Why does explaining help learning? Insight from an explanation impairment effect

Joseph Jay Williams (joseph_williams@berkeley.edu)
Tania Lombrozo (lombrozo@berkeley.edu)
Department of Psychology, University of California, Berkeley

Bob Rehder (bob.rehder@nyu.edu)
Department of Psychology, New York University

Abstract

Engaging in explanation, even to oneself, can enhance learning. What underlies this effect? Williams & Lombrozo (in press) propose that explanation exerts *subsumptive constraints* on processing, driving learners to discover underlying patterns. A category-learning experiment demonstrates that explanation can enhance or impair learning depending on whether these constraints match the structure of the material being learned. Explaining can help learning when reliable patterns are present, but actually *impairs* learning when patterns are misleading. This *explanation impairment effect* is predicted by the subsumptive constraints account, but challenges alternative hypotheses according to which explaining helps learning by increasing task engagement through motivation, attention, or processing time. The findings have both theoretical and practical implications for learning and education.

Keywords: explanation; self-explanation; learning; constraints; impairment; category learning

Most teachers and tutors have had the experience of explaining a concept to another person and achieving greater understanding as a result. How does engaging in explanation generate this beneficial effect? This question's importance is underscored by the ubiquity of the phenomenon, and by converging evidence from cognitive science, education, and cognitive development confirming that explanation plays a significant role in learning.

Explanations have been implicated in theories of how conceptual knowledge is represented and how categories are learned (Carey, 1985; Gopnik & Meltzoff, 1997; Murphy & Medin, 1985). Education researchers have demonstrated that explaining has a potent effect on students' learning and fosters deep understanding that allows generalization to novel contexts (Chi et al, 1989; Chi et al, 1994). Research in cognitive development reveals even more profound effects (Wellman & Liu, 2006; Wellman, in press). Prompting children to explain can accelerate conceptual change, such as developing an understanding of number conservation (Siegler, 2002) and false belief (Amsterlaw & Wellman, 2006).

Many extant accounts of explanation's effects have emphasized the metacognitive benefits of explanation, such as prompting learners to identify and fill gaps in their knowledge (Chi et al, 1994; Chi, 2000). Explanations may also focus learners on uncovering the causes that underlie observed outcomes (Wellman & Liu, 2006), or may enhance learning by increasing task engagement in the form of additional motivation, attention, or processing time (for discussion see Siegler, 2002).

Williams and Lombrozo (in press) propose and find empirical support for the *subsumptive constraints* account, according to which explaining exerts constraints on processing that drive people to interpret what they are learning in terms of underlying patterns and regularities. The account is motivated in part by "subsumption" and "unification" theories of explanation from philosophy (Friedman, 1974; Kitcher, 1981), which propose that good explanations show how what is being explained is an instance of a unifying pattern: explanations cite generalizations that *subsume* what is being explained. If the explanations learners generate must satisfy this constraint, explaining will drive learners to reason and construct beliefs in the service of identifying patterns. When useful regularities exist, the subsumptive constraints account predicts positive effects of explanation through the discovery of generalizations. However, this account also predicts that seeking explanations can *impair* learning if there is a mismatch between the subsumptive constraints and the material being learned—for example, in situations in which patterns are nonexistent or misleading.

This paper tests the prediction that such an *explanation impairment effect* exists. Investigating the conditions under which explanation *hurts* learning can inform theories which aim to specify the mechanisms by which explanation *helps* learning, analogous to the study of visual illusions in perception. The conditions under which human perception or cognition succeeds can be less informative than those under which it breaks down and produces errors because the latter serve as a window onto the cognitive machinery underlying perception, in the case of visual illusions, or cognition, in the case of explanation and learning.

In fact, examining explanation's detrimental effects can discriminate the subsumptive constraints account from current theories, which to date have not predicted explanation impairment effects. In particular, a *task engagement* account advocates that engaging in explanation leads learners to be more engaged with the learning task, through increased motivation, attention, or time, which should benefit learning in virtually all contexts. The task engagement account provides an intuitive explanation for the beneficial effects of explaining, positing mechanisms that extend to contexts beyond explanation.

Some studies argue that explanation has effects that go beyond task engagement, showing that its effects surpass control conditions that promote motivation, attention and processing time (e.g. Amsterlaw & Wellman, 2006; Chi et al, 1994; Williams & Lombrozo, in press). However, these studies cannot rule out the possibility that explaining simply engages these mechanisms to a greater degree than the control tasks, highlighting the difficulty of discriminating between competing accounts solely on the basis of explanation's beneficial effects.

