Effects of Explaining Anomalies on the Generation and Evaluation of Hypotheses

Joseph Jay Williams (joseph_williams@berkeley.edu), Caren M. Walker (caren.walker@berkeley.edu)
Samuel G. Maldonado (samuel.g.maldonado@gmail.com), Tania Lombrozo (lombrozo@berkeley.edu)
Department of Psychology, University of California, Berkeley, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract

We investigate the effects of explaining anomalies (i.e., observations that conflict with current beliefs) on belief revision, and in particular how explaining contributes to the rejection of incorrect hypotheses, the generation of alternative hypotheses, and the selection of a hypothesis that can account for anomalous observations. Participants learned how to rank students across courses using statistical concepts of deviation, and did so while either explaining sample rankings or writing their thoughts during study. We additionally varied whether or not candidate hypotheses about the basis for ranking were presented to participants prior to learning, and the number of sample rankings that violated intuitive misconceptions about ranking. Measures of learning and coded responses suggest that prompting people to explain can increase the rate at which they entertain both correct and incorrect hypotheses, but that explaining promotes the selection of a hypothesis that can account for anomalous observations.

Keywords: explanation; self-explanation; learning; generalization; statistics; misconceptions; anomalies.

Introduction

A critical element of successful learning is the ability to flexibly revise beliefs in light of new data and experience. For example, a mathematics student might form tentative beliefs about how to solve a novel problem, but subsequently revise these beliefs in the face of anomalous data: observations that conflict with working assumptions and therefore signal a need to revise beliefs (Chinn & Brewer, 1993; Koslowski, 1996). Here we consider how beliefs are revised in light of anomalous observations, and in particular how *explaining* such observations influences learning.

Generating explanations has been shown to promote learning across a range of tasks and domains, with evidence from experimental studies of category learning (Williams & Lombrozo, 2010), "self-explaining" in students (e.g., Fonseca & Chi, 2011), and conceptual development in children (e.g., Siegler, 2002; Wellman & Liu, 2007). These benefits are likely to derive from multiple sources, including increased engagement and increased accessibility of effective strategies (Siegler, 2002), better metacognitive monitoring (e.g., Chi et al, 1994), and the generation of inferences to fill gaps in understanding (e.g., Chi et al, 2000), among others.

In the present work we build on the *Subsumptive Constraints* account of explanation developed in prior research (Williams & Lombrozo, 2010, 2013). According to this account, explaining a particular observation drives learners to interpret it as an instance of a broad pattern or generalization, and thereby facilitates learning about regularities that apply broadly (Williams & Lombrozo, 2010; 2013; Williams, Lombrozo & Rehder, 2013).

To illustrate, consider the findings reported in Williams and Lombrozo (2010). Participants attempted to learn a new classification system involving two categories that could be differentiated by a rule with no exceptions (100% rule) or an alternative that accounted for most cases, but with two anomalies (75% rule). Participants who were prompted to explain were significantly more likely to discover the 100% rule than those prompted to *describe* the category members, *think aloud*, or engage in *free study*. These findings confirm the prediction that explaining facilitates learning about broad patterns, and also suggest that explaining could make learners especially sensitive to anomalies, as they signal that current beliefs are either false or limited in scope.

Subsequent research, however, suggests a more complicated relationship between explanation and anomalies. Williams and Lombrozo (2013) found that participants prompted to explain favored patterns consistent with prior knowledge, even when such patterns had exceptions (anomalies) that were better explained by alternative patterns. Williams, Lombrozo, and Rehder (2013) found that participants prompted to explain were more likely to *over*generalize broad patterns, effectively ignoring exceptions, even when this resulted in slower and less accurate learning.

