinn & Brewer (2001) also proposed a set of mental models for interpreting people's verbal evaluations of anomalous data and patterns of belief change,

but these were not formalized at the level specified in the causal induction models.

To directly address the linkage between concept learning research in the theory view tradition and studies of concept restructuring, I developed a task in which individuals would learn and then *revise* their hypothesized causal relations for a novel conceptual domain. The task was inspired by causal structure learning in real world domains, where one often develops a naïve, incorrect view of the underlying causal structure, and then with the accumulation of knowledge and evidence, restructures their original beliefs.

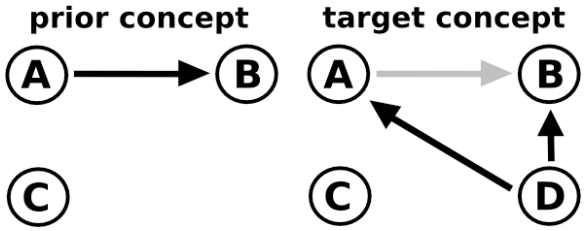

Figure 1: Diagrams of a hypothetical learner's causal representations en route to learning a common cause relation. The prior link remains in the target concept, though reduced, signifying a possible residual belief in that link.

The "common cause" scenario is one of many ways a learner may develop a prior, naïve concept and then need to restructure that concept based on new knowledge and evidence. See Figure 1 for an example. In this scenario, two variables—A and B—will appear correlated, and without further scrutiny, one may assume these variables share a direct causal relation. In fact, both A and B are caused by a third variable, the common cause. When the common cause becomes known, learners can track the relations it shares with variables A and B, and rule out the direct causal relation initially hypothesized.

This paper uses an empirical study based on the common cause scenario as a starting point to understanding the mechanisms underlying shifts in causal knowledge. Given that we currently know much about the initial learning of causal relations (i.e., the learning of the initial A causes B link), this study asks how that initial learning affects the process of concept restructuring. In particular, how does the belief in the prior concept affect later learning where one views contingency data in favor of the target explanation?

Consider the possible effects the prior concept may have on inferring the target structure. First, the prior concept may serve as an anchor, or bias, such that people show commitment to the A.B link (A.B means "A causes B") and later learning of alternative causes is more difficult. Previous work shows that prior beliefs are difficult to give up, especially when they figure centrally in other causal explanations (Chinn & Brewer, 1993).

Second, the acquisition of the prior belief may actually benefit later learning. In particular, evidence suggesting the lack of a correlation between other nodes in the system (between C&A and C&B) might draw resources away from those nodes and facilitate later search for the correct causal mechanism. This is especially true in Figure 1 since an alternative explanation for the A&B correlation is a mediating causal pathway, A.C.B. To the extent that one can rule out this "mediating cause" explanation, they might rule in the common cause explanation.

Third, both previous effects may occur. That is, learning the prior might increase one's belief in the A.B link, and independently, guide learners away from the wrong links and toward the right ones. If learners infer both the common cause and maintain a belief in the direct cause, they will have "over-explained" the occurrence of event B. Although previous work shows that people prefer simple explanations with fewer causal links (Lombrozo, 2007) and that competing causal hypotheses are considered in opposition (Lu et al., 2008), none have examined a case where learners are committed to a prior alternative conceptual structure, as is typically found in studies of conceptual change. In this case, people might over-explain to retain both possible causal pathways.

Experiment

The goal of the experiment was to determine how previously learned causal relations affect continued learning and concept revision. I created an experimental paradigm analogous to Figure 1. One group, the *change* condition, was verbally instructed on a prior structure with three nodes where A directly causes B, then in a second phase, was shown a fourth node (D) and had to infer the correct causal structure from contingency data. The control group, the *no-change* condition did not learn the prior structure and immediately attempted to infer the correct structure from contingency data with nodes A-D. The question is: How does the learning of the prior concept in the change condition affect the learning of the target concept, relative to that of the no-change condition?

