Explanatory Scope Informs Causal Strength Inferences

Samuel G. B. Johnson (samuel.johnson@yale.edu)
Angie M. Johnston (angie.johnston@yale.edu)
Department of Psychology, Yale University
2 Hillhouse Ave., New Haven, CT 06520 USA

Amy E. Toig (aet2144@columbia.edu)
Department of Psychiatry, Columbia University
1051 Riverside Dr., New York, NY 10032 USA

Frank C. Keil (frank.keil@yale.edu)
Department of Psychology, Yale University
2 Hillhouse Ave., New Haven, CT 06520 USA

Abstract

People judge the strength of cause-and-effect relationships as a matter of routine, and often do so in the absence of evidence about the covariation between cause and effect. In the present study, we examine the possibility that *explanatory power* is used in making these judgments. To intervene on explanatory power without changing the target causal relation, we manipulated explanatory scope—the number of effects predicted by an explanation—in two different ways, finding downstream consequences of these manipulations on causal strength judgments (Experiment 1). Directly measuring perceived explanatory power for the same items also revealed item-by-item correlations between causal strength and explanatory power (Experiment 2). These results suggest that explanatory power may be a useful heuristic for estimating causal strength in the absence of statistical evidence.

Keywords: Causal reasoning; explanation; diagnostic reasoning; explanatory scope.

Introduction

Causes come in all shapes and sizes. In the face of this variety, we must often assess causal structure and strength on the fly, with limited computational resources and without access to evidence about the covariation between cause and effect. In these situations, people use several cues, at least when assessing causal *structure*, including mechanism knowledge (e.g., Ahn, Kalish, Medin, & Gelman, 1995), temporal information (e.g., Lagnado & Sloman, 2006), and event structure cues (e.g., Johnson & Keil, 2014a). Less is known, however, about how causal *strength* is estimated when covariation information is unavailable. In this paper, we examine one potential cue that might be used for estimating causal strength in statistically impoverished settings—explanatory power.

According to most normative models, prior statistical evidence is not only helpful, but necessary for computing causal strength judgments for binary causes and effects. For example, one simple normative account models causal strength as the magnitude of difference the cause makes to the probability of the effect (Allan, 1980), though more complex models have since been proposed (see Perales & Shanks, 2007 for a review). Determining the relevant probabilities requires multiple observations to calculate the covariation between cause and effect.

Clearly, however, we also have intuitions about causal strength in one-shot cases where we observe a cause and effect occur just one time, and thus lack any statistical evidence. For instance, when a new legal statute has an unintended social consequence, when an earnings report affects a company's business plan, or when ingesting an unfamiliar drink causes a funny tingling, we have a sense of how powerful these causes are in producing the effects, even though we may have no statistical data at all. Where do these intuitions come from?

One possible source is a cause's *explanatory power*. Philosophers have argued that explanation plays a central role in our mental lives (Strevens, 2008) and identified a number of *explanatory virtues* (Lipton, 2004) that people use in assessing explanatory power, such as simplicity, scope, and depth (see also Lombrozo, 2012). Critically, these explanatory virtues or heuristics often are defined in one-shot cases, so beliefs about explanatory power can often be formed in the absence of statistical evidence.

Explanatory power may be a useful cue to causal strength because these factors tend to be correlated (Strevens, 2008). For instance, if a new law inadvertently incentivizes people to drive faster, then the law will be seen as a good explanation for driving speed to the extent that the law made a large causal difference to driving speed. People might capitalize on this relationship between causal strength and explanatory power in the reverse direction, to use their own perceptions of explanatory power to predict causal strength. If explanatory power can sometimes be more easily assessed than causal strength, then this may be an effective strategy for inferring causal strength, especially in one-shot cases.