Explaining seems to therefore have opposite effects: by encouraging learners to seek broad patterns, explaining can sometimes lead to greater belief revision in light of anomalies, and at other times to the anomalies being effectively dismissed or "explained away" (see also Chinn & Brewer, 1993; Khemlani & Johnson-Laird, 2012; Koslowski, 1996). As a first step towards understanding the conditions under which explanation has each effect, Williams, Walker, and Lombrozo (2012) investigated how changing the *number* of anomalous observations presented interacted with a prompt to explain. We begin by briefly reviewing the results from this study, and additionally present novel analyses concerning participants' coded explanations. We then present a new experiment aimed at differentiating two potential roles for explanation: one in the *rejection* of current hypotheses in light of anomalies, and another in the *generation* and *selection* of new hypotheses.

Explaining anomalies: Previous findings

In previous work, we explored the effects of generating explanations for observations that were anomalous with respect to learners' prior beliefs about statistical measures (Williams, Walker, & Lombrozo, 2012). Participants learned a university's ranking system by studying how pairs of students from different courses had been ranked given the students' grades and the means and standard deviations of their respective courses. The task required learners to compare student grades using concepts analogous to z-scores, and therefore to reject commonly endorsed but *non-normative* principles for ranking. These *non-normative* principles included ranking students based on the higher raw score, the greater number of points above the course mean, or closeness to the maximum course score (Schwartz & Martin, 2004).

A realistic and experimentally useful feature of this task was that participants could encounter ranked student pairs that were either *consistent* or *anomalous* with respect to the non-normative principles for ranking. In many ranked student pairs, the student who is a greater number of standard deviations above the mean will also have a higher raw score, be farther from the mean, or closest to the maximum. We call sample rankings that are consistent with all of the identified ranking principles *consistent* items, and those that are only consistent with the use of z-scores *anomalous* items because they are anomalous with respect to many participants' prior beliefs (see Fig. 1).

Williams, Walker, and Lombrozo (2012), henceforth WWL12, presented participants with five examples of ranked pairs of students to learn a university's method for ranking students. Participants' study task was either to *explain* why the higher ranked student was ranked higher, or to *write thoughts* they had while studying the pair. Of the five example pairs, there was either a *single anomaly* (and four consistent pairs) or *multiple anomalies* (four anomalies, one consistent pair). WWL12 found that belief revision was greatest when participants explained *and* received multiple anomalies. Explaining did not promote belief revision when only a single anomaly was presented, and multiple anomalies had no effect on learning unless participants explained.

While these findings suggest that explaining may be especially potent for ensuring that learners process anomalies and use them in updating beliefs, there are several reasons why explaining might have this effect.

Explaining anomalies could be improving accuracy by increasing the *rejection* of the non-normative principles that were inconsistent with the anomalies, or by increasing the *generation and selection* of the normative principle. Either of these would account for the observed belief revision, and in fact, the effects could be due to a combination of both.

To evaluate these possibilities, we report here the results of coding the written responses that participants provided in the *explain* and *write thoughts* conditions. We coded for whether participants mentioned any of the non-normative principles and whether they identified standard deviation as playing an important role in rankings.

Verbal Response Coding

Each of the five written responses participants provided during the study phase of the experiment was coded according to the following criteria: whether a response mentioned a *non-normative* principle, whether it mentioned the *relative-to-deviation* principle (i.e., standardized z-scores, whether or not participants used technical terminology to convey the idea), and whether it contained some *other* response, such as expressions of surprise or confusion, disagreement with the ranking, or mention of other features of the pairs.

Non-Normative Principles The three *non-normative* principles were incorrect but designed to correspond to intuitive statistical misconceptions. We term the principles (1) *raw-score*: the higher ranking went to the student with the higher score, irrespective of mean, average deviation, and minimum or maximum score; (2) *relative-to-average*: the higher ranking went to the student whose score was the farthest above (or least

(a) Sarah got 85% in a Sociology class, where the average score was 79%, the average deviation was 3%, the minimum score was 67%, and the maximum score was 90%.

Tom got 69% in a Art History class, where the average score was 65%, the average deviation was 8%, the minimum score was 42%, and the maximum score was 87%.

