Philosophy Compass 6/8 (2011): 539–551, 10.1111/j.1747-9991.2011.00413.x

The Instrumental Value of Explanations

Tania Lombrozo*
University of California

Abstract

Scientific and 'intuitive' or 'folk' theories are typically characterized as serving three critical functions: prediction, explanation, and control. While prediction and control have clear instrumental value, the value of explanation is less transparent. This paper reviews an emerging body of research from the cognitive sciences suggesting that the process of seeking, generating, and evaluating explanations in fact contributes to future prediction and control, albeit indirectly by facilitating the discovery and confirmation of instrumentally valuable theories. Theoretical and empirical considerations also suggest why explanations may nonetheless *feel* intrinsically valuable. The paper concludes by considering some implications of the psychology of explanation for a naturalized philosophy of explanation.

1. *Introduction*

The quest for explanations characterizes both science and everyday life. We seek to understand the origins of the universe and the intricacies of protein folding, as well as the causes of an inconvenient traffic jam or a friend's surprising behavior. Explanations are clearly a product of great value, but why?

Philosophers have proposed two quite different answers. The first is that explanations are intrinsically valuable: an end in themselves, or perhaps a means to understanding, which is itself intrinsically valuable. The second is that explanations are instrumentally valuable: a means to a more tangible and practical benefit, such as greater predictive success when it comes to negotiating a complex world.

Recent empirical findings in cognitive development, cognitive psychology, and education have begun to uncover core properties and consequences of explanation – both the act or process, and the outcome or product. These findings have implications for claims about the instrumental and intrinsic value of explanations. In particular, research increasingly points to an important role for explanation in the development and confirmation of intuitive theories about the social and physical worlds. Such theories have clear practical value in supporting prediction, intervention, and more general reasoning and inference. At the same time, this body of research reveals why explanations may *feel* like ends in themselves.

In this paper, I argue that explanations have critical instrumental value for everyday cognition, and that parallel considerations apply to explanation in science. The paper begins, in Section 2, with a brief introduction to previous claims about the intrinsic and instrumental value of explanations. The purpose is not to provide an exhaustive overview of arguments for each position, but simply to elicit the compelling insight that motivates each view. In Section 3, I turn to empirical research indicating a central role for explanation in discovery and confirmation, two of the key processes in the development of scientific and intuitive theories. Section 4 then considers why claims about the intrinsic value of explanations may nonetheless be quite compelling, again drawing on the psychology of explanation. Finally, Section 5 concludes with a more speculative discussion of the implications of preceding claims for a ‘naturalized’ philosophy of explanation.

2. Explanations as Intrinsically and Instrumentally Valuable

Scientific theories are typically taken to support three core functions: prediction, explanation, and control. But explanation is frequently singled out as unique among these functions. For example, in the eloquent introduction to his book on explanation, Michael Strevens writes:

> If science provides anything of intrinsic value, it is explanation. Prediction and control are useful, and success in any endeavor is gratifying, but when science is pursued as an end rather than a means, it is for the sake of understanding – the moment when a small, temporary being reaches out to touch the universe and makes contact. (Strevens 2008)

Carl Hempel likewise distinguishes the value of explanation from that of prediction and control, with explanation as a distinct (and potentially secondary) motivation for the scientific enterprise:

> Among the divers factors that have encouraged and sustained scientific inquiry through its long history are two pervasive human concerns which provide, I think, the basic motivation for all scientific research. One of these is man’s persistent desire to improve his strategic position in the world by means of dependable methods for predicting and, whenever possible, controlling the events that occur in it … But besides this practical concern, there is a second basic motivation for the scientific quest, namely, man’s insatiable intellectual curiosity, his deep concern to know the world he lives in, and to explain, and thus to understand, the unending flow of phenomena it presents to him. (Hempel 1962)

These perspectives suggest a value for explanation independent from any potential contribution to prediction, control, or other practical benefits. Instead, explanations may be ends in themselves, or (merely) a means to satisfying a psychological need.

An alternative approach is to tie explanation very closely to prediction and control. On this view, explanations are valuable because they somehow contribute to these more tangible functions of scientific theories. In fact, Hempel is better known for advocating a position along these lines, suggesting that:

> It is this potential predictive force which gives scientific explanation its importance: only to the extent that we are able to explain empirical facts can we attain the major objective of scientific research … to anticipate new occurrences and to control, at least to some extent, the changes in our environment. (Hempel and Oppenheim 1948)

This perspective finds an earlier advocate in philosopher-psychologist Kenneth Craik, who proposes explanations as a means to anticipating and accommodating the future:

> It is clear that, in fact, the power to explain involves the power of insight and anticipation, and that this is very valuable as a kind of distance-receptor in time, which enables organisms to adapt themselves to situations which are about to arise. (Craik 1943)

And more humorously but with a common message, Quine and Ullian suggest that ‘the hypotheses we seek in explanation of past observations serve again in the prediction of future ones. Curiosity thus has survival value, despite having killed a cat’ (Quine and Ullian 1970). (See also Van Fraassen 1980, on pragmatic functions of explanation.)