Two dependent measures assessed learners' knowledge of the causal system. First, after the prior and target learning phases, participants rated the likelihood of each possible configuration of the system (e.g., A/~B/C for the prior phase, A/~B/~C/D for the target phase). These were used to infer participants "implicit" causal models of the system via model fitting, with the idea that some predictions offered above might not hold if participants were asked directly about their beliefs in the causal links (due to experimenter demands). Second, participants were asked at regular intervals during the target learning phase which of a set of possible links they believed were true. These judgments correspond to participants "explicit" beliefs about the causal system, similar to typical causal induction measures.

Method

Participants Forty-eight University of Illinois students participated in exchange for course credit.

Materials Participants learned about a fictitious ecosystem composed of four observable properties. Each property varied probabilistically during learning, taking one of two binary values (see Figure 2 for "on" values). The first property was the population size of a new fish biologists call "tespula": **above average** or **normal**. The second property was the color of a new type of algae called "plemocyn": **very green** or **normal**. The third property was the chemical composition of barium contained in the ecosystem's water: **crystallized** or **not crystallized**. The fourth property was the cloudiness of the water: **cloudy** or **not cloudy**. I refer to the first mentioned values as the "on" values.

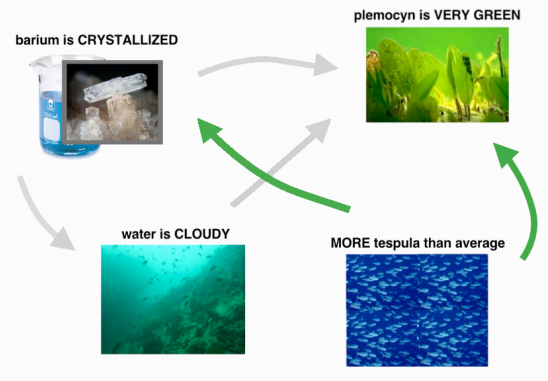

Figure 2. The causal structure of the ecosystem. Darkened links indicate that properties share a generative causal relation with causal power 0.85.

During covariation trials, the property values on each trial were determined by a causal system displayed in Figure 2. When the tespula population is more than average (base rate equal to 0.6), this will cause the barium to be crystallized with probability 0.85, and independently, the plemocyn to be green with probability 0.85. The water will be cloudy with probability 0.6. When the tespula population is average (depicted by a less colorful picture not shown in Figure 2), all other properties will be "on" with probability 0.6.

Covariation trials appeared like Figure 2, except that the property values varied probabilistically and all arrows appeared in grey. During the test phases, participants viewed all "on/off" combinations of the four properties and told to rate their likelihood (see *Procedure* section). In this phase, the arrows were completely absent.

Design Participants were divided into two groups: *change* and *no-change*, corresponding to those given a prior belief regarding the properties' causal relations and those who were not, respectively. Each group was subdivided into four counterbalance conditions, controlling for which properties were assigned to the roles in the causal system.

Procedure Prior to the experiment, participants read and signed a consent form. Participants then read instructions and completed all tasks on a computer.

Change condition: The instructions stated that the purpose of the task was to learn about a new oceanic ecosystem. Specifically, the task was to help a group of biologists to understand how the properties of the ecosystems cause one another. Three properties of that ecosystem were described—the top two properties from Figure 2 (A and B) and the bottom left property (C). The fourth property was absent during this phase. Participants were told that the biologists' current understanding was that property A causes property B (and told nothing else about C). They were also shown a picture with properties A-C and a green arrow connecting A to B. To ensure understanding, participants answered a multiple-choice question asking which properties were related and in what way. If they answered incorrectly, they repeated the instructions and re-took the question until they were correct.