Sarah was ranked more highly by the university than Tom.

(b) Sarah got 85% in a Sociology class, where the average score was 79%, the average deviation was ***8%***, the minimum score was 67%, and the maximum score was 90%.

Tom got 69% in a Art History class, where the average score was 65%, the average deviation was ***3%***, the minimum score was 42%, and the maximum score was 87%.

Tom was ranked more highly by the university than ***Sarah***.

Figure 1: **(a)** A *consistent* ranked example for which all four principles predicted the same ranking. **(b)** An *anomalous* ranked example constructed by switching the class average deviations of the consistent example from (a). The switch means that the correct relative-to-deviation ranking is now the opposite of what is predicted by the raw-score, relative-to-average, and relative-to-highest-score principles. Emphasis is added for illustration and was not provided to participants.

below) the class's mean score; (3) *relative-to-highest-score*: the higher ranking went to the student whose score was the closest to the highest score achieved in the class.

Relative-to-Deviation Principle According to this principle, the better student was the one who scored a greater number of standard (average) deviations above the mean (see Schwartz & Martin, 2004; Belenky & Nokes-Malach, 2012). This was calculated as the difference from the mean divided by the average deviation, and is closely related to a normative measure like the z-score.

Response Coding Results

Principles Cited A *task (2: explain, write thoughts)* x *number of anomalies (2: single, multiple)* x *principle type* (*non-normative, relative-to-deviation)* mixed ANOVA was conducted on the proportion of responses that mentioned each type of principle (see Fig. 2).

This analysis revealed main effects of *task*, $F(1, 272) = 43.98$, $p < 0.001$, $\eta_p^2 = 0.14$, and *number of anomalies*, $F(1, 272) = 37.15$, $p < 0.001$, $\eta_p^2 = 0.12$. Overall, explaining *increased* mention of principles, while multiple anomalies led to *decreased* mention of principles.

There was also a main effect of *principle type*, $F(1, 272) = 49.90$, $p < 0.001$, $\eta_p^2 = 0.16$, with non-normative principles mentioned more frequently than the relative-to-deviation principle. However, this effect was qualified by an interaction between *number of anomalies* and *principle type*, $F(1, 272) = 40.52$, $p < 0.001$, $\eta_p^2 = 0.13$. We therefore conducted separate *task* x *number of anomalies* ANOVAs for the two principle types.

Non-normative principles were cited *more often* by participants prompted to explain, $F(1, 272) = 19.03$, $p < 0.001$, $\eta_p^2 = 0.07$, and *less often* by those who encountered multiple anomalies, $F(1, 272) = 96.49$, $p < 0.001$, $\eta_p^2 = 0.26$, with no interaction.

The relative-to-deviation principle was also cited more often in the explain condition, $F(1, 272) = 13.14$, $p < 0.001$, $\eta_p^2 = 0.05$, with no significant effect of the number of anomalies, $F(1, 272) = 1.89$, $p = 0.17$, $\eta_p^2 = 0.01$.

Number of Different Principles Cited A *task* x *number of anomalies* ANOVA was performed on the mean number of different principles cited by each participant (see Fig. 3). Participants prompted to explain mentioned a greater number of different principles, $F(1, 272) = 16.20$, $p < 0.001$, $\eta_p^2 = 0.06$, and multiple anomalies resulted in mention of *fewer* different principles, $F(1, 272) = 31.36$, $p < 0.001$, $\eta_p^2 = 0.10$. There was also a *task* x *number of anomalies* interaction: explaining robustly increased the number of different principles mentioned in the multiple anomalies condition, $t(125) = 3.97$, $p < 0.001$, $d = 0.70$), while the effect in the single anomalies condition was not significant, $t(147) = 1.55$, $p = 0.12$, $d = 0.25$.