Claims about scientific theories and explanations have analogs for human cognition. The parallels between science and cognition are highlighted by so-called 'child as scientist' or 'person as scientist' approaches (e.g., Kelley 1967; Carey 1985; Gopnik and Meltzoff 1997), according to which children and adults – like scientists – construct more or less coherent bodies of belief concerning particular phenomena, where these bodies of belief are responsive to empirical evidence and serve the critical functions of supporting prediction, explanation, and control. As with science, one can question the value of explanations for cognition and propose similar candidate answers. In psychology, most answers have followed Craik in positing an instrumental value for explanations. For example, Fritz Heider, an influential social psychologist, noted:

> If I find sand on my desk, I shall want to find out the underlying reason for this circumstance. I make this inquiry not because of idle curiosity, but because only if I refer this relatively insignificant offshoot event to an underlying core event will I attain a stable environment and have the possibility of controlling it. (Heider 1958)

In other words, explanations allow us to relate individual events to broader generalizations that support prediction and intervention, and thus serve an important instrumental function in guiding future interactions with the world. Lombrozo and Carey (2006) advocate a similar proposal, titled 'Explanation for Export', according to which explanations facilitate the generation of 'exportable' or broadly applicable beliefs by highlighting information likely to support future prediction and intervention (see also Lombrozo 2010).

If explanations have an instrumental value – whether in science or for everyday cognition – this value should manifest in consequences susceptible to empirical study. For science, explanation should have measurable effects on other aims of scientific inquiry, such as the efficiency with which theories are developed, the breadth and accuracy of their predictions, or their utility in supporting interventions. For human cognition, explanations should have measurable effects on behavior, potentially mediated by the efficient and effective development of useful intuitive theories. To my knowledge, this empirical approach to assessing the value of explanations has not been adopted within philosophy of science. However, a growing literature within psychology attests to the value of explanations for everyday cognition, with lessons that potentially generalize to science. The following section reviews key evidence from this burgeoning area of research.

3. Explanation as a Mechanism for Discovery and Confirmation

Explanation – the process – and explanations – the products – have been shown to have a variety of substantive cognitive consequences (for reviews, see Keil 2006; Lombrozo 2006, forthcoming). Engaging in explanation can facilitate learning (e.g., Chi et al. 1994, Williams and Lombrozo 2010), guide exploration (e.g., Legare forthcoming), alter which features of an item are deemed most important in assessing category membership (Ahn et al. 2002; Chin-Parker et al. 2006; Lombrozo 2009), and influence decision making (e.g., Hastie and Pennington 2000). Having an explanation and the content of that explanation can influence how a property is generalized from known to unknown cases (e.g., Sloman 1994; Rehder 2006; Lombrozo and Gwynne forthcoming), how data are interpreted (e.g., Koslowski 1996), and how subjective probabilities are assigned (e.g., Koehler 1991). These findings and others like them suggest an intimate relationship between explanation, learning, and inference.

Explanations clearly *reflect* an individual's beliefs about what is being explained, but in addition the very process of seeking, producing, and evaluating explanations plays a role in *generating and shaping* the content and structure of an individual's beliefs (Lombrozo 2006, forthcoming). In the traditional vocabulary of philosophy of science, explanation (both the product and the process) can contribute to discovery and confirmation, at least when it comes to intuitive theories in the minds of individual reasoners. This is a natural consequence of the hypothesis that explanation has instrumental value, whether explanations contribute to prediction and control directly, or indirectly via the generation of useful theories. What follows is a brief review of recent evidence supporting the claim that explanation impacts discovery and confirmation.

3.1. EXPLANATION AND DISCOVERY

The role of explanation in discovery is illustrated by a common experience: explaining something to someone else – or even to oneself – seems to generate greater understanding. This phenomenon is known as the self-explanation effect, and has been documented in both laboratory and classroom settings (for review, see Fonseca and Chi 2010). For example, compared with students who study instructional materials twice, students prompted to explain aspects of the circulatory system perform better on tests that assess information contained within the materials and also on those that assess information that can be inferred from the materials (Chi et al. 1994). Learning by explaining shares notable characteristics with thought experiments: the reasoner seems to discover something genuinely new in the absence of novel data from the external world (e.g., Gendler 1998).

What accounts for the effects of explanation on discovery? Here I focus on two potential approaches that can be motivated on the basis of theories of explanation from philosophy of science: subsumption and unification theories, on the one hand, and causal and causal mechanism theories, on the other (for review, see Woodward 2010). Both approaches assume substantive constraints on what counts as an adequate or satisfying explanation, and that these constraints in turn influence discovery via the inferences one draws, the evidence one seeks, and the way in which information is represented. The approaches differ in the constraints they emphasize: a preference for more unifying or subsuming regularities versus a focus on causal structure and mechanisms.