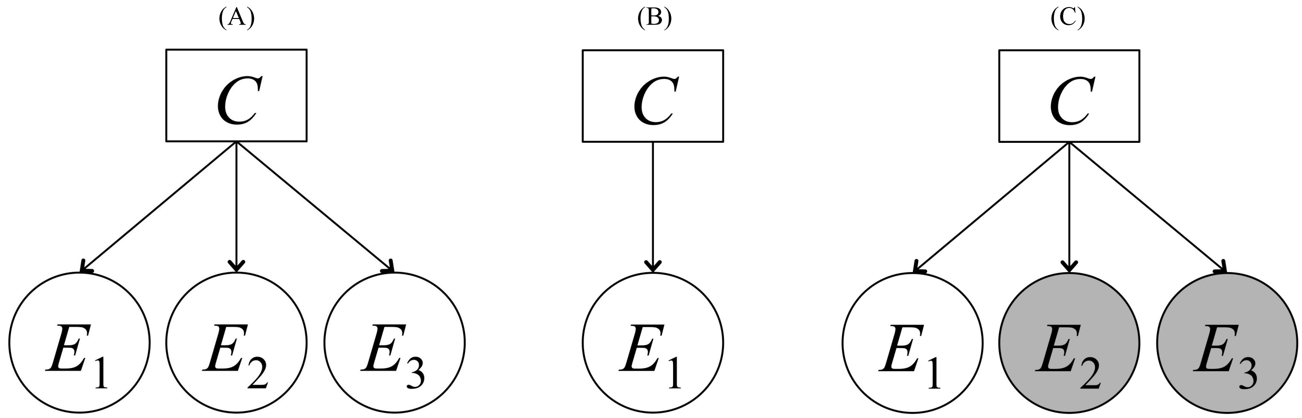

Figure 1: Explanatory structures used in Experiments 1 and 2. (A) *Good* structure (wide manifest scope), (B) *Bad-Manifest* structure (narrow scope), (C) *Bad-Latent* structure (wide latent scope). White circles indicate observed predictions and gray circles indicate latent predictions.

However, this tight correspondence between causal and explanatory power also creates a methodological problem, because if we find a close link between explanatory power and causal strength, it could be due to the influence of causal strength on explanatory power, rather than the reverse. To solve this problem in our experiments, we used cases with identical causal claims, but manipulated whether an explanatory virtue was present or absent. We were thus able to intervene directly on explanatory power without changing the underlying causal relationship.

We followed this strategy for two explanatory virtues. In Experiment 1A, we manipulated the *manifest scope* of an explanation to test the influence on causal judgments, and in Experiment 1B, we manipulated *latent scope*. In Experiment 2, we directly measured explanatory power for the items used in Experiment 1, to examine the correspondence between explanatory and causal judgments. Across manipulations and across items, we expected perceived causal strength to depend on perceived explanatory power.

Experiments 1A and 1B

The quality of an explanation depends not only on its relationship with what it is explaining, but also on the other predictions it makes. The set of predictions made by an explanation is known as its *scope*. For example, a disease called Ferraro's Disorder (C) might have three characteristic symptoms—hair loss (E_1), weight gain (E_2), and night terrors (E_3). When diagnosing the cause of Randy's hair loss (E_1), we benefit from knowing about E_2 and E_3 in assessing whether Ferraro's Disorder is the best explanation (see Figure 1-A). That is, an explanation's scope is used in determining explanatory power. We can partition an explanation's scope into two parts—its *manifest scope*, consisting of all the observed predictions (Read & Marcus-Newhall, 1993), and its *latent scope*, consisting of all the potential but unverified predictions (Khemlani, Sussman, & Oppenheimer, 2011). For example, suppose we know that Randy had hair loss (E_1), but we do not know whether or not he also had weight gain (E_2) or night terrors (E_3). In this case, E_1 is in the *manifest scope* of Ferraro's Disorder, and E_2 and E_3 are in the *latent scope* (see Figure 1-C). People tend to prefer explanations with wide manifest scope, accounting for as many actual observations as possible (Read & Marcus-Newhall, 1993) and narrow latent scope, accounting for as few potential but as-yet-unobserved effects as possible (Khemlani et al., 2011; but see Johnson, Rajeev-Kumar, & Keil, 2014 for boundary conditions).