Identifying explanation impairment effects is also of clear practical importance, as educators must know when prompted or spontaneous explanation will be detrimental (see also Kuhn & Katz, 2009). Moreover, a deeper understanding of the process by which explaining helps learning can inform educational interventions. If explaining simply boosts students' engagement with the task of learning or increases metacognitive awareness, then it can be expected to produce an 'all-purpose' benefit for learning. But if it helps through more specific mechanisms, such as constraining learners to find underlying principles, then it will be more helpful in some contexts than in others. Its effect may depend on the content being learned, learners' prior knowledge, and other factors.

As in previous work (Williams & Lombrozo, in press) , to investigate explanation our study utilizes category learning, which has been studied extensively and lends itself to carefully controlled artificial materials, permitting rigorous tests of competing accounts. Moreover, previous research supports the idea that explanation can and does play a role in category learning. When learners possess prior knowledge that explains why category features co-occur, they discover patterns underlying category membership and learn to classify items more quickly (Bott & Heit, 2000; Kaplan & Murphy, 2000; Murphy & Allopenna, 1994; Rehder & Ross, 2001; Wattenmaker, Dewey, Murphy, & Medin, 1986). There is also evidence that explanations influence the relative importance of features in learning novel categories (Lombrozo, 2009), and that explaining category membership can influence which features are used in categorization (Chin-Parker et al, 2006). Understanding how explaining influences category learning can thus shed light on the acquisition and representation of conceptual knowledge.

Experiment

Our category learning experiment tested the prediction that explanation can help or hinder learning, depending on the relationship between the material being explained and the subsumptive constraints imposed by explanation. Participants learned about two artificial categories of vehicles by classifying unlabeled items and then receiving feedback on their classification. After feedback and while studying the labeled item, participants in the *explain* condition were prompted to provide an explanation (out loud) for the item's category membership. In contrast, participants in the *classify* condition were free to use any study strategy and simply prompted to share what they were thinking out loud.

The category structures supported at least two bases for categorization, which are illustrated in Table 1 (materials adapted from Kaplan & Murphy, 2000). First, each of the 5 items in each category had a unique color feature. Remembering the 10 idiosyncratic color features always permitted accurate classification of all 10 items. Second, each item contained a feature that was associated with the unifying thematic pattern of jungle vehicles (e.g., drives in jungles, lightly insulated) or arctic vehicles (e.g., drives on glaciers, heavily insulated). In the *reliable pattern* condition, the theme could also be used to perfectly classify 10 out of 10 items based on the presence of an arctic or jungle vehicle feature. However, in the *misleading pattern* condition, the theme led to accurate classification for only 8 out of 10 items, and incorrect classification for the remaining 2 items. The experiment therefore used a 2 (study condition: *explain* vs. *classify*) x 2 (pattern type: *reliable* vs. *misleading*) design.

Dax	Kez
Theme Feature (1)	
Made in Norway	Made in Africa
Has Treads	Has Wheels
Heavily Insulated	Lightly Insulated
Used in Mountain Climbing	Used on Safaris
Drives on Glaciers	Drives in Jungles
Idiosyncratic Color Feature (1)	
Blue	Cyan
Silver	Magenta
Purple	Olive
Red	Maroon
Yellow	Lime
Irrelevant Features (3)	
Two doors/four doors	
Manual transmission/Automatic transmission	
Vinyl seats/Cloth seats	

Table 1. Features associated with each category. Each category item contained one theme feature, one idiosyncratic color feature, and three irrelevant features that were not diagnostic of category membership.

The subsumptive constraints account predicts that engaging in explanation should drive participants to discover and utilize the theme whether it is reliable or misleading, as the theme is more subsuming than the idiosyncratic color features. However, use of the theme should help learning when it is reliable but perpetuate classification errors when it is misleading, thereby *impairing* learning. In contrast, if explanation helps learning by boosting task engagement through increased motivation, attention, or processing time, it should produce a benefit regardless of pattern type.

Method

Participants There were 240 participants (60 in each of four conditions) from the UC Berkeley community who participated for monetary reimbursement or course credit.