Next, participants entered the *prior learning phase* where they viewed a sequence of 30 "snapshots" of the ecosystem. Each snapshot depicted a particular on/off configuration of properties A-C. Each snapshot appeared with a frequency proportional to its likelihood, which was determined using the probabilities given in the *Materials* section. To compute the probability of a particular snapshot, one computes the probability of each node taking its presented on/off value (conditional on the parent nodes) and then takes the product. Rehder (2003) describes this procedure building on Cheng's (1997) causal power theory, showing that the probability of node N being "on" is $1-(1-b_N)\prod(1-m_{CN})^{Con}$, where b_N is the probability of some unobserved background cause leading to the presence of node N, m_{CN} is the probability that node C generates the presence of N, and Con is an indicator variable equal to 1 when feature C is "on" and 0 otherwise. The snapshot frequencies were identical for all participants, but the order was random and different for each. Note that the causal system from Figure 2 creates a correlation between properties A and B, which supports the belief that A causes B when the status of property D is not visible.

After the 30 snapshots, participants entered the *prior likelihood rating phase* where they viewed each possible snapshot and were told to rate how likely the ecosystem is to look like the snapshot. They were also told, "when making the judgments, be sure to keep in mind the fact that the biologists think that [property A] causes [property B]." Ratings were given by moving a vertical bar up and down a scale, where the highest position indicated "VERY likely" and the lowest indicated "NOT likely."

Then, participants entered the *target learning phase*. They were told that the biologists discovered an important new aspect of the ecosystem, property D, and now they are wondering if their previous belief that A causes B was "wrong or perhaps missing something." They viewed a diagram similar to Figure 2 except with no links darkened, and were told their next task was to help the biologists figure out which of the shown potential causal relationships were true. Participants would learn which causes were true by viewing snapshots like those in the prior learning phase. The instructions also clarified that each property may occur

without being caused by another observed property (i.e., even if X causes Y, Y may appear in the absence of X) and that the links were not necessarily deterministic (e.g., if X causes Y, Y is simply more likely to appear in the presence of X). Finally, they were told that in addition to viewing the snapshots, they would sometimes be making predictions about which of the causes are true. Later during learning, the computer would give feedback about whether their hypotheses were close to or far from the true structure.

After every 10 snapshots participants were asked to guess which of the possible links were true. They were shown the picture in Figure 2 but with no links darkened, and told to click on the links to make their guess. Links darkened when selected. To assist with learning, participants were given indirect feedback regarding their link choices starting on their 4th hypothesis trial (after 40 snapshots)[1]. They were never told the status of any particular link choice (e.g., that the A.B link was right or wrong). Instead, they were told that the hypothesis was VERY GOOD, GOOD, WEAK, or VERY WEAK, indicating that 5, 4, [3 or 2], [1 or 0] links were correct, respectively. Participants were not told the correspondences between the feedback and number of accurate links. On the final hypothesis, participants were told, "This is your LAST PREDICTION. On the next trial, make your best guess as to what causes what."

Finally, in the *target likelihood rating phase*, participants again rated the likelihood of all possible snapshots of the ecosystem but this time with nodes A-D.

No-change condition: The no-change condition was identical to the change condition, but the prior learning phase and the prior likelihood ratings phase were excluded. The instructions immediately introduced participants to all four aspects of the ecosystem and the five possible links. Participants then began the target learning phase.

Results and Discussion

Hypotheses First, I present the results from the hypotheses participants made during the target learning phase. Each link was analyzed separately. Hierarchical logistic regression was used to evaluate the effects of condition and hypothesis trial on link choice. The "hierarchical" component refers to a random intercept term, which was used to model the between-participant variability in overall response tendency.

Results are plotted in Figure 3. To reduce inter-trial variability, I blocked the trials, except for the final trial: 1-4 (without feedback), 5-11 (feedback 1st half), 12-17 (feedback 2nd half), and 18 (the final trial). Main effects and interactions were assessed using Wald tests and likelihood ratio comparisons, but only Wald tests are reported. Likelihood tests led to similar interpretations.