In Experiment 1, we capitalized on these explanatory preferences to manipulate explanatory power, looking for downstream consequences for causal strength judgments. Participants read vignettes depicting a causal relationship, where we varied features of the explanatory structure, such that some explanations were "good" and others were "bad." In Experiment 1A, we contrasted explanations with wider manifest scope (the *Good* condition) to explanations with narrower manifest scope (the *Bad-Manifest* condition). For example, a *Good* item read:

> When someone has Ferraro's Disorder, they lose hair, gain weight, and have frequent night terrors.
>
> Three months ago, Randy developed Ferraro's Disorder. Because he has Ferraro's Disorder, Randy lost hair, gained weight, and had frequent night terrors.

That is, Ferraro's Disorder (C) has three effects in its scope—hair loss (E_1), weight gain (E_2), and night terrors (E_3), of which all three were observed (see Figure 1-A). In contrast, a *Bad-Manifest* item read (see Figure 1-B):

> When someone has Ferraro's Disorder, they lose hair.
>
> Three months ago, Randy developed Ferraro's Disorder. Because he has Ferraro's Disorder, Randy lost hair. We also know that Randy gained weight and had frequent night terrors, but we don't know why.

Here, Randy still has the same three symptoms, but two of them are unexplained by Ferraro's Disorder since it now has only one effect in its scope—hair loss (E_1). On the basis of previous results on explanatory preferences (e.g., Read & Marcus-Newhall, 1993), we expected that Ferraro's Disorder would be seen as a more powerful explanation for Randy's hair loss when it explained all of Randy's symptoms (in the *Good* condition), rather than just one (i.e., in the *Bad-Manifest* condition).

In Experiment 1B, we contrasted the *Good* explanations from Experiment 1A with *Bad-Latent* explanations that predicted the same effects as the *Good* explanations, but where some of the predicted effects were latent rather than manifest. This version read (see Figure 1-C):

> When someone has Ferraro's Disorder, they lose hair, gain weight, and have frequent night terrors.
>
> Three months ago, Randy developed Ferraro's Disorder. Because he has Ferraro's Disorder, Randy lost hair. We don't know if he gained weight or had frequent night terrors.

In this version, Ferraro's Disorder accounts for three potential observations, with E_1 observed (in the manifest scope) and E_2 and E_3 unknown (in the latent scope). We expected Ferraro's Disorder to be seen as a more powerful explanation of Randy's hair loss when it made predictions that were manifest rather than latent. Indeed, we might expect the *Bad-Latent* version to be less powerful than even the *Bad-Manifest* version, because both versions involve the same observations (E_1 only) but the *Bad-Latent* version also predicts E_2 and E_3, whereas the *Bad-Manifest* version does not (see Figure 1). That is, the *Bad-Latent* version differs from the *Bad-Manifest* version only in having wider latent scope, which makes explanations less powerful even when the observations are held constant (Khemlani et al., 2011).

In all three versions of the item, Ferraro's Disorder is said to have caused Randy's hair loss, but the extent to which Ferraro's Disorder is judged as a *powerful* causal explanation should differ across conditions due to our scope manipulations. Therefore, if people use explanatory power as a way to estimate causal strength, causal strength ratings should differ across conditions. In contrast, if only the reverse were true—that causal strength merely influences explanatory power—then these manipulations should have no effect on causal judgments.

Method

Participants We recruited 100 participants from Amazon Mechanical Turk for Experiment 1A, and another 100 participants for Experiment 1B. Zero participants from Experiment 1A and one participant from Experiment 1B were excluded from data analysis because they incorrectly answered more than 33% of a series of check questions.

Design Items were causal explanations drawn from eight categories covering a variety of everyday and scientific topics (e.g., medicine, sports, chemistry). Each category contained two items, which were conceptually similar but differed in content (e.g., the two medical items described different fictitious diseases with distinct symptoms). Each participant thus completed a total of 16 items. In Experiment 1A, participants saw the *Good* version of one item from each category and the *Bad-Manifest* version of the other item from each category, in a counterbalanced manner (see above for example wordings). In Experiment 1B, participants saw the *Good* version of one item from each category and the *Bad-Latent* version of the other item. In both experiments, each *Good* item was always presented adjacent to a *Bad* item from a different category, forming eight pairs; the order within each pair was randomized, as was the order of the pairs.