Materials Each category was represented by five items, for a total of ten items. Each item was described by a list of five features (see Table 1): one *idiosyncratic color* feature (e.g. blue), one *theme*-related feature from either the arctic vehicle theme (e.g. heavily insulated) or the jungle vehicle theme (e.g. lightly insulated), and three *irrelevant* features that (a) occurred equally often in each category and so were not diagnostic and (b) were unrelated to the arctic/jungle themes (e.g. two doors). The pairing of theme and idiosyncratic color features was randomly chosen in each block of 10 items. The idiosyncratic color features were perfectly predictive of category membership (10 out of 10 items). The theme-related features were perfectly predictive (10 out of 10) in the *reliable* pattern condition, but predictive for only 8 out of 10 items in the *misleading* pattern condition. In each block, a different pair of theme features was randomly chosen to be misleading.

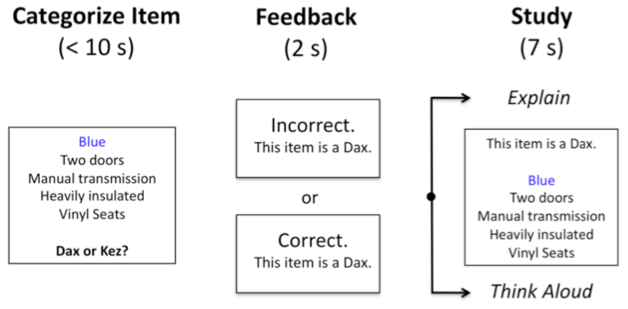

Figure 1: Structure of a single learning trial: item presentation and classification, feedback, and study.

Procedure The *reliable* and *misleading* conditions were run sequentially: data were first collected for 120 participants in the *misleading* pattern condition, then for 120 in the *reliable* condition. The *explain* and *classify* conditions were randomly interleaved, with participants randomly assigned to one or the other study condition. The experiment consisted of learning, test, and explicit report phases.

Learning phase. The structure of a learning trial is shown in Figure 1. On each learning trial an item description was presented as a list of five features, and participants had up to 10 seconds to categorize it as a "Dax" or a "Kez." The idiosyncratic color feature was always displayed on the first line in the color it named (e.g. the feature "red" was shown in red). All other features were presented below it in a random order, and shown in black. Feedback was provided after categorization, and the item was shown with the correct category label for 7 seconds. During this study period, participants in the *explain* condition were prompted (for example) to "Explain why this might be a Dax," and those in the *classify* condition were prompted with: "This item is a Dax. (Remember to say out loud whatever you are thinking.)" In both conditions participants spoke out loud to a voice recorder.

A random ordering of all 10 items constituted a block. Participants completed the experiment when they reached the learning criterion of correctly categorizing all 10 items in a single block, or the maximum of 15 blocks.

Classification test. Each of the 10 *idiosyncratic* and 10 *theme* features was individually presented onscreen and participants categorized it as belonging to a Dax or Kez, rated confidence in their decision (from 1 to 10), and how typical the feature was of its chosen category (1 to 7).[1] Idiosyncratic and theme features were presented in separate, randomly ordered blocks.

Ten *conflict* items were then presented in which an idiosyncratic feature was pitted against a theme feature. Features were paired so that using the idiosyncratic color features to categorize would generate an opposite response to using the theme features.[2]

Explicit report. At the end of the experiment participants were asked what differences might exist between categories and about their strategy for categorization; responses were typed onscreen.

Results

Learning measures, discovered differences between categories, and accuracy in the classification test are shown in Table 2. Significant differences between the *explain* and *classify* conditions are bolded.

	Reliable Pattern		Misleading Pattern	
Measures	Explain	Classify	Explain	Classify
Learning				
Perc. Reaching Criterion	93%	88%	**48%**	**75%**
Mean No. Blocks	6.9	7.9	**11.5**	**10.2**
Discovered differences between categories (from explicit reports)				
Theme Features	**62%**	**43%**	**28%**	**10%**
Color Features	**37%**	**57%**	**45%**	**70%**
Classification test accuracy				
Theme Features	**0.83**	**0.74**	**0.70**	**0.60**
Color Features	0.78	0.83	**0.81**	**0.89**
Conflict Items	**0.40**	**0.55**	**0.63**	**0.83**

Table 2. Measures of learning, discovered differences between categories, and classification test accuracy, as a function of *study condition* and *pattern type*. Significant differences between study conditions are bolded.

[1] These measures mirrored the results on classification accuracy and are not discussed further.

[2] After the classification test in the *reliable* condition, eight *transfer theme* features that were related to the arctic/jungle themes but had not been studied in the learning phase were presented for individual categorization and in *transfer conflict* items. Performance on these items was similar to those with studied theme features and are not discussed further.