The first approach, called the *subsumptive constraints account*, proposes that explanations guide learners to interpret what they are trying to explain in terms of broad, unifying generalizations (Williams and Lombrozo 2010). Like subsumption and unification accounts of explanation in philosophy (e.g., Friedman 1974; Kitcher 1989), this account suggests that successful explanations relate what is being explained to general explanatory patterns or schemata. Explicitly representing such patterns or the relationship between such patterns and what is being explained (the explanandum) in turn results in a kind of discovery: the learner either 'discovers' properties or inferential consequences of what she already believed, or changes the format in which such beliefs are represented to make them more amenable to application in the current and future situations.

The subsumptive constraints account is supported by recent evidence from category learning. In a series of experiments reported in Williams and Lombrozo (2010), experimental participants were tasked with learning about two novel categories of robots from an alien planet. They were provided with four exemplars from each category, carefully constructed to exhibit two regularities: one that was relatively obvious and supported classification with 75% accuracy (called the 75% rule), and a more subtle regularity that supported classification with 100% accuracy (the 100% rule). As participants studied the exemplars, they were prompted to either *explain* their category membership, *describe* the members, *think aloud*, or engage in *free study*. The exemplars were then removed, and a variety of tests were administered to assess whether each participant had discovered the 75% rule, the 100% rule, or both. If explanation drives learners toward broader or more subsuming regularities, then participants who explain should be more likely than those in other conditions to discover the 100% rule, no matter that all participants were exposed to the same data.

As predicted, the results revealed that participants who explained were significantly more likely to discover the 100% rule than were those who described, thought aloud, or engaged in free study. While all participants were encouraged to discover some basis for categorization and a majority discovered the 75% rule, those who explained were more likely to persist in seeking and to find a more unifying regularity that supported classification in all cases. This suggests that engaging in explanation encourages learners to seek and adopt beliefs that support better – in this case more subsuming or unifying – explanations, and that participants do not always engage in this process spontaneously in the absence of a prompt to explain. A counterintuitive prediction of the subsumptive constraints account is therefore that in the absence of a subsuming or unifying pattern that can adequately account for all cases, prompting participants to explain can actually delay learning, and this is in fact the case (Williams et al. 2010).

In line with approaches to explanation that emphasize causation and causal mechanisms (e.g., Salmon 1984; Machamer et al. 2000), engaging in explanation can also promote the discovery of causal structure. This has been demonstrated both directly and indirectly. Indirect evidence comes from a growing developmental literature suggesting that explanations encourage children to posit unobserved causal factors. For example, Legare et al. (2009) presented preschool children with events in which a character in a vignette must choose between two food options, such as a vanilla milkshake and a strawberry milkshake, where one has been contaminated by a foreign substance, such as a leaf. Children were asked to both predict which item the character in the vignette would choose to consume, and to explain why a particular item was chosen. The authors found an 'explanation advantage', with children succeeding in effectively taking contamination into account more often in explanation than prediction. Most telling for present purposes, children often invoked causal factors in their explanations – such as germs or 'yucky stuff' – that they failed to spontaneously posit or consult as a guide to prediction. Such findings suggest that engaging in explanation can foster causal reasoning that does not occur spontaneously, and in this case generate a discovery, or more explicit recognition of the relationship between particular causal factors and behavior (see also Amsterlaw and Wellman 2006).

More direct evidence for the role of explanation in focusing learners on causal structure and causal mechanisms comes from a study in which preschool children learned about a novel mechanical device and were explicitly tested on their functional–mechanical understanding after a brief training period. Legare and Lombrozo (forthcoming) presented preschool children with a novel toy constructed from several colorful gears that connected a crank to a rotating fan. Children interacted with the toy for matched amounts of time while being prompted to either *explain* or *observe* the toy (additional conditions in which children explored the toy are not discussed here). The toy was then removed, and children's knowledge about it was assessed with several measures, where some involved functional–mechanical understanding (e.g., recognizing that gear size and shape were relevant for making the toy work), and others involved properties of the toy that were irrelevant to its causal structure and mechanism (e.g., remembering the color of a particular gear). The study revealed a striking interaction between training condition and learning, with children who explained outperforming those who observed on measures of functional–mechanical understanding, but those who observed outperforming those who explained on measures irrelevant to functional–mechanical understanding.

The cases considered so far – in particular, the findings from Williams and Lombrozo (2010), Williams et al. (2010), and Legare and Lombrozo (forthcoming) – reveal three related lessons. First, explanation is a powerful mechanism for learning. Merely being prompted to explain can influence the efficiency and content of learning. Second, however, explanation is not an all-purpose engine for discovery: under some conditions, explanation can delay discovery (Williams et al. 2010), and under some conditions, explanation privileges functional–mechanical understanding at the expense of memory for causally irrelevant details (Legare and Lombrozo forthcoming). Finally, the power and limitations of explanation go hand in hand. Engaging in explanation may influence discovery precisely by constraining the hypotheses a reasoner entertains, with consequences that can be either beneficial or detrimental depending on the relationship between those constraints and the structure one hopes to discover.

