Explanations and Causal Judgments are Differentially Sensitive to Covariation and Mechanism Information

Nadya Vasilyeva (Vasilyeva@Berkeley.Edu)
Tania Lombrozo (Lombrozo@Berkeley.Edu)
Department of Psychology, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract

We report four experiments demonstrating that judgments of explanatory goodness are sensitive both to covariation evidence and to mechanism information. Compared to judgments of causal strength, explanatory judgments tend to be *more* sensitive to mechanism and *less* sensitive to covariation. Judgments of understanding tracked covariation least closely. We discuss implications of our findings for theories of explanation, understanding and causal attribution.

Keywords: explanation; covariation; mechanism; causal strength; understanding

Suppose you work for an art museum, where you're tasked with tracking museum statistics, including which visitors visit each gallery, and which visitors make optional donations to the museum in the extra donation box near the exit. In organizing your data, you stumble upon a correlation between two factors: museum visitors who visit the portrait gallery are much more likely to make an optional donation than those who do not. How do you make sense of this relationship? Does the visit to the portrait gallery *explain* why some visitors make donations? Are you persuaded there's a strong *causal relationship* between visiting the portrait gallery and making a donation?

In answering these questions, at least two pieces of additional information may be relevant. First, how strong is the correlation between visiting the portrait gallery and making a donation? If the covariation evidence suggests a perfect association you'll likely respond differently from a case in which the association is weak. Second, is there a plausible mechanism linking the candidate cause to the effect? On the face of it the answer may be "no," but suppose you learn of research in social psychology that exposure to faces (and particularly to eyes) triggers mechanisms associated with the maintenance of a pro-social reputation, increasing cooperative behavior (Bateson, Nettle, & Roberts, 2006). Would this alter your response?

Decades of research on causal learning have pinpointed both covariation and mechanism information as relevant to causal claims (e.g., Cheng & Novick, 1990; Koslowski, 1996; Park & Sloman, 2014), with some debate as to their relative contributions for different causal judgments (Ahn, Kalish, Medin & Gelman, 1995; Danks, 2005; Newsome, 2003). However, little is known about how these factors influence judgments of how good an *explanation* is, or about whether and how explanatory and causal judgments diverge with respect to the relative influence of covariation versus mechanism information. Addressing these questions is of interest for several reasons.

First, both philosophers and psychologists are interested in identifying "explanatory virtues" – characteristics that make for better explanations, such as simplicity, scope, and a specification of mechanism (e.g., Lipton, 2004). Research suggests that people find explanations more satisfying when they are simple and broad (for a review, see Lombrozo, 2012), with additional evidence that explanations are more likely to be inferred when they are more strongly supported by probabilistic evidence (Lombrozo, 2007). However, it's unknown whether explanations are also judged *better* when they are merely supported by stronger evidence, without some other explanatory relationship, such as a known causal mechanism, also in place. The influence of mechanisms on judgments of explanation "goodness" is also unknown, despite many suggestions that explanations and mechanisms are closely related (e.g., Ahn & Kalish, 2000; Bechtel & Abrahamsen, 2005; Machamer, Darden, & Craver, 2000).

Second, it's important to consider why mechanism information is valuable in the first place, whether for causal or explanatory judgments. For starters, mechanism information could affect the interpretation of covariation data, making people more confident that a correlation in fact supports the candidate causal relationship and is not, for example, the result of a common cause. Once a causal relationship is established, information about causal mechanisms will typically support predictions (Douglas, 2009) and interventions (Woodward, 2000): we can predict who will make donations by knowing whether they visited the portrait gallery, and we can make people more likely to donate by increasing their visits to that part of the museum. Mechanism information can also support broader generalizations from one case to another. In learning the mechanism in our museum example, we become better able to predict whether visiting a sculpture garden will have the same effect (it should depend on whether the sculptures have eyes), and on whether the effect might extend to other museum transactions (such as recycling one's museum badge versus buying a souvenir). According to accounts that link the function of explanation to generalization (Lombrozo & Carey, 2006), one might predict an especially strong effect of mechanism information on explanation judgments.

Third, the relationship between causal explanations and bare statements of the causal relationship they presuppose is largely unknown. For instance, in explaining museum donations by appeal to the portrait gallery, are we committing to any more or less than the claim that visiting the portrait gallery causally contributes to museum donations?