Measures of learning. The mean number of blocks to reach the learning criterion is shown in Table 2, and frequency histograms in Figure 2, as a function of *study condition* and *pattern type*. A 2 (study condition*: explain* vs. *classify*) x 2 (pattern type: *reliable* vs. *misleading*) ANOVA on the number of blocks to learn revealed a significant interaction: the effects of explanation differed depending on whether the pattern was reliable or misleading, $F(1, 236) = 6.33$, $p < 0.05$.[3]

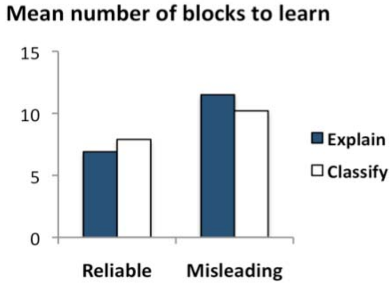

Figure 1: Mean number of blocks to reach the learning criterion of correctly categorizing all 10 items in a block, as a function of *study condition* and *pattern type*. Maximum number of blocks is 15.

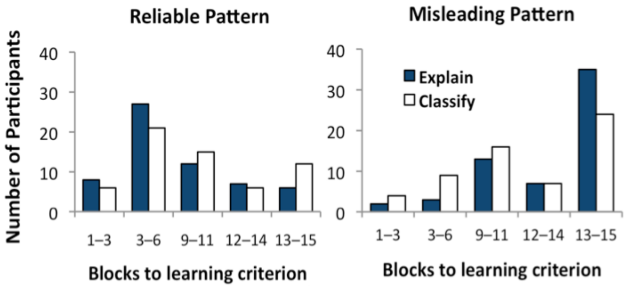

Figure 2: Frequency histogram of the number of blocks to reach learning criterion, as a function of *study condition* and *pattern type*. Bin size is three blocks.

The main effect of *pattern type* was also significant, $F(1, 236) = 44.49$, $p < 0.05$, suggesting that the *misleading* pattern slowed learning, although this interpretation should be qualified because participants were not randomly assigned to these two conditions. When the thematic pattern was reliable, there was a non-significant trend for the *explain* group to learn faster than the *classify* group, $t(118) = 1.43$, $p = 0.16$.[4] When it was misleading, the *explain* group took *longer* to learn, $t(118) = 2.11$, $p < 0.05$. In fact, the number of participants who learned how to classify (reached the learning criterion of correctly categorizing one block of 10 items) was lower in the *explain* condition than the *classify* condition, $\chi 2$ (1) = 5.4, $p < 0.05$. As predicted by the subsumptive constraints account, explanation's effects interacted with the structure of what was being learned, and actually *impaired* learning when a misleading pattern was present.

Discovered differences between categories. To test whether explaining exerted its effects through discovery of the theme, participants' explicit reports about the differences between categories and their categorization strategy were coded for mention of the theme-related and color features (see Fig. 3).[5] Participants in the *explain* condition more often reported theme features as a difference between categories than those in the *classify* condition, whether the pattern was reliable, $\chi 2$ (1) = 4.04, $p < 0.05$, or misleading, $\chi 2$ (1) = 9.79, $p < 0.05$. Participants in the *classify* condition more often reported color features (*reliable* pattern: $\chi 2$ (1) = 4.82, $p < 0.05$; *misleading* pattern: $\chi 2$ (1) = 4.48, $p < 0.05$). Explaining increased learning of theme-related category differences and decreased learning of theme-unrelated (color) features.

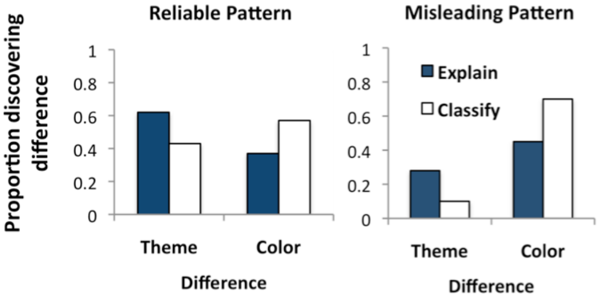

Figure 3: Proportion of participants whose explicit reports revealed discovery that theme and color features differed across categories, as a function of *study condition* and *pattern type*.

Classification test results. Accuracy in classifying theme and color features presented in Figure 4 shows that the *explain* and *classify* groups' different knowledge of theme versus color features also manifested itself in categorization performance. A 2 (study condition: *explain* vs. *classify*) x 2 (feature type: *theme* vs. *color*) repeated measures ANOVA on accuracy revealed a significant interaction for both the *reliable*, $F(1, 118) = 3.96$, $p < 0.05$, and *misleading*, $F(1, 118) = 9.85$, $p < 0.05$, conditions. Participants who explained learned which category the theme features were associated with better than those who classified, with the reverse pattern for color features.