The main effect of block on choosing the A.B link was significant, $\chi^2(1)=10.23$, $p<0.01$, suggesting that learning did occur, as participants selected this incorrect link less over time. The main effect of condition was marginally significant, $\chi^2(1)=3.44$, $p=0.06$, revealing an early and late bias in the change condition to select the prior link. The interaction was marginally significant, $\chi^2(1)=3.33$, $p=0.07$.

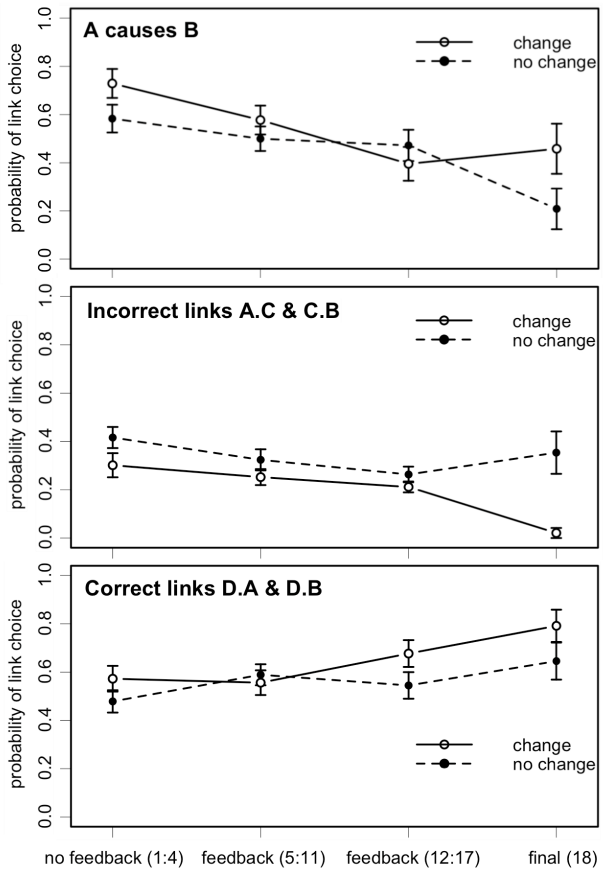

Figure 3. The probability of a participant including a link in their hypotheses during the target learning phase. Error bars are standard errors (binomial variance for final block).

Because the difference in conditions for the A.B link was non-monotonic over blocks, two separate regressions were fit to bocks 1-3 and blocks 3-4. The interaction between trial and condition was significant for blocks 1-3, $\chi^2(1)=9.36$, $p<0.01$, and for blocks 3-4, $\chi^2(1)=6.81$, $p<0.01$. Finally, the difference in conditions on just the final hypothesis was assessed using Fisher's exact test, which did not reach significance, $p>0.1$.

The interactions between trial and condition for the A.B link have two implications. First, although the change condition began selecting A.B more than the no-change condition, this difference went away by the third block as both conditions learned to not select A.B. Second, the difference in conditions increased from blocks 3 to 4.

[1] Feedback was added to improve learning based on the results of a pilot study and previous work showing poor learning for 3-4 node structures given only covariation data (e.g., Lagnado & Sloman, 2004; Steyvers et al., 2003). Feedback is natural in real world learning and is usually provided by confirming or disconfirming predictions made on the basis of hypothesized causal relations. The feedback in this task can be viewed as a proxy for the outcome of multiple such predictions.

Relative to the no-change condition, the change condition was more likely to retain a belief in the prior concept in their final judgment, despite both groups having chosen this link equally often during the final block of feedback.

The incorrect links A.C and C.B were analyzed together. The interaction between block and condition was not significant, $\chi^2(1)<1$. The main effect of block was significant, $\chi^2(1)=23.02$, $p<0.01$. The main effect of condition was also significant, $\chi^2(1)=4.49$, $p<0.05$, even when considering only the final hypothesis (Fisher's exact test, both ps<0.01). This advantage for the change condition is sensible; they are likely attributable to the extra learning in the change group during the prior learning phase. The scientists' tentative theory regarding the ecosystem implied no causal relation between node C and either A or B. Further, the 30 covariation trials suggested little correlation between these nodes, corroborating the scientists' view.