Identifying factors that differentially influence "matched" explanation and causation claims is a good strategy for beginning to address this question. If explanation claims can be reduced to the corresponding causal claims, we might anticipate differences in the absolute value of ratings assigned to each claim, but ratings for the different claims should respond similarly to manipulations of covariation strength and the presence of a mechanism.

For a similar reason, our experiments consider claims about understanding, e.g., how well people feel they understand the relationship between visiting the portrait gallery and making a museum donation. On some accounts, understanding amounts to a grasp of causes and/or explanations (e.g., Strevens, 2008), but empirical research has not considered how judgments of understanding relate to causal strength or explanation quality.

To investigate these issues, the experiments that follow manipulate the strength of covariation evidence and the specification of a mechanism, and elicit judgments about explanation "goodness," causal strength, and understanding. To preview our results, we find that judgments of causal strength are more responsive to covariation than either explanation or understanding judgments, while explanation judgments are more sensitive to the specification of a full mechanism than are causal judgments. In the general discussion we consider the implications of these results for the issues raised above.

Experiment 1a

Experiment 1 presented participants with two factors that were selected such that they would not suggest an obvious causal relationship. Participants received evidence about the covariation between these factors that suggested no relationship, a weak relationship, a moderate relationship, or a strong (deterministic) relationship. We also manipulated whether they received information about a possible mechanism.

Method

Participants Four-hundred-and-ninety-two participants were recruited on Amazon Mechanical Turk in exchange for $1.45. In all experiments, participation was restricted to users with an IP address within the United States and an approval rating of at least 95% based on at least 50 previous tasks. An additional 217 participants were excluded for failing a comprehension check for covariation tables (18), failing a memory check (199), or both (27).

Materials, Design, and Procedure Participants first completed a practice session in which they were introduced to covariation tables and received two problems that tested for comprehension. They were given feedback and requested to correct wrong responses. Participants who gave up on comprehension questions without providing the correct responses were excluded from further analysis.

Next, participants were presented with eight cause-effect pairs, selected to minimize prior beliefs about their relationship. Half of the participants were provided with a hypothetical mechanism connecting the cause and the effect. Below is sample text from one item:

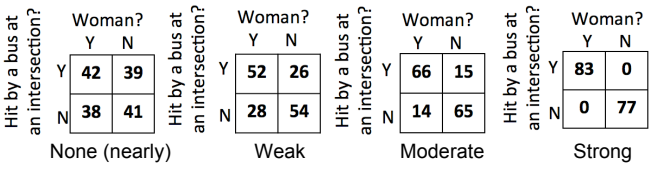
> 160 cyclists participated in a large survey. The survey included many questions. Two of the questions asked: a. whether or not the cyclist is a woman b. whether or not the cyclist has ever been hit by a bus at an intersection. These two things may or may not be related.
>
> *No mechanism:* In fact, the researchers who designed the survey didn't have any particular hypotheses about their relationship.
>
> *Full Mechanism:* When designing the survey, the researchers thought that they would be related as follows: Women are encouraged to obey rules more than men, so they stop at intersections for red lights more frequently than men do. This puts them in bus drivers' blind spot, so they get hit by buses more often than men.

Each cause-effect pair was also accompanied by a covariation table showing nearly no covariation, weak covariation, moderate covariation, or strong covariation (see Table 1). Covariation levels rotated through cause-effect pairs across participants, and each participant saw two cause-effect pairs for each level of covariation. A small amount of noise was introduced into the covariation data in the second set of tables to avoid presenting participants with identical tables.

Participants were assigned to one of the three judgment conditions: causal strength, explanatory goodness, or sense of understanding. Judgment questions were phrased either at the type or token level.[1] Below are sample judgments for the cyclist item, with token wording in brackets:

> [One of the respondents to the survey was LP, who is a woman. LP was hit by a bus at an intersection.]
> Based on the information you have, …
>
> *Causal strength:* do you think there exists a causal relationship between [LP] being a woman and [LP] getting hit by a bus at an intersection? No causal relationship (1) – Very strong causal relationship (9)
>
> *Explanatory goodness:* please rate how good you think

Table 1. Sample covariation matrices from Experiments 1-2. Conditions correspond to ΔP = .04, .33, .64 and 1.