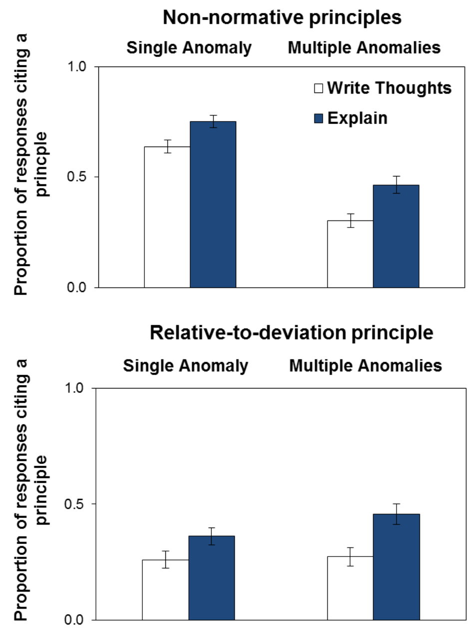

Figure 2: Data from WWL12: Mean proportion of responses citing either a non-normative principle (upper panel) or the relative-to-deviation principle (lower panel).

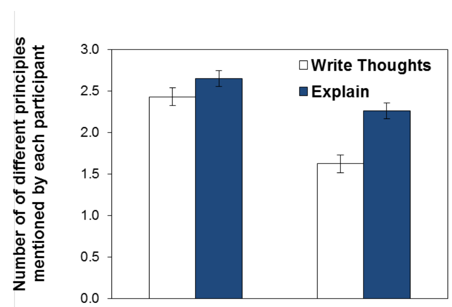

Figure 3: Data from WWL12: Mean number of different principles mentioned by each participant.

Summary and Discussion

The results of coding responses from WWL12 suggest that the effects of explanation on learning are not principally a consequence of *rejecting* principles in light of anomalies, at least not in this kind of task. Explaining increased the rate at which participants mentioned the correct relative-to-deviation principle, but also how often participants mentioned *non-normative* principles, and how many different principles were cited. Instead, it appears that explanation played an important role in the generation of multiple hypotheses and the selection of the correct hypothesis from among them.

We now present a new experiment that aims to better understand the role of explanation in *generating* the correct hypothesis as opposed to *evaluating and selecting* the correct hypothesis from among candidates. In order to do so, we replicate the basic design of WWL12 with an additional manipulation: whether or not participants are presented with a fixed set of candidate hypotheses, including the relative-to-deviation principle, prior to learning.

Experiment

Our experiment manipulated whether participants were asked to explicitly consider potential ranking principles before engaging in learning. Specifically, participants in the *exposure* condition were presented with descriptions of five candidate principles and rated their plausibility. Participants in a *no exposure* condition completed the task without this initial presentation of candidate hypotheses, effectively replicating WWL12 (see Bonawitz & Griffiths, 2010, for a similar manipulation).

As in WWL12, we additionally varied whether participants received instructions to *explain* or to *write thoughts*, and whether they encountered a *single anomaly* or *multiple anomalies* during study.

If the main role of explanation in WWL12 was to facilitate the *generation* of candidate hypotheses – and therefore of the relative-to-deviation principle – then the exposure manipulation should mimic effects of explanation in the write thoughts condition, and potentially eliminate differences across study conditions. In contrast, if explaining principally or additionally plays a role in the evaluation and selection of the correct hypothesis (i.e., the relative-to-deviation principle, which accounts for all observations), then we should observe effects of explanation even in the *exposure* condition.

Methods

Participants Seven-hundred-and-twenty-seven members of the Amazon Mechanical Turk community participated in exchange for monetary compensation. Four-hundred-and-eighty additional participants were excluded for failing an instructional manipulation check adapted from Oppenheimer et al. (2009) and designed to evaluate whether participants were reading instructions. The number of excluded participants did not differ as a function of condition, all *p*s > 0.10.

Materials & Procedure The materials and procedure mirrored WWL12, except as noted.