But do these lessons extend from relatively mundane, everyday beliefs to scientific theories? One reason to think so comes from the parallels between science and everyday cognition that motivate the 'person as scientist' perspective. A second comes from the simple observation that scientists are people: scientific reasoning must be a product of the same cognitive mechanisms that support everyday cognition, albeit potentially refined and influenced by scientific communities and institutions. Perhaps a more serious challenge comes from the fact that most laboratory experiments on the role of explanation in discovery involve simple changes in belief, not radical conceptual revisions of the kind that characterize scientific progress. Here, too, however, there's reason to endorse continuity with science. Effects of explanation have been shown to accelerate difficult conceptual transitions in childhood, including an understanding of numerical conservation (Siegler 2002) and acquisition of false belief understanding in theory of mind (Wellman and Lagattuta 2004; Amsterlaw and Wellman 2006). These examples of conceptual change in childhood provide a closer analog for scientific theory change (Carey 1985), and further reason to expect the influence of explanation in intuitive theory formation identified here to extend to discovery in science.

3.2. EXPLANATION AND CONFIRMATION

The previous section suggests that generating explanations can influence the hypotheses a reasoner discovers and entertains. But another role for explanation is in the evaluation of those hypotheses, and in particular the assignments of degrees of belief. This idea has roots in philosophy, and in particular in the proposal that an explanation's quality is a guide to its probability (see Lipton 2004). For example, Harman (1965) characterizes a process of 'inference to the best explanation' as follows:

> In making this inference, one infers, from the fact that a certain hypothesis would explain the evidence, to the truth of that hypothesis … one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true. (Harman 1965)

In other words, one assesses the merit of candidate hypotheses as explanations for a particular explanandum, and infers the truth of one hypothesis on the basis of this assessment.

If explanatory considerations figure in the evaluation of degrees of belief, or subjective probabilities, then the factors that influence the perceived quality of explanations should also serve as cues to the explanations' probability. A variety of such factors, or 'explanatory virtues', have been proposed, including an explanation's simplicity, scope, and fruitfulness. For the explanatory virtue studied most extensively to date, simplicity, the hypothesized relationship between an explanation's quality and its perceived probability is borne out.

Lombrozo (2007) presented adult participants with a task in which an alien's two symptoms could be explained either by appeal to a single disease, call it D_1, or to the conjunction of two diseases, D_2 and D_3 (see also Read and Marcus-Newhall 1993; Lagnado 1994). Quantifying simplicity as the number of causes invoked in an explanation, D_1 is the simpler explanation. Perhaps unsurprisingly, participants overwhelmingly preferred this explanation when it was just as likely as the conjunction of D_2 and D_3. However, a majority of participants continued to prefer the simpler explanation when the baserates for the three diseases were such that the conjunction of D_2 and D_3 was *more likely* than D_1. Participants seemed to treat simplicity as commensurate with frequency information, generating a pattern of judgments consistent with the hypothesis that they assigned a higher prior probability (about 80% versus 20%) to the simpler explanation, but updated this probability appropriately in light of the baserate information. A suitably modified task generated comparable results with preschool-aged children (Bonawitz and Lombrozo forthcoming). The findings suggest that in the face of probabilistic uncertainty, both children and adults use an explanation's simplicity is a guide to belief.

A variety of additional explanatory virtues have been found to influence explanatory preferences, although few studies consider the impact of these preferences on confirmation. For example, there is evidence that people prefer explanations that account for a larger number of observations (Thagard 1989; Read and Marcus-Newhall 1993; Preston and Epley 2005), that appeal to functions and goals (Kelemen 1999; Lombrozo et al. 2007), and that fit more coherently into a narrative structure (Pennington and Hastie 1992). For narrative structure, a suggestive finding relating explanatory preferences to subjective probability is that the order in which testimony is presented to mock jurors – a factor that should affect the ease with which the testimony can be used to reconstruct a temporal sequence – influences verdicts on the corresponding trial (Pennington and Hastie 1992). It is likely that other explanatory virtues similarly affect direct and indirect assessments of probability.

Going beyond explanatory virtues and their potential impacts on the probability of the explanans, a variety of psychological findings suggest that the mere existence of an explanation can influence the probability assigned to an *explanandum*. Specifically, explaining a hypothetical outcome or relationship (e.g., why people who are risk-seeking might make better firefighters) increases the subjective probability of that outcome or relationship (Anderson and Sechler 1986; see Koehler 1991 for review). Similarly, the content of generated or provided explanations can influence how probability is assigned to one claim in light of another (Sloman 1994; Rehder 2006; Lombrozo and Gwynne forthcoming). For example, suppose you learn about a novel reptile called a brollig, which eats a diet containing a mineral that causes stripes, and the stripes are an adaptation for camouflage. If you then learn that brolligs' stripes are very thin, how would this influence your belief in the claim that other reptiles – such as those that have stripes via a different mechanism, or those that have stripes for a different adaptive purpose – also have stripes that are very thin? Lombrozo and Gwynne (forthcoming) find that this judgment is guided by how an individual explains brolligs' stripes: participants who preferentially explain the stripes by appeal to the mechanism (diet) are more likely than those who explain by appeal to the function (camouflage) to generalize the property of thin stripes on the basis of shared mechanisms relative to shared functions.