[1] In Experiment 1a, participants who were presented with judgments in the token format gave higher ratings (M=5.65) than those presented with the type format (M=5.19, F(1,480)=10.94, p=.001, η_p^2=.022); however, the effect of format was not significant in Experiment 1b (F(1, 470)=.611, p=.435), and it did not interact with any other variables in any experiment, so all reported analyses collapse across this factor.

the following explanation is: Why do some cyclists get hit by buses at intersections? Because they are women. [Why was LP hit by a bus at an intersection? Because LP is a woman.] Very bad explanation (1) – Very good explanation (9)

Sense of understanding: do you feel you understand the relationship between [LP] being a woman and [LP] getting hit by a bus at an intersection? Very weak sense of understanding (1)– Very strong sense of understanding (9).

The order of trials was randomized for each participant. Finally, as a memory check, participants sorted causes from distractors and matched them with effects; those who made one or more errors were excluded from further analyses.

Results and Discussion

Are explanation ratings sensitive to covariation and mechanism information? Explanatory goodness ratings were subjected to a 4 (covariation: none, weak, moderate, strong) x 2 (mechanism: none, full) mixed ANOVA. This revealed a main effect of covariation evidence, $F(3,474)=118.16$, $p<.001$, $\eta_p^2=.428$ (all repeated contrasts $ps<.001$), with stronger ratings the stronger the covariation: $M_{none}=3.01$, $M_{weak}=4.79$, $M_{moderate}=5.56$, $M_{strong}=6.43$. There was also a main effect of mechanism, $F(1,158)=7.62$, $p=.006$, $\eta_p^2=.046$, with explanations rated as better when a full mechanism was provided ($M=5.29$ vs. $M=4.61$). In addition, this effect interacted with covariation, $F(3,474)=3.26$, $p=.021$, $\eta_p^2=.020$: a full mechanism significantly increased ratings when the covariation was absent ($M_{diff}=1.29$, $t(158)=4.56$, $p<.001$), but at higher levels of covariation this effect did not reach significance (weak: $M_{diff}=.55$, $t(158)=1.74$, $p=.083$; moderate: $M_{diff}=.75$, $t(158)=2.32$, $p=.022$; strong: $M_{diff}=.13$, $t(158)=.315$, $p=.753$, Bonferroni-corrected $p_{crit}=.013$). Because this interaction was not significant in subsequent experiments, we are inclined to attribute this effect in Experiment 1a to random variation in the data.

Are explanation, causation, and understanding ratings differentially affected by covariation information? For each participant we calculated the slope of ratings as a function of increasing covariation strength, and we compared mean slopes across the three judgment types in a one-way ANOVA, revealing a significant effect, $F(2,489) = 16.92$, $p<.001$, $\eta_p^2 = .065$. As shown in Figure 1, the mean slope of causal ratings *(*$M=1.41$) was higher than the slope of explanatory goodness ratings ($M=1.10$, Tukey HSD $p=.010$), which was in turn higher than the slope of understanding ratings ($M=.80$, $p=.013$).

Are explanation, causation, and understanding ratings differentially affected by mechanism information? A 2 (mechanism: none, full) x 3 (judgment: causal strength, explanatory goodness, sense of understanding) ANOVA on ratings revealed a main effect of mechanism, $F(1,486)=25.57$, $p<.001$, $\eta_p^2=.050$, with higher ratings for a full mechanism ($M=5.77$) than no mechanism ($M=5.07$), as well as a main effect of judgment, $F(2,486)=32.10$, $p<.001$, $\eta_p^2=.117$, with higher ratings for understanding ($M=6.23$; Tukey HSD $ps<.001$) than either causal strength ($M=5.14$) or explanation goodness ($M=4.95$), which did not differ from each other ($p=.500$). The interaction was not significant, $F(2,486)=.24$, $p=.785$.

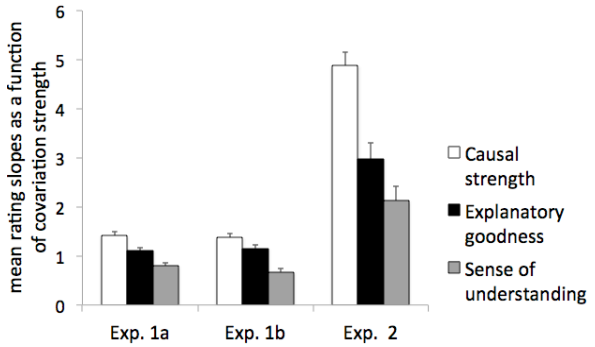


Figure 1: Mean covariation slopes as a function of judgment type in Experiments 1a, 1b and 2. Error bars: 1SE.