The correct links D.A and D.B were also analyzed together. The interaction between block and condition was not significant, $\chi^2(1)=2.63$, $p=0.10$. The main effect of block was significant, $\chi^2(1)=15.74$, $p<0.01$. The main effect of condition was not significant, $\chi^2(1)=1.34$, $p>0.10$, though there was a tendency for to change condition to choose these links more often.

Likelihoods judgments Likelihoods judgments were used to infer participants' latent causal representations via model fitting. Causal model theory (CMT; Rehder, 2003) and a version of causal support (Griffiths & Tenenbaum, 2005) were fit to each individual's data. Only the results from CMT are presented here, since they were very similar to the results from causal support.

Causal model theory fits were obtained via maximum likelihood estimation. Each fit yields an estimate of nine free parameters: the strength of each potential causal relation in Figure 2, plus an estimate of the probability that some unobserved background node causes each feature. The fitting routine worked by assuming that the participants' likelihood judgments were guesses about the relative frequency of the snapshots, should they be sampled again. Thus, 100 new snapshots were created with frequencies proportional to the normalized likelihood judgments of each participant. The MLE parameter values were those that maximized the likelihood of the snapshots.

The fits to CMT are presented in Table 1. Fitted background probabilities did not differ between the groups, but estimates of causal strength were different, and in the same direction as the differences present in the hypotheses data. First, the difference for link A.B was significant, $t(46)=2.39$, $p<0.05$, reinforcing the non-significant trend in the hypotheses data. This implies that the change condition represents the prior link stronger than the no-change condition, and this difference is robust for the more implicit measure, the likelihoods, where causal strength is not queried directly.

The conditions did not differ significantly in their representation of the incorrect links A.C and C.B, but the differences in the correct links were marginally significant: the change condition represented the D.A link more strongly, $t(46)=1.82$, $p=0.06$, as well as the D.B link, $t(46)=1.93$, $p=0.08$. In addition, when averaging the strength of the correct links, the difference in conditions was reliable, $t(46)=2.32$, $p<0.05$.

Table 1. Average causal strengths (standard deviations).

	No change	Change	*p*-values
Link A.B	0.06 (0.08)	0.14 (0.14)	0.02
Link A.C	0.10 (0.11)	0.06 (0.08)	0.12
Link C.B	0.07 (0.12)	0.05 (0.06)	0.50
Link D.A	0.20 (0.15)	0.28 (0.16)	0.08
Link D.B	0.25 (0.16)	0.35 (0.20)	0.06
Average of D links	0.22 (0.13)	0.31 (0.14)	0.02

The latter result is in line with a predictions stated earlier that the change group may benefit from the prior learning phase by observing the lack of a correlation between nodes A&C and between nodes C&B. Recall that links A.C and C.B constitute an alternative explanation of the A/B correlation; i.e., that A causes C causes B. Put simply, this set of links may be considered in opposition to the common cause links D.A and D.B in order to avoid over-explaining node B. If so, a reduced belief in the former may increase one's belief in the latter.

The idea that alternative causes compete or inhibit one another has empirical backing (Rehder & Milovanovic, 2007) and is made explicit in recent models of causal induction (Lu et al., 2008). In the current study, to evaluate the relation between choices of links involving the two explanations, I used a hierarchical linear regression with number of correct links chosen as the dependent variable and number of incorrect links as the predictor. The predictor variable was separated into two parts: the participant-level effect (the average number of A.C and C.B links chosen by a participant) and the within-participant effect (the number of links chosen on a given hypothesis minus the participant's average). These variables address different questions: the former asks whether participants who choose more incorrect links on average tend to choose more correct links; the latter asks whether on a given trial the number of incorrect links chosen affects the number of correct links chosen.

