# Explanation-based Mechanisms for Learning: An Interdisciplinary Approach

**Michelene Chi (mtchi@asu.edu)**
Department of Psychology, Arizona State University

**Gerald DeJong (mrebl@illinois.edu)**
Department of Computer Science, University of Illinois at Urbana-Champaign

**Cristine Legare (legare@psy.utexas.edu)**
Department of Psychology, University of Texas At Austin

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

## Abstract

The significant role of explanation in learning and generalization is ubiquitous and well documented: explanation promotes student learning in educational settings, drives conceptual development in young children, is accorded a central role in theories of conceptual representation, and has a long history in artificial intelligence. Despite this, relatively little is known about the precise mechanisms that underlie explanation's effects, and there is a paucity of discourse between the disciplines of cognitive science that study explanation. This interdisciplinary symposium brings together key researchers from education, development, cognitive psychology and computer science to synthesize the progress from these disciplines, forging connections between ongoing research programs to identify promising future directions.

**Keywords:** explanation, self-explanation, understanding, learning, transfer, generalization, explanation-based learning

Investigating the processes that foster learning and generalization is one of the core questions in cognitive science: how does a child, adult, or computer acquire knowledge and understanding that is readily generalized to novel situations, rather than lying inert? Explanation has been repeatedly shown to play a key role across a range of core cognitive science disciplines, including education, cognitive development, cognitive psychology, and artificial intelligence.

In real-world educational contexts, researchers have documented a *self-explanation effect*: whether studying math, physics or biology, learners who generate explanations for what they are learning (spontaneously or in response to prompts) are more likely to acquire accurate concepts and transfer their knowledge to novel problems and situations (Chi et al, 1994). Developmentally, generating explanations for the behavior or reasoning of others has been shown to foster conceptual change in young children's understanding of theory of mind and number conservation (Wellman & Liu, 2006; Siegler, 2002).

Research in cognitive psychology also invokes explanation as a powerful force in conceptual representation and revision, even when explanations are not explicitly solicited. In particular, theories of conceptual representation accord a central role to explanation in understanding how concepts and causal relationships between concepts are represented, especially in knowledge-rich domains (e.g. Murphy & Medin, 1985; Lombrozo, 2009). Understanding how explaining promotes learning can thus shed light on concept acquisition and causal learning, and in particular the role of prior knowledge in these processes.

Finally, work on *explanation-based learning* in artificial intelligence and machine learning has isolated different aspects of explanation and formally characterized how these can support generalization (DeJong & Mooney, 1986; Mitchell et al, 1986). However, the developments in these formalisms have not been fully exploited to interpret current educational and psychological work on explanation and learning, nor been directed at understanding current findings and issues in these fields (for a rare exception, see Ahn, Brewer & Mooney, 1992).

Despite extensive documentation of the powerful effects of explanation in education and development and the relevance of explanation and learning to current research, little is known about why and how explaining exerts its effects (Lombrozo, 2006). This symposium provides a timely forum for addressing this gap in current theories by bringing together researchers to synthesize empirical findings in education and development, theories and models from artificial intelligence, and investigation using the experimental methodology of cognitive psychology.

Michelene Chi synthesizes research in education to give insight into the nature of self-explanation by contrasting it with other learning activities. Cristine Legare provides a developmental perspective on how explanations selectively guide children's exploration and learning, particularly of causal mechanisms. Joseph Williams and Tania Lombrozo provide a novel account of explanation's role in generalization in terms of pattern discovery, drawing on philosophical theories and methodology in cognitive psychology. Gerald DeJong presents current machine learning research on how explanations can integrate statistical and logical inference.

## The constructive nature of self-explanation

Michelene Chi conducted the pioneering work on explanation in education with seminal papers on the self-explanation effect. Her talk reviews and synthesizes research in education to shed light on why self-explanation is beneficial, through a comparison to other active, constructive, and interactive learning activities (Chi, 2009). The constructive nature of explanation is proposed as an

---

* Denotes organizers.

important factor and provides insight into when explanation will be most effective in educational settings.

## In search of a mechanism: The unique and selective benefits of explanation for learning

Cristine Legare reviews evidence that children preferentially explain events that have the potential to teach them something new: inconsistency with prior knowledge triggers children's explanatory reasoning by motivating a search for underlying causal mechanisms (Legare, Gelman, & Wellman, 2010) and explaining inconsistent outcomes guides exploratory, hypothesis-testing behavior (Legare, in press). She presents recent developmental research comparing explanation to other cognitive processes such as exploration and observation (Legare & Lombrozo, under review) to provide additional evidence that identifying causal mechanisms is a crucial function of explanation and provides support for the proposal that explanation has unique and selective benefits for learning.

## A subsumptive constraints account of explanation's role in generalization

Understanding what explanations are – their structure and content – can shed light on why explaining supports generalization. Williams & Lombrozo (2010) draw on theories from philosophy of science to propose a *subsumptive constraints* account, according to which explaining exerts the selective constraint of driving learners to find patterns or regularities that underlie or generate (*subsume*) what they are explaining. Experiments on category learning demonstrate the first self-explanation effect using rigorously controlled artificial materials from cognitive psychology and provide direct evidence that explaining "Why?" drives the discovery of abstract regularities, which then provide the basis for generalization to novel contexts. The double-edged nature of this constraint is demonstrated in an *explanation impairment effect*: if people seek explanations when only misleading regularities are present, explaining impairs their learning. This provides evidence against a primarily motivational or attentional account of explanation's effects.

## A novel perspective on explanation-based learning

Gerald DeJong has been an early architect of and major contributor to machine learning and artificial intelligence research on explanation-based learning, from its earliest stages to the present. His talk and current work (DeJong, 2006) explore how explanations may serve as a mechanism for integrating the strengths of statistical inference – robust induction despite the uncertainty in empirical observations – and symbolic and logical reasoning – the representational and inferential richness obtained from structured representations supplied by prior or expert knowledge.

The goal of this symposium is to shed light on the powerful impact of explanation on learning and generalization by bringing together leading researchers from education, cognitive development, cognitive psychology and machine learning and AI. Four talks and an extended discussion period by the full panel of presenters (moderated by the discussant Tania Lombrozo) will synthesize recent advances in the empirical discoveries and theoretical models in each discipline, promote cross-fertilization of pertinent ideas, and identify promising directions for future research.

## References

Ahn, W., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 391-412.

Chi, M.T.H. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73-105.

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

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

DeJong, G., & Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine learning*, *1*(2), 145–176.

DeJong, G. (2006). Toward robust real-world inference: A new perspective on explanation-based learning. *Proceedings of the Seventeenth European Conference on Machine Learning*, (pp. 102-113).

Legare, C.H. (in press). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*.

Legare, C.H., Gelman, S.A., & Wellman, H.W. (in press). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development*.

Legare, C.H. & Lombrozo, T. (under review). The unique and selective benefits of explanation for learning in early childhood.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10,* 464-470.

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

Mitchell, T. M., Keller, R. M., & Kedar-Cabelli, S. T. (1986). Explanation-based generalization: A unifying view. *Machine learning*, *1*(1), 47–80.

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

Rozenblit, L.R. & Keil, F.C. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science, 26,* 521-562.

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.

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik, & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 261-279). Oxford: Oxford University Press.

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