Experiment 1b

Experiment 1a found that explanations were judged better the stronger the corresponding covariation evidence, and when a full mechanism was provided. We also found that explanation judgments were less sensitive to covariation evidence than were causal judgments, but more sensitive than understanding judgments. The effect of mechanism did not differ significantly across judgment types.

In Experiment 1b we tested whether the specification of a full mechanism was necessary to observe a mechanism effect, or whether it would suffice to state that *some* mechanism connected the two factors. If people suffer from an "illusion of explanatory depth" (Rozenblit & Keil, 2002) and make do with quite skeletal mechanistic understanding (Keil, 2003), one might anticipate a boost in judgments from even a mechanism sketch or placeholder, and that this would be greater for explanation than causal judgments. We therefore duplicated the structure of Experiment 1a, but replacing detailed mechanism descriptions with a "mechanism pointer" - the statement that the factors in question are related via some unspecified mechanism.

Method

Participants Four-hundred-and-eighty-two participants were recruited on Amazon Mechanical Turk in exchange for $1.45. An additional 198 participants were excluded for failing a comprehension check for covariation tables (17), failing a memory check (181), or both (27).

Materials, Design, and Procedure were the same as in Experiment 1a, with the exception of the mechanism statement: the full mechanism was replaced with a general statement that there exists some multi-step pathway connecting the cause to the effect, omitting all other details.

Mechanism pointer: When designing the survey, the researchers thought they would be related by a multi-step pathway connecting being a woman to being hit by a bus

at an intersection: Women and men behave differently, and the differences in their behavior on the road result in a different probability of getting hit by a bus at an intersection.

Results and Discussion

Are explanation ratings sensitive to covariation and mechanism information? Explanatory goodness ratings were subjected to a 4 (covariation: none, weak, moderate, strong) x 2 (mechanism: none, pointer) mixed ANOVA. This revealed a main effect of covariation, $F(3,516)=146.50$, $p<.001$, $\eta_p^2 = .460$, with higher ratings the stronger the evidence: $M_{none}=2.43$, $M_{weak}=4.27$, $M_{moderate}=4.91$, $M_{strong}=6.01$, all repeated contrasts ps<.001). The main effect of mechanism did not reach significance, $F(1,172)=2.71$, $p=.102$, although the difference was in the predicted direction: no mechanism $M=4.33$, mechanism pointer $M=4.73$. The interaction was not significant, $F(3,516)=1.09$, $p=.352$.

Are explanation, causation, and understanding ratings differentially affected by covariation information? As in Experiment 1a, covariation slopes were analyzed as a function of judgment in a one-way ANOVA, revealing a significant effect, $F(2,479)=20.41$, $p<.001$, $\eta_p^2=.079$. As shown in Figure 1, the ordering of mean slopes mirrored Experiment 1a, but the difference between the slopes of causal ($M=1.37$) and explanatory ($M=1.15$) judgments did not reach significance (Tukey HSD $p=.111$). The slope for understanding ratings ($M=.67$) was significantly lower than that for causal or explanatory ratings (ps<.001).

Are explanation, causation, and understanding ratings differentially affected by mechanism information? A 2 (mechanism: none, pointer) x 3 (judgment: causal strength, explanatory goodness, sense of understanding) ANOVA on ratings revealed that providing a mechanism pointer did not significantly raise ratings, $F(1,476)=1.82$, $p=.178$: $M_{none}=5.03$ versus $M_{pointer}=5.26$, suggesting that a "skeletal" mechanism is insufficient to affect judgments. There was again a main effect of judgment, $F(2,476)=32.86$, $p<.001$, $\eta_p^2=.121$ ($M_{caus}=5.06$, $M_{expl}=4.52$, and $M_{und}=5.94$, all different, Tukey HSD $ps \leq .007$) and no interaction, $F(2,476)=1.03$, $p=.360$.

