Cognitive Psychology 66 (2013) 55–84

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Explanation and prior knowledge interact to guide learning

Joseph J. Williams *, Tania Lombrozo

Department of Psychology, University of California at Berkeley, United States

ARTICLE INFO

Article history:
Accepted 19 September 2012
Available online 23 October 2012

Keywords:
Explanation
Self-explanation
Prior knowledge
Learning
Generalization
Category learning

ABSTRACT

How do explaining and prior knowledge contribute to learning? Four experiments explored the relationship between explanation and prior knowledge in category learning. The experiments independently manipulated whether participants were prompted to explain the category membership of study observations and whether category labels were informative in allowing participants to relate prior knowledge to patterns underlying category membership. The experiments revealed a superadditive interaction between explanation and informative labels, with explainers who received informative labels most likely to discover (Experiments 1 and 2) and generalize (Experiments 3 and 4) a pattern consistent with prior knowledge. However, explainers were no more likely than controls to discover multiple patterns (Experiments 1 and 2), indicating that effects of explanation are relatively targeted. We suggest that explanation recruits prior knowledge to assess whether candidate patterns are likely to have broad scope (i.e., to generalize within and beyond study observations). This interpretation is supported by the finding that effects of explanation on prior knowledge were attenuated when learners believed prior knowledge was irrelevant to generalizing category membership (Experiment 4). This research provides evidence that explanation can serve as a mechanism for deploying prior knowledge to assess the scope of observed patterns.

1. Introduction

Children, adults, and students of all ages face the common challenge of discovering useful information and then generalizing it to novel contexts. While learning and generalization engage a variety of

* Corresponding author. Address: Department of Psychology, University of California at Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, United States. Fax: +1 510 642 5293.
E-mail address: joseph_williams@berkeley.edu (J.J. Williams).

0010-0285/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cogpsych.2012.09.002

cognitive processes, researchers across several fields have recognized an important role for explanation (Lombrozo, 2012). For example, prompting young children to explain observations that challenge their intuitive theories can accelerate conceptual development (e.g., Amsterlaw & Wellman, 2006; Siegler, 1995; Wellman & Liu, 2007), and prompting students to explain why a fact is true or why a solution to a problem is correct can improve both learning and transfer to novel problems (e.g., Chi, de Leeuw, Chiu, & LaVancher, 1994; Fonseca & Chi, 2011). How and why does explaining have these effects? In particular, how does explaining guide discovery and generalization?

We propose that explaining recruits a set of criteria for what constitutes a good explanation, and that these criteria in turn act as constraints on learning and generalization (Lombrozo, 2012). For example, explanations are typically judged better if they are simple (Lombrozo, 2007; Read & Marcus-Newhall, 1993) and have what we refer to as broad *scope* – appealing to features, principles, or patterns that accurately apply to numerous instances across a range of contexts (Pennington & Hastie, 1992; Preston & Epley, 2005; Read & Marcus-Newhall, 1993). In this paper we focus on scope to consider whether the act of generating explanations makes learners more likely to discover and generalize patterns with broad scope. For example, in trying to explain why peafowl at the zoo vary in color, one might discover that males (peacocks) tend to be colorful while females (peahens) tend to be drab. This discovery and the reasoning behind it could in turn support inferences about unobserved peafowl, such as the generalization that all male and female peafowl are likely to conform to this pattern, and not just the particular species observed at the zoo.

The idea that explaining makes learners more sensitive to scope predicts that explaining should increase the extent to which learners consult prior knowledge.[1] Learning poses a challenging inductive problem, and prior knowledge can serve as an important cue to which patterns are likely to have broad scope. For example, an explanation for variation in peafowl coloration that appeals to a generalization over sex (males versus females) could be preferred over one formulated over size (larger versus smaller) because prior knowledge favors the former as more likely to generalize beyond the peahen sample observed. So if explaining changes the criteria that learners adopt in generating or evaluating hypotheses by leading them to privilege patterns with broad scope, then explaining should recruit prior knowledge in evaluating the scope of candidate patterns. In addition to testing this prediction, we consider whether such an effect (if found) results from a special relationship between explanation and prior knowledge or instead from a more general effect, such as a global increase in how much information explainers discover and retain.

By focusing on the relationship between explanation and prior knowledge, we gain unique leverage in addressing two important questions in cognitive science: how explanation impacts learning and generalization, and when and how prior knowledge is brought to bear on learning. In addition to bridging research on explanation and prior knowledge, we bridge two research traditions by examining questions about explanation and learning (typically studied by educational psychologists) in the context of artificial category learning (typically studied by cognitive psychologists). In the remainder of the introduction we briefly review past work from each of these traditions before presenting the key theory, questions, and predictions that motivate the four experiments that follow.

1.1. General and selective effects of explanation on learning

Research in education has investigated the role of explanation in learning in the context of the "self-explanation effect": the phenomenon whereby explaining, even to oneself, can improve learning. Effects of self-explanation have been documented in domains from biology to mathematics, from elementary school through university, and under a variety of methods for eliciting explanations (e.g., Aleven & Koedinger, 2002; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi et al., 1994; Crowley & Siegler, 1999; Graesser, Singer, & Trabasso, 1994; Nokes, Hausmann, VanLehn, & Gershman, 2011; Renkl, 1997; Rittle-Johnson, 2006; Siegler, 2002). This diversity is matched by a wide range of proposals concerning how explanation affects learning. For example, a prompt to explain could encourage the

[1] Throughout the paper we use the term "prior knowledge" to indicate a learner's beliefs or commitments, whether or not they are true. That is, our use of the term "knowledge" is non-factive.