In sum, the criteria by which explanations are evaluated inform the probability assigned to both explanans and explanandum. Explanations can also influence how belief in one claim is updated in light of belief in another. These findings suggest a close relationship between explanation and confirmation. In particular, explanatory considerations may serve as a cue to an explanation's plausibility or utility, especially in the absence of more direct or reliable cues. While it remains to be determined whether this reliance on explanatory considerations is warranted (see e.g., Lipton 2004), there is little question that as a matter of psychological fact, explanations and the mechanisms involved in their generation and evaluation do inform degrees of belief. And given the prevalence of inference to the best explanation in science, often involving explicit appeals to simplicity or other explanatory virtues in arbitrating theoretical disputes, it seems very likely that the psychological mechanisms uncovered in the lab operate in scientific practice.

3.3. INTERIM SUMMARY

The findings reviewed in this section suggest that explanation plays an important role in the discovery and confirmation of everyday beliefs, and by extension of intuitive and scientific theories. In particular, psychological mechanisms involved in the generation and evaluation of explanatory hypotheses may guide reasoners toward those hypotheses that make for better explanations, where an explanation's quality is influenced by the extent to which it subsumes or unifies, whether it invokes causation or causal mechanisms, and a host of explanatory virtues such as simplicity. As a result, explanation has widespread and systematic consequences for a reasoner's beliefs, consistent with the hypothesis that explanations have instrumental value and serve the functions of supporting discovery and confirmation.[1]

4. Why Claims about the Intrinsic Value of Explanations are So Compelling

Although everyday experience provides ample opportunity to observe the instrumental benefits of explanation for learning and generalization, the phenomenology of explanation may nonetheless incline people toward the view that explanations are ends in themselves. Both theoretical and empirical considerations suggest why this might be the case.

First, suppose we take seriously the proposal that explanations serve the function of supporting future prediction and intervention, and do so by facilitating the discovery and confirmation of useful theories. Because the mechanisms that underlie the generation and evaluation of explanations cannot assess an explanation's future utility directly, these mechanisms must instead be sensitive to properties of explanations that are reliable – if imperfect – cues to future utility. These cues may be precisely those properties of explanation that constrain discovery and confirmation: subsumption and unification, a privileged role for causal structure and causal mechanisms, simplicity, and other explanatory virtues. Reasoners may not have introspective access to *any* of the factors that underlie the generation or evaluation of explanations, but to the extent they do, they are likely to recognize these cues, and not the ultimate functions they serve, as the bases for explanatory preferences. As a result, explanations may appear valuable in virtue of their inherent properties, and not in virtue of their instrumental utility.

Second, the phenomenological satisfaction that accompanies an explanation may actually play a motivational role for engaging in important but effortful reasoning. This proposal has been developed most provocatively by Alison Gopnik (2000), who compares explanation to orgasm. Specifically, she suggests that just as orgasm provides an incentive to engage in an activity with a clear (evolutionary) instrumental value for an individual (i.e., reproduction), so the satisfaction of explanation provides an incentive to engage in the kind of theory formation that allows individuals to navigate a causally complex world. For this motivational system to work, individuals need not engage in sexual activity with the goal of reproducing – in fact, the proximate goal is more likely orgasm. Similarly, individuals need not engage in explanation with the goal of generating theories that support useful predictions and interventions, but may instead do so to achieve the phenomenological satisfaction that comes from an explanandum well explained. This, too, should contribute to the sense that explanations are pursued for their own sake.

Finally, a growing body of research suggests that individuals have relatively poor metacognition when it comes to explanation. That is, individuals are often blissfully unaware of the properties and shortcomings of their own explanations. Rozenblit and Keil (2002), for example, document a phenomenon they call the Illusion of Explanatory Depth. They find that people systematically overestimate the depth of their own explanatory understanding, be it for how a zipper zips or a helicopter flies. Relatedly, Trout (2002, 2007, 2008) argues that the sense of understanding that accompanies an explanation is a poor guide to the truth of that explanation. Interestingly, however, the act of generating explanations can improve metacognitive evaluations: people's overestimates of the quality of their own explanations is greatly reduced after being asked to provide said explanations (Rozenblit and Keil 2002), and explaining a text improves readers' accuracy in monitoring their own understanding (Griffin et al. 2008). So while the process of generating and evaluating explanations may actually improve an individual's ability to identify and use explanatory cues as a guide to discovery and confirmation, the unreflective *phenomenology* of explanation may not be firmly anchored to the properties of explanations or their consequences that would allow individuals to accurately track (and thus appreciate) their instrumental value.

These theoretical and empirical considerations help explain why explanations have a seductive appeal as intrinsically valuable ends in themselves. Of course, the fact that explanations also have instrumental value does not itself undermine the legitimacy of claims to intrinsic value. But recognizing the psychological factors that contribute to the plausibility of the latter position helps account for its intuitive appeal in the face of everyday experience and experimental evidence for widespread and important instrumental benefits.