Procedure For each of the 16 items, participants first read the explanation (as worded above for each condition, with the effects color-coded to make the paragraphs easier to parse). Participants then answered a *causal structure* question (e.g., "Do you think that Randy having Ferraro's Disorder caused him to lose hair?") formatted as a yes/no forced-choice, to ensure that participants acknowledged the existence of a causal relationship. On the next screen, they rated *causal strength* ("How *strong* do you think this relationship is between Randy having Ferraro's Disorder and Randy losing hair?") on a scale from 1 ("Extremely Weak") to 9 ("Extremely Strong"), with the passage from the previous page at the bottom of the screen as a reminder. They were asked to make this strength rating regardless of their response to the structure question, because we wanted to discourage participants from adopting the strategy of answering "no" to the initial question in an effort to shorten the task.

Results and Discussion

Both manipulations of explanatory structure influenced causal strength judgments, and this effect was larger in Experiment 1B than in Experiment 1A (see Figure 2). For each participant, we averaged across the items for which that participant answered "yes" to the initial causal structure question ("Do you think that [*C*] caused [E_1]?"). These strength ratings were higher for the *Good* items than for the *Bad-Manifest* items in Experiment 1A (M =

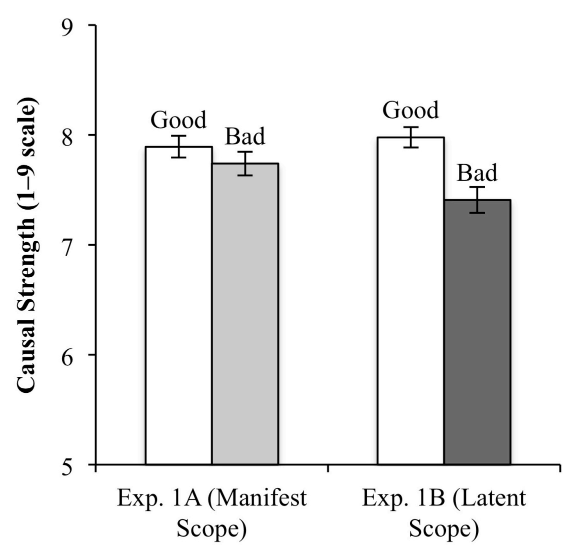

Figure 2: Results of Experiment 1. Bars represent ±1 SE.

7.89, $SD = 0.99$ vs. $M = 7.74$, $SD = 1.08$; $t(99) = 2.28$, $p = .025$, $d = 0.23$), and were higher for the *Good* items than for the *Bad-Latent* items in Experiment 1B ($M = 7.98$, $SD = 0.92$ vs. $M = 7.41$, $SD = 1.17$; $t(98) = 6.22$, $p < .001$, $d = 0.62$. The difference in strength ratings between the *Good* and the *Bad-Latent* items was larger than the difference between the *Good* and the *Bad-Manifest* items, leading to an interaction between item type (good or bad explanation) and manipulation (manifest scope or latent scope), $F(1,198) = 13.37$, $p < .001$, $\eta_p^2 = .06$, as depicted in Figure 2. Item analyses were also conducted, using the average for each item among those participants who responded "yes" to the initial causal question. These analyses show that the effects generalized well across items, $t(15) = 2.82$, $p = .013$, $d = 0.70$ for Experiment 1A and $t(15) = 9.84$, $p < .001$, $d = 2.46$ for Experiment 1B.

These results show that explanatory scope—a feature known to affect the perceived quality of an explanation—influences judgments of causal strength. This result held using two distinct manipulations of scope: A contrast between the *Good* and *Bad-Manifest* items, where the observations were the same but the scope differed, and a contrast between the *Good* and the *Bad-Latent* items, where the scope was the same but the observations differed. These findings are consistent with the idea that explanatory power is used to estimate causal strength.