The *conflict* items pitted an idiosyncratic color feature against a theme feature in a categorization decision, and the proportion of items categorized in accordance with the color features was defined as the *conflict score*. This measure was

[3] To address concerns about non-normality, we sorted the number of blocks to learning into five bins of three blocks (as in the histogram in Figure 1) and performed an ordinal regression with *study condition* and *pattern type* as factors. This analysis also found a significant interaction.

[4] To address concerns about non-normality, all *t*-tests reported in this paper were checked with non-parametric Mann-Whitney U tests, which generated the same conclusions.

[5] Agreement between two independent coders was 84% and reported results are for the first coder.

larger for the *classify* condition than the *explain* condition, whether the pattern was reliable, $t(118) = 2.00$, $p < 0.05$, or misleading, $t(118) = 3.42$, $p < 0.05$.

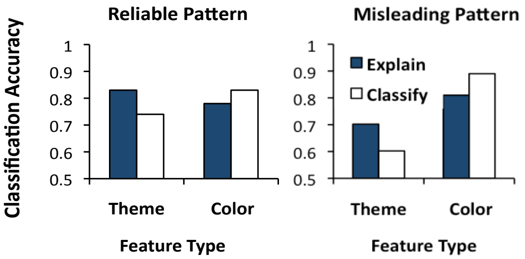

Figure 4: Accuracy in classifying theme and color features, as a function of *study condition* and *pattern type*.

Discussion

While most past research has documented explanation's positive effects, we found an explanation impairment effect: when a misleading pattern was present, explaining category membership impaired learning to categorize. Although counterintuitive, this impairment confirms a prediction of the subsumptive constraints account (Williams & Lombrozo, in press), according to which explaining exerts constraints that drive learners to interpret what they are learning in terms of underlying patterns.

The experiment provides evidence for an interaction between the subsumptive constraints exerted by explanation and the structure of the category. When compared to merely thinking aloud during study, explaining category membership further drove participants to rely on a unifying thematic pattern in categorization rather than use idiosyncratic features, even though both conditions engaged in the demanding task of classification learning with feedback. This produced (nonsignificant) positive consequences for learning when the thematic pattern was reliable, and (significant) negative consequences when it was misleading. Our explanation impairment effect provides evidence against a task engagement account of why explaining helps learning: if explaining merely increases motivation, attention, or processing time, it should not have impaired learning when the pattern was misleading.

A critical reader might have the intuition that the results of this experiment are unsurprising: prompting participants to explain tells them to find a pattern, which helps or harms learning depending on its existence. However, no previous account of explanation and learning has explicitly proposed that explaining constrains people to find patterns or predicted an impairment, instead focusing on metacognitive monitoring, identifying gaps in knowledge, or motivation and attention. A prompt to explain could have made participants attend more to their errors, justify individual categorizations by appeal to the salient and objective color features, or increased motivation to find a reliable basis for categorization. The subsumptive constraints account motivates our specific design and accounts for *why* people feel compelled to seek underlying patterns in response to explanation prompts.

Another criticism could be that this impairment effect is an artifact of an artificial lab task involving a "misleading" theme. However, our goal was precisely to characterize the conditions under which explanation's subsumptive constraints are detrimental. The finding that eliciting explanations impairs learning in any context is novel and consequential for current theories. Moreover, real-world cases involving misleading regularities and suspicious coincidences abound (Griffiths & Tenenbaum, 2007) and provide a promising direction for examining this effect outside the lab. It should be noted that *deeper processing* was not considered under the umbrella of task engagement. We do not see deeper processing as a specific competing account (like motivation) because we interpret the subsumptive constraints account as a specific proposal about the *nature* of the deeper processing explanation evokes.

Evidence for the subsumptive constraints account over the task engagement account has potential implications for education. If explaining does not merely produce an 'all-purpose' enhancement but exerts particular constraints on learning, more research is needed to understand the contexts in which self-explanation interventions are most effective and when they may be detrimental. First, one important question is how the explanation impairment effect varies with the quality of the explanatory pattern, that is, how misleading it is. In our misleading condition, the themes were misleading but only *partially* so: Classifying on the basis of theme features alone could result in moderately good accuracy (80%). The size of the learning impairment may have increased if the themes were even more misleading, but it is equally plausible that it would have *decreased* because subjects might choose to discard use of an explanatory pattern that is yielding poor performance (Murphy & Kaplan, 2000). The extent to which explaining may encourage learners to perseverate on a very low quality explanatory pattern remains to be determined.