Pre-Test. Participants were presented with ten unranked student pairs and judged how likely the university would be to rank one student above another, on a nine point scale ranging from "Definitely student [X]" to "Definitely student [Y]," with a midpoint of "Equally Likely."

Unlike WWL12, six pre- and post-test items pitted the relative-to-deviation principle against a single one of the non-normative principles, with the other two non-normative principles predicting that the students were equally ranked. Of the ten pairs, two pitted the relative-to-deviation principle against the raw-score principle; two against the relative-to-average; and two against the close-to-highest-score. Four pairs were like the anomalous study pairs in pitting the relative-to-deviation principle against all three non-normative principles.

Pre-Exposure to Principles. In the *exposure* condition, after the pre-test and before the study phase, participants were shown an example pair of students and told who was ranked higher. This ranking was *consistent*, similar to the example in Figure 1a. Participants were then presented with five potential rules the university could use to rank students, and asked to judge, on a scale from 1-7, how likely it was that the university used that particular rule. The rules included all four principles discussed above, as well as an additional *average-plus-deviation* principle[1], which favored whichever student was the greater number of percentage points above the average plus average deviation.

Study. Each of the five ranked examples was presented onscreen for exactly 90 seconds in a format similar to Figure 1a and 1b. Participants in the *explain* condition were prompted to explain why the higher-ranked student was ranked more highly, typing their explanation into a text box onscreen. Participants in the *write thoughts* control condition were told to type their thoughts during study into an equivalent text box.

Post-Test. The post-test was identical to the pre-test, but all student names and grades were changed, with five points added to each grade to generate novel numbers while preserving the way in which the items pitted the principles against each other.

Additional Measures. Additional questions were asked at the end of the experiment (e.g., demographics, sufficient time for task, strategy) but are not discussed here in the interest of space.

Results

Learning Pre-test accuracy did not differ significantly as a function of condition (all *p*s > 0.2, mean = -.90); we subsequently consider the *change* in pre- to post-test accuracy as our measure of learning.

A *task* x *number of anomalies* x *exposure* ANOVA on the pre- to post- test change in accuracy found main effects of explanation, $F(1, 719) = 15.06$, $p < 0.001$, $\eta_p^2 = 0.02$, and number of anomalies, $F(1, 719) = 29.59$, $p < 0.001$, $\eta_p^2 = 0.04$, with no significant effect of exposure, $F(1, 719) = 1.81$, $p = 0.18$, $\eta_p^2 < 0.01$, nor interactions (see Fig. 4). Participants prompted to explain showed greater learning than those who were not so prompted (whether they observed one or multiple anomalies), and

[1] We thank Daniel Belenky (personal communication) for suggesting this as an additional principle that participants might find compelling and spontaneously employ.

participants who saw multiple anomalies learned more as well (whether or not they explained).

Principles Cited To analyze the relative frequencies with which non-normative and normative principles were cited in each response, we conducted a repeated-measures ANOVA with principle type (non-normative principle, relative-to-deviation principle) as a within-subjects factor and *task* (2), *number of anomalies* (2), and *exposure* (2) as between-subjects factors. This analysis revealed a four-way interaction, $F(1, 717) = 8.01$, $p < 0.01$, $\eta_p^2 = 0.01$. We therefore conducted separate *task* x *exposure* x *principle* ANOVAs for the single anomaly and multiple anomalies conditions.

In the single anomaly condition, this analysis revealed that explaining promoted overall mention of principles, $F(1, 448) = 13.44$, $p < 0.001$, $\eta_p^2 = 0.03$, and that non-normative principles were mentioned more frequently than the relative-to-deviation principle, $F(1, 448) = 67.30$, $p < 0.001$, $\eta_p^2 = 0.13$. There were no other significant effects – in particular, the effect of exposure was not significant, $F(1, 448) < 0.01$, $p = 0.99$, $\eta_p^2 < 0.01$.