An alternative interpretation is that participants' answers to the strength questions may not have reflected their beliefs about causal strength at all, but instead their beliefs about causal structure. On this account, it is not necessary to invoke explanatory power to explain the causal strength ratings, because these ratings were really covert structure judgments about participants' confidence in the *existence* of a causal relationship. In fact, some previous findings in causal induction can be modeled in this manner (Griffiths and Tenenbaum, 2005), and a critic might account for our results in a similar way. Suppose we were uncertain about whether a causal relationship exists in a particular case—for example, we know that Randy lost hair but we do not know whether it was due to Ferraro's disorder or to some other cause. In such cases, the other effects predicted by a common cause explanation can be used to corroborate the cause's presence—that is, Randy's weight gain and night terrors could be used as evidence that Randy has Ferraro's disorder. There is thus not only more explanatory power in the *Good* conditions than in the *Bad* conditions, but also more evidence for Ferraro's disorder as the cause of Randy's hair loss, which would also lead to higher casual strength judgments according to this alternative account.

However, we took several measures to guard against this possibility. Critically, we sought to eliminate possible uncertainty over the causal structure by stating explicitly that the cause occurred (e.g., "Three months ago, Randy developed Ferraro's disorder") and that it caused the effect in the token case (e.g., "Because he has Ferraro's disorder, Randy lost hair"). We also added a separate structure question prior to the strength question ("Do you think that Randy having Ferraro's disorder *caused* him to lose hair?"), creating pragmatic pressure to interpret the strength question as a distinct query from the structure question. This also allowed us to include in our analyses only causal strength ratings following affirmative responses to the initial structure question.

Despite the explicit statement of causal structure, there was nonetheless a minority of "no" responses to the structure questions, and this minority was somewhat larger for the *Bad-Manifest* and *Bad-Latent* items than for the *Good* items. "No" responses were marginally more frequent for the *Bad-Manifest* items than for the *Good* items in Experiment 1A ($M = 8.75\%$ vs. $M = 6.50\%$; $t(99) = 1.84$, $p = .069$, $d = 0.18$), and significantly more frequent for the *Bad-Latent* items than for the *Good* items in Experiment 1B ($M = 9.85\%$ vs. $M = 3.54\%$; $t(98) = 3.68$, $p < .001$, $d = 0.37$). However, this subset of causal deniers cannot account for the differences in causal strength ratings across conditions. Even if we consider just those participants who responded "yes" to the causal structure question for all 16 items (about 55% of participants), the effects of our manipulations on causal strength held up for both Experiment 1A (*M*s = 8.23 vs. 8.14; $t(56) = 1.95$, $p = .056$, $d = 0.26$) and for Experiment 1B (*M*s = 8.19 vs. 7.87; $t(54) = 3.77$, $p < .001$, $d = 0.51$). Therefore, covert structure inferences in the place of causal strength judgments cannot explain our findings.

A further challenge to our interpretation, however, comes from the relative sizes of our manifest and latent scope manipulations. Previous studies have generally found larger effects for manipulations of manifest scope (e.g., Read & Marcus-Newhall, 1993) than of latent scope (e.g., Khemlani et al., 2011), whereas we found a larger effect for our latent scope manipulation. We note, however, that the manipulation of latent scope used in previous studies (e.g., Khemlani et al., 2011) is analogous to the difference between our *Bad-Manifest* and *Bad-Latent* conditions, rather than between our *Good* and *Bad-Latent* conditions (i.e., the comparison between Figure 1-B and 1-C, rather than 1-A and 1-B). In those studies, the observations being explained were the same but the scope differed, with the wide latent scope explanation making additional, unverified predictions not made by the narrow latent scope explanation. Similarly, only E_1 was observed in both our *Bad-Manifest* and *Bad-Latent* conditions, but the *Bad-Latent* version had E_2 and E_3 in its scope, while the *Bad-Manifest* version did not. This suggests that the *Bad-Manifest* version may have been seen as a more powerful explanation than the *Bad-Latent* version, in addition to the *Good* version being seen as more powerful than the *Bad-Manifest* version. This would lead to a larger effect in Experiment 1B (contrasting *Good* and *Bad-Latent*) than in Experiment 1A (contrasting *Good* and *Bad-Manifest*), as we found. Nonetheless, we measured perceived explanatory strength directly in Experiment 2 to provide experimental corroboration for this account.