5. Naturalizing Explanation: A Brief, Speculative Coda

The discussion so far has focused on descriptive claims about the role of explanation in everyday, human cognition. In some ways, these claims are closely tied to traditional questions in philosophy of science and epistemology. In particular, much of the research concerning explanation's contributions to discovery is motivated by – or at least consistent with – philosophical accounts of explanation that invoke unification or causation, with a similar relationship between inference to the best explanation and empirical research on explanation and confirmation. I've also suggested why empirical findings concerning individuals' intuitive theories are likely to extend to scientific explanations within communities of scientists. However, many theories of explanation from philosophy are not intended as descriptive claims about human cognition, but as normative claims about which explanations are in fact good or warranted, in science or in general. Does the research reviewed above have any implications for these more traditional projects? I conclude by briefly considering two implications: the first for a naturalized philosophy of explanation, the second for pluralism about explanation.

While approaches to naturalism vary, most share the commitment that philosophy should be continuous with the sciences. However, specifying the nature of the relationship between philosophical and scientific claims – especially when the philosophical claims involve normativity – is a challenge. One strategy is to identify the function of a particular concept or practice, and to derive a kind of instrumental normativity by considering what *best* satisfies that function. For example, it could be that explanations facilitate future prediction and intervention, or (as I've suggested here) play a role in discovery and confirmation, which in turn produces theories that support future prediction and intervention. Having identified explanation's key function or functions (F) empirically, one can then ask a quasi-normative question: given that explanations have the function of accomplishing F, what *ought* they to be like such that F is optimally achieved? More concretely, how ought we to generate or evaluate explanations as individuals to (say) maximize the benefits of our intuitive theories, and how ought we to generate and evaluate explanations as scientists to maximize scientific progress?

Considering this quasi-normative question provides a benchmark against which actual explanatory practices can be evaluated, and more generally supports a differentiation between descriptive and quasi-normative projects. I refer to these as 'quasi-normative' because they are nonetheless grounded in descriptive claims about the function or functions of explanation. In a sense, the normativity derives from a conditional: 'If explanations are to optimally contribute to F, how ought they to be generated or evaluated?' Empirical considerations, whether they come from psychology, the history or anthropology of science, or elsewhere, thus play a critical role in constraining this quasi-normative approach.

One virtue of this proposal is that it situates explanation with respect to other aspects of science and inference, many of which are arguably better developed. In a recent review of the philosophy of scientific explanation, James Woodward advocates a similar goal, noting that:

> … writers on explanation have not always paid adequate attention to how explanation itself is connected to or interacts with (or is distinct from) other goals of inquiry – for example, what the connection is between explanatory goodness and other frequently proposed goals for inquiry such as evidential support, prediction, control of nature, simplicity, and so on. One result is that it is sometimes unclear how to assess the significance of our intuitive judgments about the goodness of various explanations or to determine what turns on our giving one judgment rather than another. (Woodward 2010)

Another contemporary writer, Heather Douglas, likewise ties explanation very closely to other goals of inquiry, in this case to prediction:

> What makes an explanation scientific is not that it fits within one of these particular models or that it avoids some of the conceptual pitfalls that have littered the explanation landscape over the past decades. What makes an explanation scientific is that it is useful for producing that other important goal of science: testable predictions. (Douglas 2009)

Douglas goes on to draw a valuable insight concerning the prospects for a unified or universal theory of explanation. If explanations are functionally defined in terms of their contributions to testable predictions (on Douglas' view) or the discovery and confirmation of useful theories (on mine), there is no strong reason to expect all explanations to conform to a common structure. Causal, subsumption, unification, and causal mechanism theories of explanation may all identify legitimate kinds of explanations that contribute to these functions, perhaps in different ways or in different contexts. Such considerations open the door to a robust explanatory pluralism.

These concluding remarks do little more than gesture toward promising directions in need of further development. I present them here in the spirit of furthering a conversation about the relationship between empirical evidence and philosophy in general, and about the prospects for a naturalized philosophy of explanation in particular. A promising first step is to recognize the instrumental value of explanations both for science and for everyday cognition.

Acknowledgement

The author thanks M. Strevens and J. D. Trout for comments on an earlier draft of this paper.

Short Biography

Tania Lombrozo's research combines the empirical tools of cognitive psychology with the conceptual tools of analytic philosophy to address foundational questions about human cognition. Her work focuses on explanation, causation, conceptual representation, and moral reasoning, with publications in such journals as *Cognition*, *Cognitive Psychology*, *Psychological Science*, and *Trends in Cognitive Sciences*. In 2009, she received the Mary C. Potter Award from Women in Cognitive Science (WICS) and, in 2010, the Stanton Prize from the Society for Philosophy and Psychology, both recognizing early career contributions. Lombrozo is currently an Assistant Professor in the Department of Psychology at the University of California, Berkeley, as well as an affiliate of the Department of Philosophy and a member of the Institute for Cognitive and Brain Sciences. She holds a PhD in Psychology from Harvard University, and a BA in Philosophy and a BS in Symbolic Systems from Stanford University.