In the multiple anomalies condition, there was a *task* x *exposure* x *principle* interaction, $F(1, 269) = 5.28$, $p < 0.05$, $\eta_p^2 = 0.02$. Participants who explained and were exposed to the hypotheses beforehand were more likely to mention the relative-to-deviation principle over the non-normative principles, relative to those who explained without exposure. A *task* x *exposure* ANOVA for just non-normative principles revealed a main effect of explaining, $F(1, 269) = 5.66$, $p < 0.05$, $\eta_p^2 = 0.02$, and a significant interaction, $F(1, 269) = 7.76$, $p < 0.05$, $\eta_p^2 = 0.03$. For the relative-to-deviation principle, there was only a main effect of explaining. $F(1, 269) = 10.03$, $p < 0.01$, $\eta_p^2 = 0.04$.

Number of Different Principles Cited The average number of different principles cited was analyzed with a 2 (task) by 2 (number of anomalies) by 2 (exposure) ANOVA, which revealed more principles in the *explain* condition than the *write thoughts* condition, $F(1, 712) = 26.73$, $p < 0.001$, $\eta_p^2 = 0.04$, with a marginal effect of exposure, $F(1, 712) = 2.81$, $p = 0.09$, $\eta_p^2 < 0.01$. There was also an interaction between task and exposure, $F(2, 712) = 4.51$, $p < 0.05$, $\eta_p^2 = 0.01$: explanation's boost in number of principles cited was considerably attenuated when participants were exposed to the principles before study. This finding suggests that participants did attend to the *exposure* task, even though it did not affect learning.

Relationship Between Coded Responses and Learning To investigate the relationship between participants' responses to the explain and write thoughts prompts and their learning as reflected on the post-test, we examined correlations and partial correlations between response types and accuracy. The largest contributor to post-test accuracy was the proportion of responses citing the relative-deviation-principle, $r(725) = 0.60$, $p < 0.001$, followed by the *negative* effect of the proportion of responses citing the non-normative principles, $r(725) = -0.38$, $p < 0.05$. Even conditioning on pre-test accuracy, task, number of anomalies, and exposure, post-test accuracy was positively correlated with citing the

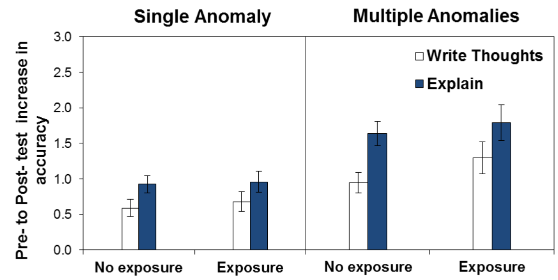

Figure 4: Change in accuracy from pre- to post-test.

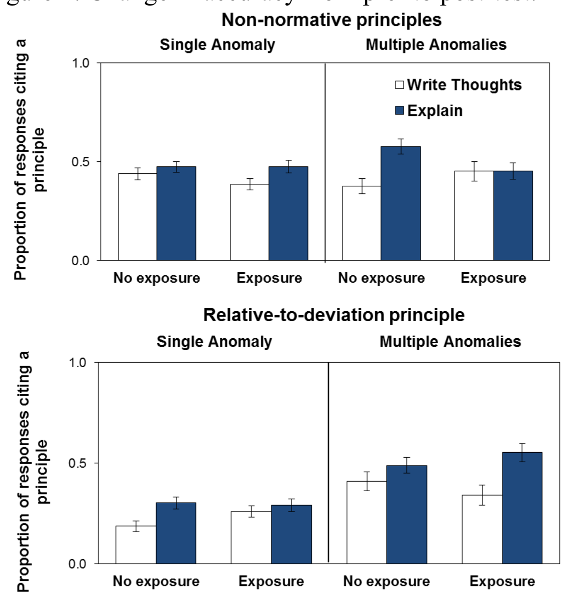

Figure 5: Mention of non-normative principles and the relative-to-deviation principle (per-response).