Experiment 2

In Experiment 2, participants read the same items used in Experiment 1, providing judgments of "how satisfying" they perceived C to be as an explanation for E_1 (in line with previous research on explanatory preferences; e.g., Khemlani et al., 2011). Given our results in Experiment 1, we would expect the *Good* items (with wide manifest scope) to be seen as the best explanations, followed by the *Bad-Manifest* items (with narrow manifest and latent scope), and then by the *Bad-Latent* items (with wide latent scope). This design also allowed us to look for item-by-item correlations between explanatory power and causal strength, to explore whether the causes that most powerfully explained their effects were also thought to be the strongest causes of their effects.

Method

Participants We recruited 60 participants from Amazon Mechanical Turk. Two participants were excluded because they incorrectly answered more than 33% of the check questions.

Procedure Participants saw each of the 16 items used in Experiment 1, where each item was randomly assigned to either its *Good* version, its *Bad-Manifest* version, or its *Bad-Latent* version (see Experiment 1 for example wordings). After reading the item, participants completed an *explanatory power* rating (e.g., "To what extent do you think that Randy having Ferraro's Disorder is a satisfying explanation for why Randy lost hair?") on a scale from 1 ("Not at all satisfying") to 9 ("Very satisfying").

Results and Discussion

As shown in Figure 3, ratings of explanatory power differed across the different versions of each item as expected. The *Good* items were rated more satisfying (M = 7.69, SD = 1.09) than the *Bad-Manifest* items (M = 7.43, SD = 1.27), $t(57) = 2.18$, $p = .034$, $d = 0.29$, which were rated more satisfying than the *Bad-Latent* items (M = 6.68, SD = 1.64), $t(57) = 3.54$, $p = .001$, $d = 0.46$. Thus,

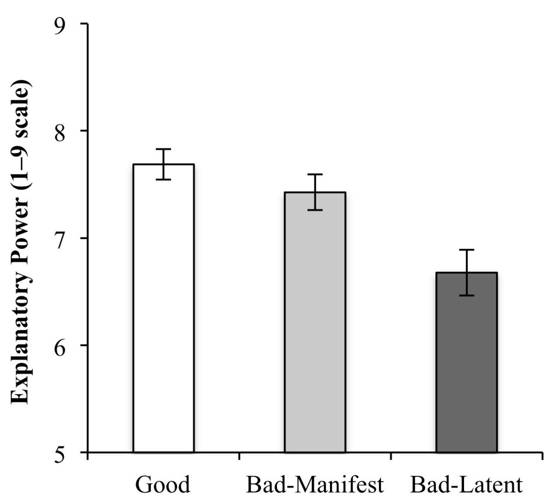

Figure 3: Results of Experiment 2. Bars represent ±1 SE.

the manifest scope manipulation was a weaker manipulation of explanatory power compared to the latent scope manipulation. This is consistent with our findings in Experiment 1, where the latent scope manipulation had a relatively large effect on causal strength ratings, but the manifest scope manipulation had a smaller effect.

We also examined the correlations on an item-by-item basis between explanatory power ratings and causal strength judgments from Experiment 1 (causal strength judgments for the *Good* items were averaged across Experiments 1A and 1B). Across all 48 combinations of item and version, the correlation between explanatory power and causal strength was highly robust, $r(46) = .65$, $p < .001$. Looking just within each version of each item, this correlation was significant within the *Bad-Manifest* items, $r(14) = .70$, $p = .002$, and was positive but non-significant within the *Good* items, $r(14) = .19$, $p = .48$, and the *Bad-Latent* items, $r(14) = .42$, $p = .11$. These positive correlations are consistent with the idea that explanatory power is used to infer causal strength, and also help to assuage a potential concern about Experiment 1—that although we manipulated the scope as a way to intervene on explanatory power, the effect of scope on causal strength judgments may have been mediated by some factor other than explanatory power. Such alternative explanations would be unable to explain these positive item-by-item correlations within each version.

General Discussion

Even though life often fails to lay out statistical information in a digestible way, we are nonetheless able to infer causal strength from individual cases. To see whether this ability might in part draw on our explanatory capacities, we manipulated the explanatory virtues present in otherwise identical causal explanations and looked for downstream effects on causal strength judgments. In Experiment 1A, causes that accounted for more observations (i.e., causes of wider manifest scope) were judged stronger than causes that accounted for fewer observations. In Experiment 1B, causes whose predictions were all verified (i.e., causes of narrow latent scope) were judged stronger causes than those that made unverified predictions. In Experiment 2, the same manipulations had similar effects on explanatory power judgments, and item-by-item differences in perceived explanatory power were positively correlated with differences in causal strength judgments. These effects all suggest that explanatory power is used to estimate causal strength in token cases where covariation information is unavailable.