Experiment 2

Although providing detailed mechanisms in Experiment 1a boosted all ratings, the effect was weaker than we expected, which could have masked differences across judgments. In particular, it is possible that by presenting Experiments 1a and 1b as studies about the way people understand data tables, taking participants through an extensive practice session focusing on covariation tables, and manipulating covariation within subjects (while judgment and mechanism varied between subjects) we artificially drew attention to the covariation manipulation at the expense of the mechanism information. To address these concerns, we conducted Experiment 2, in which we minimized task features that drew attention to the covariation tables, hoping that it would set an "even playing field" for covariation and mechanism manipulations. We also combined the mechanism manipulations from Experiments 1a and 1b into a single variable with three levels (full mechanism, mechanism pointer, and no mechanism) and manipulated it within subjects, along with two levels of covariation (none, strong).

Method

Participants Two-hundred-and-fifty-one participants were recruited on Amazon Mechanical Turk in exchange for $1.55. An additional 81 participants were excluded for failing a memory check.

Materials, Design and Procedure Mechanism information (none, pointer, full) and covariation strength (none, strong) were manipulated within subjects, and rotated through items across participants. The type of judgment (explanation goodness, causal strength, sense of understanding) was manipulated between subjects.

The materials and procedure were the same as in Experiments 1a and 1b, with the following exceptions: the number of items (cause-effect pairs) was reduced to 6 and the practice session was shortened, as the comprehension questions about covariation tables were removed to avoid pragmatic cues that covariation evidence should be prioritized over mechanism information during the task. All questions were presented in the token format.

Results and Discussion

Are explanation ratings sensitive to covariation and mechanism information? Explanatory goodness ratings were subjected to a 2 (covariation: none, strong) x 3 (mechanism: none, pointer, full) repeated-measures ANOVA. This revealed a main effect of covariation, $F(1,85)=77.69$, $p<.001$, $\eta_p^2=.478$, with higher ratings for strong covariation ($M=5.74$) than no covariation ($M=2.76$). There was also a main effect of mechanism, $F(2,170) = 15.71$, $p<.001$, $\eta_p^2=.156$. Repeated contrasts indicated that ratings increased significantly from no mechanism ($M=3.69$) to a mechanism pointer ($M=4.34$) to a full mechanism ($M=4.72$), all ps<.05. The effects of mechanism and covariation did not interact, $F(2,170)=.341$, $p=.712$.

Are explanation, causation, and understanding ratings differentially affected by covariation information? As in Experiment 1a, covariation slopes were analyzed as a function of judgment in a one-way ANOVA, revealing a significant effect, $F(2,248)=21.27$, $p<.001$, $\eta_p^2=.146$. As shown in Figure 1, the ordering of mean slopes was the same as in Experiments 1a and 1b. The covariation slope for causal strength ratings ($M=4.87$) was significantly higher than the slopes for explanatory goodness ($M=2.97$) and understanding ratings ($M=2.12$, Tukey HSD ps<.001); the difference between the latter two was not significant ($p=.115$).

Are explanation, causation, and understanding ratings differentially affected by mechanism information? A 3 (mechanism: none, pointer, full) x 3 (judgment: causal

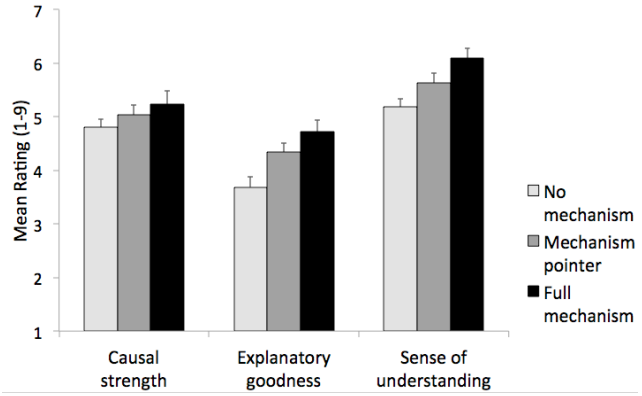


Figure 2: Mean ratings as a function of mechanism and judgment type in Experiment 2. Error bars: 1SE.