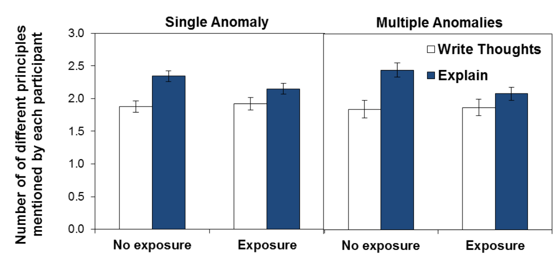

Figure 6: Number of different principles mentioned by each participant.

relative-to-deviation principle, $r(640) = 0.45$, $p < 0.001$, and negatively correlated with citing non-normative principles, $r(640) = -0.26$, $p < 0.001$. These findings suggest that coded responses reflected learning, and are at least consistent with the stronger claim that producing the responses was itself a causal factor in driving learning.

Discussion

The current study found that participants who were prompted to explain reliably outperformed those in a *write thoughts* control condition when it came to learning how a university ranked students, a task that required some understanding of population variance or deviation. Although the current data suggest a trend for a larger effect of explanation in the multiple (vs. single) anomaly condition, the interaction was not significant, as it was in WWL12, where explanation facilitated belief revision significantly more when there were multiple anomalies rather than a single anomaly. With respect to one of the main issues that motivated this research – i.e., specifying the conditions under which explaining leads to greater versus less belief revision – our findings are therefore inconclusive.

Nonetheless, the current work provides novel data from participants' coded responses to the explain and write thoughts prompts, which shed light on the role of explanation in rejecting incorrect hypotheses, generating candidate hypotheses, and selecting the correct hypothesis. If it were the case that explaining anomalous observations made learners more likely to reject hypotheses that failed to account for those observations, then we might have expected that prompting participants to explain would lead them to mention non-normative principles less often than participants in the control condition..Instead, we found that participants prompted to explain were *more likely* to produce non-normative principles, and also more likely to produce a larger number of different principles. This result – found in WWL12 and replicated again here – suggests that explanation instead played a role in the generation and selection of the correct hypothesis concerning ranking.

Our new experiment helped isolate effects of explanation due to hypothesis generation from those of hypothesis selection. We found that "generating" candidate hypotheses for learners did not mimic effects of explanation; explanation improved learning even when candidate hypotheses were provided in both study tasks. This finding suggests that explaining may be playing an important role in the comprehension or *selection* of the correct hypothesis (see also Siegler, 2002).

Acknowledgments

This research was supported by an NSF CAREER grant awarded to TL. (DRL-1056712).

References

Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences*, *21*(3), 399-432.

Bonawitz, E. B., & Griffiths, T. L. (2010). Deconfounding hypothesis generation and evaluation in Bayesian models. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2260-2265).

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

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

Fonseca, B. & Chi, M.T.H. (2011). The self-explanation effect: A constructive learning activity. In Mayer, R. & Alexander, P. (Eds.), *The Handbook of Research on Learning and Instruction* (pp. 270-321). New York, USA: Routledge Press.

Khemlani, S. & Johnson-Laird, P. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica, 139,* 486-491.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. The MIT Press.

Legare, C.H. (2012). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development*.

Lipton, Peter (2004). Inference to the Best Explanation. New York, NY: Routledge.

Oppenheimer, D. M., Meyvisb, T., & Davidenkoc, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867-872.

Schwartz, D.L., & Martin, T. (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction*, 22(2), 129-184.

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. *Causal learning: psychology, philosophy, and computation*, 261–279.

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

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology, 66,* 55-84.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*. Advance online publication.

Williams, J. J., Walker, C. M., & Lombrozo, T. (2012). Explaining increases belief revision in the face of (many) anomalies. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society (pp. 1149-1154). Austin, TX: Cognitive Science Society.