Notes

* Correspondence: Department of Psychology, University of California, Berkeley, 3210 Tolmon Hall, Berkeley, CA 94720, USA. Email: lombrozo@berkeley.edu.

[1] On many accounts of 'function', to claim that these consequences are the *function* of explanation requires a further commitment that the mechanisms that underlie these effects result either from natural selection or from learning (e.g., Wright 1976; Lombrozo and Carey 2006). Defending such a commitment goes beyond the scope of the current paper; for present purposes, it suffices to note that such consequences are consistent with the hypothesis that explanations have instrumental value.

Works Cited

Ahn, W., et al. 'Effect of Theory-Based Feature Correlations on Typicality Judgments.' *Memory & Cognition* 30 (2002): 107–18.

Amsterlaw, J. and H. Wellman. 'Theories of Mind in Transition: A Microgenetic Study of the Development of False Belief Understanding.' *Journal of Cognition and Development* 7 (2006): 139–72.

Anderson, C. A. and E. S. Sechler. 'Effects of Explanation and Counterexplanation on the Development and Use of Social Theories.' *Journal of Personality and Social Psychology* 50 (1986): 24–34.

Bonawitz, E. B. and T. Lombrozo. 'Occam's Rattle: Children's Use of Simplicity and Probability to Constrain Inference.' forthcoming.

Carey, S. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press, 1985.

Chi, M. T. H., N. de Leeuw, M. H. Chiu and C. LaCancher. 'Eliciting Self-Explanations Improves Understanding.' *Cognitive Science* 18 (1994): 439–77.

Chin-Parker, S., O. Hernandez, and M. Matens. 'Explanation in Category Learning.' *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Eds. R. Sun and N. Miyake. Mahwah, NJ: Erlbaum, 2006. 1098–103.

Craik, K. *The Nature of Explanations*. Cambridge, UK: Cambridge University Press, 1943.

Douglas, H. E. 'Reintroducing Prediction to Explanation.' *Philosophy of Science* 76 (2009): 444–63.

Fonseca, B. A. and M. T. H. Chi. 'Instruction Based on Self-Explanation.' *The Handbook of Research on Learning and Instruction*. Eds. R. Mayer and P. Alexander, New York, NY: Routledge Press, 2010. 296–321.

Friedman, M. 'Explanation and Scientific Understanding.' *Journal of Philosophy* 71 (1974): 5–19.

Gendler, T. S. 'Galileo and the Indispensability of Scientific Thought Experiments.' *British Journal for the Philosophy of Science* 49 (1998): 397–424.

Gopnik, A. 'Explanation as Orgasm and the Drive for Causal Knowledge: The Function, Evolution, and Phenomenology of the Theory-Formation System.' *Explanation and Cognition*. Eds. F. Keil and R.A. Wilson. Cambridge, MA: MIT Press, 2000. 299–324.

—— and A. Meltzoff. *Words, Thoughts and Theories*. Cambridge, MA: MIT Press, 1997.

Griffin, T. D., J. Wiley, and K. W. Thiede. 'Individual Differences, Rereading, and Self-Explanation: Concurrent Processing and Cue Validity as Constraints on Metacomprehension Accuracy.' *Memory & Cognition* 36 (2008): 93–103.

Harman, G. 'The Inference to the Best Explanation.' *Philosophical Review* 74 (1965): 88–95.

Hastie, R. and N. Pennington. 'Explanation-Based Decision Making.' *Judgment and Decision Making: An Interdisciplinary Reader*. 2nd ed. Eds. T. Connolly, H. R. Arkes, and K. R. Hammond. Cambridge, UK: Cambridge University Press, 2000. 212–28

Heider, F. *The Psychology of Interpersonal Relations*. New York, NY: Wiley, 1958.

Hempel, C. 'Explanation in Science and in History.' *Frontiers of Science and Philosophy*. Ed. R. G. Colodny. Pittsburgh, PA: University of Pittsburgh Press, 1962. 7–33.

Hempel, C. G. and P. Oppenheim, 'Studies in the Logic of Explanation.' *Philosophy of Science* 15 (1948): 135–75.

Keil, F. C. 'Explanation and Understanding.' *Annual Review of Psychology* 57 (2006): 227–54.

Kelemen, D. 'Function, Goals and Intention: Children's Teleological Reasoning about Objects.' *Trends in Cognitive Sciences* 3 (1999): 461–8.

Kelley, H. H. 'Attribution Theory in Social Psychology.' *Nebraska Symposium on Motivation*, Vol. 15. Ed. D. Levine. Lincoln, NE: University of Nebraska Press, 1967. 192–240.

Kitcher, P. 'Explanatory Unification and the Causal Structure of the World.' *Scientific Explanation*. Eds. P. Kitcher, W. Salmon. Minneapolis, MN: University of Minnesota Press, 1989. 410–505.

Koehler, D. J. 'Explanation, Imagination, and Confidence in Judgment.' *Psychological Bulletin* 110 (1991): 499–519.