strength, explanatory goodness, sense of understanding) mixed ANOVA on ratings showed a significant main effect of mechanism, $F(2,496)=29.39$, $p<.001$, $\eta_p^2=.106$. Repeated contrasts showed that ratings increased significantly from no mechanism ($M=4.54$) to a mechanism pointer ($M=4.99$, $p<.001$) to a full mechanism ($M=5.34$, $p=.001$). Ratings were also significantly affected by judgment, $F(2,249)=18.67$, $p<.001$, $\eta_p^2=.131$: all judgments were significantly different from each other ($M_{caus}=4.25$, $M_{expl}=5.02$, $M_{und}=45.65$, Tukey $ps\leq.022$). Although the interaction did not reach significance, $F(2,496)=1.64$, $p=.162$, the pattern of means in Figure 2 suggested that providing the full mechanism had the most pronounced effect on explanation. This was confirmed by a significant ANOVA on full vs. no-mechanism difference scores, $F(2,248)=3.22$, $p=.042$, $\eta_p^2=.025$: a full mechanism produced a larger boost in explanation ratings than causal ratings (M_{diff} 1.04 vs. .44, Tukey $p=.043$); understanding received an intermediate boost ($M_{diff}=.91$, p's$\geq.143$). In contrast, the difference between the pointer and no-mechanism did not vary across judgments, $F(2,248)=1.35$, $p=.262$. This pattern is also consistent with Experiment 3 which finds that explanation goodness ratings are significantly more responsive to full mechanism information than causal strength ratings.

Experiment 3

Focusing on explanation ratings versus causal strength ratings and on the contrast between no mechanism and a full mechanism, Experiment 2 produced a double dissociation, with explanation ratings more sensitive than causal ratings when it came to mechanisms, and causal judgments more sensitive than explanation judgments when it came to covariation. While the differential effect of covariation was also found in Experiments 1a and 1b, the effect of mechanism information was not. We therefore sought to replicate the interactions between mechanism and judgment in Experiment 2 before drawing strong conclusions. We also tied the mechanism more closely to each judgment by embedding the mechanism information in the body of the explanation and causation statements themselves.

Method

Participants Ninety-one participants were recruited on Amazon Mechanical Turk in exchange for $1.00. An additional 16 participants were excluded for failing a memory check.

Materials, Design and Procedure Experiment 3 included the following changes from Experiment 2: the mechanism information was included in the body of the explanation or causal statement (e.g., explanation with a mechanism pointer: "MP was hit by a bus at an intersection because MP is a woman, and there exists a multi-step pathway that connects being a woman to being hit by a bus: women and men behave differently, and the differences in their behavior on the road result in a different probability of getting hit by a bus at an intersection."); the covariation variable was dropped; the understanding judgment was dropped; and both judgment type (causal strength, explanation goodness) and mechanism (none, pointer, full) were manipulated within subjects. Judgments were blocked, with the order of blocks randomized across participants. Prior to the second block, participants were invited to "pay attention to the changed rating scale." Mechanism levels were randomized within each judgment block. Items rotated through conditions across participants.

Results and Discussion

A 3 (mechanism: none, pointer, full) x 2 (judgment: explanation goodness, causal strength) repeated-measures ANOVA revealed a significant main effect of mechanism, $F(2,180)=48.71$, $p<.001$, $\eta_p^2=.351$, with ratings increasing from no mechanism ($M=1.71$) to a mechanism pointer ($M=2.20$) to a full mechanism ($M=3.43$, repeated contrasts $ps\leq.001$), and no main effect of judgment, $F(1,90)=.99$, $p=.323$. Critically, there was also a significant interaction between mechanism and judgment, $F(2,180)=3.06$, $p=.049$, $\eta_p^2=.033$. As shown in Figure 3, the differences across mechanism conditions were more pronounced for explanatory than causal judgments. As in Experiment 2, this interaction was driven by the difference between the no mechanism and full mechanism conditions: the comparison of full minus no-mechanism difference for explanation vs. casual ratings was significant, $t(90)=2.18$, $p=.032$, but the pointer vs. no-mechanism difference did not vary across judgments ($t(90)=.04$, $p=.971$).