These results add to previous research documenting ways that prior knowledge influences causal judgments. For example, covariation information has a stronger effect on causal judgments when the causal candidate is believable than when it is unbelievable (Fugelsang & Thompson, 2000) and both laypeople and working scientists discount data that is inconsistent with their hypotheses (Fugelsang, Stein, Green, & Dunbar, 2004).

The present results differ from these prior findings, however, in demonstrating that even in the absence of *any* data, beliefs about explanatory power can influence judgments of causal strength. Instead, these findings may be more closely related to analogical mappings from one causal token to another (Holyoak, Lee, & Lu, 2010), where covariation information is not used at all. The tight coupling between causal and explanatory strength seems to license a heuristic wherein one can be used as a proxy for the other in any token case, without necessarily referring to the statistics of the broader reference class. This may explain in part why people often prefer explanatory information over statistical information when evaluating causal tokens (e.g., "Jim's smoking caused him to get lung cancer"), but prefer statistical information when evaluating causal types ("A person's smoking causes them to get lung cancer"; Johnson & Keil, 2014b).

An open question for future research is whether a similar explanatory heuristic could be at play in causal *structure* inference as well. The present studies were not designed with this question in mind (we specified that causal relationships existed in all cases), but participants in Experiment 1 were indeed somewhat more likely to reject the structure claim (e.g., "Randy having Ferraro's disorder caused him to lose hair") when explanatory quality was lower. Future studies might test this possibility by manipulating explanatory virtues that do not normatively license different inferences in structurally ambiguous cases.

Conclusion

We must often infer both the shapes and sizes of causal relationships without contingency information at our disposal, and in such cases we must rely on prior knowledge and cues from the environment to make these inferences. The present results show that the explanatory power of a causal relationship is one guide we use to make these strength inferences. This finding underscores the importance of research on explanatory preferences by documenting downstream consequences of explanatory reasoning for causal inference. Just as the explanatory structures in our minds mirror the causal structure in the world, so do our causal perceptions mirror our explanatory intuitions. More thorough understanding of both sides of this feedback loop will be needed to ground this circle between mind and environment.

Acknowledgments

This research was supported by a grant from the National Institutes of Health to F.C. Keil. We thank the members of the Yale Cognition & Development Lab for their valuable feedback.

References

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

Allan, L.G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*, 147–149.

Fugelsang, J.A., Stein, C.B., Green, A.E., & Dunbar, K.N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology*, *58*, 86–95.

Fugelsang, J.A., & Thompson, V.A. (2000). Strategy selection in causal reasoning: When beliefs and covariation collide. *Canadian Journal of Experimental Psychology*, *54*, 15–32.

Griffiths, T.L., & Tenenbaum, J.B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 334–384.

Holyoak, K.J., Lee, H.S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, *139*, 702–727.

Johnson, S.G.B., & Keil, F.C. (2014a). Causal inference and the hierarchical structure of experience. *Manuscript under review*.

Johnson, S.G.B., & Keil, F.C. (2014b). Pluralism in causal learning and representation. *Manuscript under review.*

Johnson, S.G.B., Rajeev-Kumar, G., & Keil, F.C. (2014). Inferred evidence in latent scope explanations. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S.S., Sussman, A.B., & Oppenheimer, D.M. (2011). *Harry Potter* and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*, 527–535.

Lagnado, D.A., & Sloman, S.A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 451–460.

Lipton, P. (2004). *Inference to the best explanation* (2nd Edition.). London: Routledge.

Lombrozo, T. (2012). Explanation and abductive inference. In K.J. Holyoak & R.G. Morrison (Eds.), *Oxford handbook of thinking and reasoning*. Oxford, UK: Oxford University Press.

Perales, J.C., & Shanks, D.R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin & Review*, *14*, 577–596.

Read, S.J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*, 429–447.

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