Koslowski, B. *Theory and Evidence: The Development of Scientific Reasoning*. Cambridge, MA: MIT Press, 1996.

Lagnado, D. 'The Psychology of Explanation: A Bayesian Approach.' Masters Thesis, Schools of Psychology and Computer Science, University of Birmingham, 1994.

Legare, C. H. 'Exploring Explanation: Explaining Inconsistent Evidence Informs Exploratory, Hypothesis-Testing Behavior in Young Children.' *Child Development* (forthcoming).

—— and T. Lombrozo. 'The Unique and Selective Benefits of Explanation for Learning in Early Childhood.' forthcoming.

——, H. M. Wellman, and S. A. Gelman. 'Evidence for an Explanation Advantage in Naïve Biological Reasoning.' *Cognitive Psychology* 58 (2009): 177–94.

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

Lombrozo, T. 'Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions.' *Cognitive Psychology* 61 (2010): 303–32.

——. 'Explanation and Abductive Inference.' *The Oxford Handbook of Thinking and Reasoning*. Eds. K.J. Holyoak and R.G. Morrison. Oxford, UK: Oxford University Press, forthcoming.

——. 'Explanation and Categorization: How "why?" Informs "What?".' *Cognition* 110 (2009): 248–53.

——. 'Simplicity and Probability in Causal Explanation.' *Cognitive Psychology* 55 (2007): 232–57.

——. 'The Structure and Function of Explanations.' *Trends in Cognitive Sciences* 10 (2006): 464–70.

—— and S. Carey. 'Functional Explanation and the Function of Explanation.' *Cognition* 99 (2006): 167–204.

—— and N. Gwynne. 'Explanation and Inference: Functional and Mechanistic Explanations Guide Property Generalization.' forthcoming.

——, D. Kelemen, and D. Zaitchik. 'Inferring Design: Evidence of a Preference for Teleological Explanations in Patients With Alzheimer's Disease.' *Psychological Science* 18 (2007): 999–1006.

Machamer, P., L. Darden, and C. F. Craver. 'Thinking about Mechanisms.' *Philosophy of Science* 67 (2000): 1–25.

Pennington, N. and R. Hastie. 'Explaining the Evidence: Tests of the Story-Model for Juror Decision Making.' *Journal of Personality and Social Psychology* 62 (1992): 189–206.

Preston, J. and N. Epley. 'Explanations Versus Applications: The Explanatory Power of Valuable Beliefs.' *Psychological Science* 16 (2005): 826–32.

Quine, W. V. O. and J. S. Ullian. *The Web of Belief*. New York: Random House, 1970.

Read, S. J. and A. Marcus-Newhall. 'Explanatory Coherence in Social Explanations: A Parallel Distributed Processing Account.' *Journal of Personality and Social Psychology* 65 (1993): 429–47.

Rehder, B. 'When Causality and Similarity Compete in Category-Based Property Induction.' *Memory & Cognition* 34 (2006): 3–16.

Rozenblit, L. R. and F. C. Keil. 'The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth.' *Cognitive Science* 26 (2002): 521–62.

Salmon, W. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press, 1984.

Siegler, R. S. 'Microgenetic Studies of Self-Explanations.' *Microdevelopment: Transition Processes in Development and Learning*. Eds. N. Granott and J. Parziale. New York, NY: Cambridge University, 2002. 31–58.

Sloman, S. A. 'When Explanations Compete: The Role of Explanatory Coherence on Judgments of Likelihood.' *Cognition* 52 (1994): 1–21.

Strevens, M. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press, 2008.

Thagard, P. 'Explanatory Coherence.' *Behavioral and Brain Sciences* 12 (1989): 435–67.

Trout, J. D. 'The Psychology of Scientific Explanation.' *Philosophy Compass* 2.3 (2007): 564–91.

——. 'Scientific Explanation and the Sense of Understanding.' *Philosophy of Science* 69 (2002): 212–33.

——. 'Seduction Without Cause: Uncovering Explanatory Neurophilia.' *Trends in Cognitive Sciences* 12 (2008): 281–2.

Van Fraassen, B. *The Scientific Image*. Oxford, UK: Oxford University Press, 1980.

Wellman, H. M. and K. H. Lagattuta. 'Theory of Mind for Learning and Teaching: The Nature and Role of Explanation.' *Cognitive Development* 19 (2004): 479–97.

Williams, J. J. and T. Lombrozo. 'The Role of Explanation in Discovery and Generalization: Evidence from Category Learning.' *Cognitive Science* 34 (2010): 776–806.

——, ——, and ——. 'Why Does Explaining Help Learning? Insight from an Explanation Impairment Effect.' *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Eds. S. Ohlsson and R. Catrambone. Austin, TX: Cognitive Science Society, 2010. 2906–11.

Woodward, J. 'Scientific Explanation.' *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*. Ed. E. N. Zalta, 2010 [Online]. Retrieved on 15 April 2011 from: http://plato.stanford.edu/archives/spr2010/entries/scientific-explanation/.

Wright, L. *Teleological Explanation*. Berkeley, CA: University of California Press, 1976.