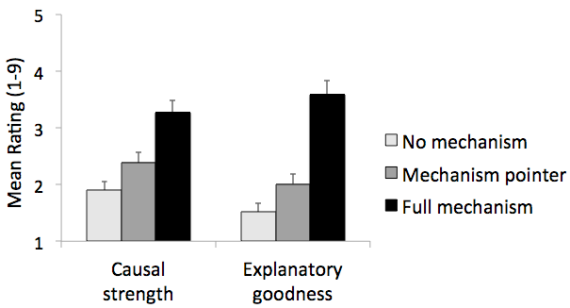


Figure 3: Mean ratings as a function of mechanism and judgment type in Experiment 3. Error bars: 1SE.

General Discussion

In four experiments we demonstrate that judgments of explanation "goodness" are sensitive to both covariation evidence and mechanism information. Comparing explanation to other judgments, we observed a consistent dissociation: explanation judgments were less responsive to the degree of covariation in the data than were causal judgments. In contrast, specifying a full mechanism had a stronger effect on explanations than on causal judgments in Experiments 2 and 3, which drew less attention to the covariation tables. Of the three judgment types, "sense of understanding" was least responsive to covariation. Overall, our results indicate that these three types of judgments differ systematically when it comes to the role of covariation data and the effects of specifying a full mechanism.

Returning to the issues raised in the introduction, our findings support some tentative conclusions and raise additional questions for further study. First, we find that explanations are judged better when supported by stronger covariation evidence or by the specification of a mechanism, and that the benefits of stronger evidence are not limited to cases in which a mechanism is also specified. It would be interesting to know whether these two factors affect explanation ratings for different reasons – for example, covariation might be valuable for purely evidential reasons, while the specification of a mechanism could be a genuine "virtue" in addition to having evidential import.

Second, full mechanism information does appear to have a larger effect on explanation goodness ratings relative to causal strength ratings, as might be expected on the view that explanations are especially geared towards generalization (Lombrozo & Carey, 2006), which full mechanism information supports. More speculatively, it could also be that reduced sensitivity to covariation emerges for a similar reason: a certain degree of resistance to over-fitting the data from a single sample could help achieve more reliable generalizations (and indeed, Williams, Lombrozo, & Rehder, 2013 show that explanation encourages a search for broad patterns despite inconsistent data).

Third, our findings suggest that explanatory goodness cannot be reduced, in any straightforward way, to judgments of causal strength. Similarly, ratings of understanding diverge from those of either explanation or causation. Our findings thus call for caution when characterizing one of these judgments in terms of another, and also raise questions about the extent to which different kinds of explanatory and causal judgments could diverge. For instance, evaluating explanatory "goodness" could diverge from evaluations of explanation probability, and evaluations of causal structure could diverge from those of strength.

In sum, we demonstrate that judgments of causal strength, explanatory goodness and, to some extent, understanding respond differently to covariation and full mechanism information. Explanations surpass causal judgments in their sensitivity to a full mechanism, and the pattern is reversed for covariation. Our results present a challenge for proposals that characterize explanations as identifying causes, and characterize understanding in terms of grasping causal relationships and/or explanations. More importantly, these patterns of divergence can begin to help us understand the different roles of these judgments in our cognitive lives.

Acknowledgments

This work was supported by the Varieties of Understanding Project, funded by the John Templeton Foundation.

References

Ahn, W.-K., & Kalish, C. (2002). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition*. MIT Press.

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

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*, 412–414.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58,* 545-567.

Danks, D. (2005). The supposed competition between theories of human causal inference. *Philosophical Psychology*, *18*, 259–272.

Douglas, H. E. (2009). Reintroducing prediction to explanation. *Philosophy of Science, 76*(4), 444-463.

Keil, F.C. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Science*, 7, 368-373.

Koslowski, B. (1996). *Theory and Evidence*. MIT Press.

Lipton, P. (2004). *Inference to the best explanation*. Psychology Press.

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

Lombrozo, T. (2012). Explanation and abductive inference. *Oxford Handbook of Thinking and Reasoning*, 260–276.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking About Mechanisms. *Philosophy of Science, 67*(1), 1–25.

Newsome, G. L. (2003). The debate between current versions of covariation and mechanism approaches to causal inference. *Philosophical Psychology*, *16*(1), 87–107.

Park, J., & Sloman, S. (2014). Causal explanation in the face of contradiction. *Memory & Cognition*, *42*(5), 806–20.

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

Strevens, M. (2008). *Depth: an account of scientific explanation.* Harvard University Press.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*(4), 1006-1014.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science*, *51*, 197 –254.