Second, it is also important to assess the benefits of explanation *relative* to alternative learning activities, such as elaboration, direct instruction, or analogical comparison, and to examine how their complementary strengths and limitations can be combined. Williams and Lombrozo (in press) found that explaining drove discovery of underlying patterns but resulted in *worse* memory for details than describing. This is problematic because elaborating information in memory and receiving direct instruction may be more valuable at an early stage of learning. The subsumptive constraints account suggests that explaining will not necessarily be useful throughout a study episode (as would be predicted if it promoted task engagement), but will have its strongest effects when learners have already acquired factual background knowledge and need to discover and understand principles that underlie these facts. Successful demonstrations of the self-explanation effect may involve precisely such cases.

Other interesting directions for future research include the role of explanation in generating beliefs about both correct and misconceived underlying principles, in the effects of anomalies in belief revision (Chi, 2000), and in the deployment of prior knowledge (Chi et al, 1994; Williams & Lombrozo, in press). Such research will also be practically important for avoiding classroom manifestations of explanation impairment effects. For example, Kuhn and Katz (2009) suggested that requests for explanations on one task led children to later justify their knowledge of causal relationships by explaining how the relationships could exist, rather than citing observed evidence.

This is the first experiment to examine category learning through classification and feedback with (and without) additional prompts to explain. The learning differences generated by explaining suggest that category learning may involve processes beyond those that reduce immediate classification error. Bott et al. (2007) report that people learned about a thematic pattern underlying category membership (the same used in this experiment) in the *absence* of classification errors – a surprising violation of the classic *blocking effect* – while in the current experiment explaining drove learning about this pattern *despite* classification errors. A deeper understanding of these and other learning phenomena may be gained by considering the contribution of both classification error *and* the construction of knowledge that satisfies the constraints of explanation, whether it is prompted or spontaneous. For example, participants' spontaneous explanations may shed light on how prior knowledge is deployed, and when category learning is driven by explicit rule use versus bottom-up exemplar-based processing that reduces classification error.

The current research emphasizes the importance of subsumptive constraints in explanation's effects on learning, and demonstrates the value of explanation impairment effects for identifying the mechanisms by which explaining enhances learning. We are beginning to explore the relationship between prior knowledge and explanation (Williams & Lombrozo, in press (b)) and expect further investigation, in category learning and other learning contexts, to reveal a complex interaction between the constraints imposed by explanation, prior knowledge, and the structure of what is being explained.

Acknowledgments

We thank the Concepts and Cognition Lab, and in particular Preeti Talwai, for providing feedback on this project, assistance with analyzing data, and preparation of this paper. This work was partially supported by the McDonnell Foundation Collaborative Initiative on Causal Learning, and JJW was supported by an NSERC post-graduate scholarship.

References

Amsterlaw, J., & Wellman, H. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development, 7,* 139-172.

Bott, L., Hoffman, A., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General, 136,* 685-699.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

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

Chi, M.T.H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), Advances in Instructional Psychology, Hillsdale, NJ: Lawrence Erlbaum Associates. 161-238.

Chin-Parker, S., Hernandez, O., & Matens, M. (2006). Explanation in category learning. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1098–1103*)*. Mahwah, NJ: Erlbaum.

Heit, E. & Bott, L. (2000). Knowledge selection in category learning. In D. Medin (Ed.), *Psychology of Learning and Motivation, 39,* 163-199. Academic Press.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 829-846.

Kitcher, P. (1981). Explanatory Unification. *Philosophy of Science, 48*, 507-31.

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

Lombrozo, T. (2009). Explanation and categorization: How "why?" informs "what?". *Cognition, 110*, 248-253.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 904-919.

Murphy, G. L., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A*, 962-982.

Rehder, B. & Ross, B.H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1261-1275.

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 University.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18*, 158-194.

Wellman, H. M. (in press). Reinvigorating explanations for the study of early cognitive development. *Child Development Perspectives.*

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. *Causal learning: psychology, philosophy, and computation*, 261–279.

Williams, J. J., & Lombrozo, T. (in press). The role of explanation in discovery and generalization: evidence from category learning. *Cognitive Science*.

Williams, J. J., & Lombrozo, T. (in press (b)). 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: Cognitive Science Society.