The effects of the two predictors were evaluated via model comparison. A model excluding the between-participant effect did not fit worse than a model including both effects, $\chi^2(1)=0.26$, $p>0.10$. However, a model excluding the within-participants effect did fit worse than the model with both effects, $\chi^2(1)=52.03$, $p<0.01$, suggesting that causal links involved in competing explanations inhibit one another on a trial-by-trial basis. To my knowledge this is the first evidence showing that competition occurs at the level of entire explanations (i.e., sets of causes), beyond simply individual causal relations.

Conclusion

The goal of this paper was to show that some aspects of concept restructuring might result from basic causal learning mechanisms, thus bridging the formal approach to concept learning with the conceptual change literature. In a novel learning task, participants first developed a prior conceptual belief and were then prompted to revise that concept through contingency learning. Results showed that the prior learning phase led participants to retain their original belief despite evidence against it but also led to enhanced learning of the target causal structure. That is, despite learning of the target, individuals retained the belief in the prior at the cost of over-explaining. Further evidence showed that when revising one's beliefs, alternative causal explanations are considered in opposition, building on the predictions of recent models for simpler causal structures.

Conceptual change surely involves many processes and representations, only some of which are the learning and revising of causal structures (and within that, only some of which are learning from contingency data; Ahn et al., 1995). For example, people also revise their taxonomic hierarchies (Thagard, 1992) and accrue domain-specific knowledge (Carey, 1985). In addition, full-blown conceptual change presumably requires the restructuring of numerous causal hypotheses and may result in emergent representations inherently unalike the prior beliefs. However, current models incorporate powerful learning mechanisms that are capable of such large-scale changes (Kemp & Tenenbaum, 2008). The hope is that improved cross-talk between formal, empirical, and developmental studies will help to build an integrated view of concept learning and conceptual change.

Acknowledgments

Thanks to Brian Ross, John Hummel, Jose Mestre, Woo-kyoung Ahn, Bill Brewer, Noah Goodman, Frank Keil, Tania Lombrozo, Bob Rehder, Pat Shafto, Dan Navarro, and three reviewers for their very helpful comments.

References

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation vs. mechanism information in causal attribution. *Cognition*, 54, 299-352.

Carey, S. (1985). Conceptual Change in Childhood. Cambridge, MA: Bradford Books, MIT Press.

Cheng, P.W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367-405.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63, 1-49.

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, 19, 323-393.

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, 20(10), 1155-1191.

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.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 354-384.

Fugelsang, J., Stein, C., Green, A., & Dunbar, K. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology*, 58, 132-141.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *JEPLMC*, 26, 829-846.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *JEP:G*, 131(4), 451-476.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *PNAS*. 105(31), 10687-10692.

Kuhn, T. S. (1962). The Structure of Scientific Revolutions, 1st. ed. Chicago: University of Chicago Press.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, 55(3), 232-257.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, 115, 955-982.

Marsh, J. K., & Ahn, W. (2006). Order effects in contingency learning: The role of task complexity. *Memory & Cognition*, 34(3), 568-576.

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.

Lagnado, D. & Sloman, S.A. (2004). The advantage of timely intervention. *JEPLMC*, 30, 856-876.

Rehder, B., & Milovanovic, G. (2007). Bias toward sufficiency and completeness in causal explanations. Proceedings of the 29th Annual Conference of the Cognitive Science Society.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *JEPLMC*, 29, 1141-59.

Schulz, L.E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology*, 43(5), 1124-1139.

Steyvers, M., Tenenbaum, J., Wagenmakers, E.J., Blum, B. (2003). Inferring Causal Networks from Observations and Interventions. *Cognitive Science*, 27, 453-489.

Thagard, P. (1992). Conceptual Revolutions. Princeton University Press.

Vosniadou, S. (2008). International handbook of research on conceptual change. New York: Routledge.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, 43, 337-375.