Experimental philosophy of explanation rising.
The case for a plurality of concepts of *explanation*

Abstract This paper brings together results from the philosophy and the psychology of explanation in order to argue that there are multiple concepts of *explanation* in human psychology. Specifically, it is shown that pluralism about *explanation* coheres with the multiplicity of models of explanation available in the philosophy of science, and is supported by evidence from the psychology of explanatory judgment. Focusing on the case of a norm of explanatory power, the paper concludes by responding to the worry that if there is a plurality of concepts of *explanation*, one will not be able to normatively evaluate what counts as good explanation.

1. Introduction

While all explanations answer some why- or how-question, significant variation is observed across contexts in what is accepted as an explanation, in what type of explanatory information is sought, and in what norms are assumed to govern good explanation. This apparent variation motivates a central question in the philosophy and psychology of explanation: How many concepts of *explanation* do we have in our psychology?

According to pluralists, we have more than one concept of explanation; and this plurality is reflected in the plurality of philosophical models of explanation. Monists oppose pluralism claiming that we have only one concept of explanation; the plurality of models observed in philosophy would be grounded in this one concept. Monism appears to be the predominant view in the philosophy of science, where all major models of explanation "are 'universalist' in aspiration—they claim that a single, 'one size' model of explanation fits all areas of inquiry in so far as these have a legitimate claim to explain" (Woodward 2014).[1] A third answer comes from eliminativists, according to whom there is no distinct concept of explanation in our psychology; the variety of philosophical models would be grounded in a heterogeneous, indistinct suite of psychological processes and representations used and reused in a variety of complex capacities spanning causal, confirmation-theoretic, counterfactual, deductive, inductive, and probabilistic reasoning.

In what follows, I am going to scout some of the work in the philosophy and psychology of explanation that supports pluralism. I draw on results from what can be called *experimental philosophy of explanation* with the hope of furthering a conversation "about the prospects for a naturalized philosophy of explanation" (Lombrozo 2011, 549).

Specifically, I begin with the idea that philosophical models of explanation ought to be calibrated on people's explanatory judgments (Section 2). When room is made for psychology to constrain philosophical models of explanation, pluralism will be strongly supported by results in psychology, and in philosophy (Section 3). I conclude by showing that pluralism does not entail that

[1] This claim requires qualification. Some philosophical models of explanation are intended to be tailored to particular disciplines or domains of inquiry. Some model is intended to apply *only* to explanation in mathematics (e.g., Steiner 1978)—assuming that there are explanations in mathematics (Mancosu 2001). The mechanistic model of explanation is intended to be tailored to explanation in the life sciences, which are thought to be organized around the goal of supplying mechanistic explanations (Bechtel & Richardson 1993; Machamer, Darden, & Craver, 2000; Craver 2007). While the mechanistic model does not claim that it fits all domains of inquiry, it arguably remains universalist in aspiration to the extent that it aims to fit all adequate explanations in the domains of biology or of neuroscience (e.g., Kaplan & Craver 2011).

no model or norm of explanation can be justified. This conclusion is borne out by the case of a norm of explanatory power (Section 4).

2. Philosophical explications of *explanation*

No philosophical model of explanation incorporates empirical evidence from the psychology of explanation as a serious constraint on theorizing. This is a deficiency. Results from studies of explanatory cognition ought to be used at least to calibrate the method of explication, which is traditionally employed to develop philosophical models of explanation.

All main philosophical models of explanation aim to provide explications of the concept of *explanation* (Hempel & Oppenheim 1948; Hempel 1965; Salmon 1984; Kitcher 1989; Woodward 2003; Craver 2007; Strevens 2008). These models purport to clarify the meaning and usage of the word 'explanation,' taking account of practices of explanation in scientific and non-scientific contexts, and paradigmatic examples of explanation. The tendency "has been to assume that there is a substantial continuity between the sorts of explanations found in science and at least some forms of explanation found in more ordinary non-scientific contexts," which explains why philosophical models of explanation often move back and forth between scientific examples and more homey cases of explanation (Woodward 2014).[2] Philosophical models also aim to regiment usage of 'explanation,' making recommendations about what one ought to mean by an explanatory claim, about what counts as good and bad explanation and about what purposes explanation ought to serve. Philosophical models are not purely normative, neither are they simply descriptive.

A preliminary step in the task of explicating *explanation* is to clarify what kind of concept the philosophical model aims to target. A model may aim at finding an explicatum of 'explanation' not as used in the English phrases 'explaining a meaning' or 'explaining a recipe,' but as used in the phrases 'explaining why some objects float in water and others sink' or 'explaining why the tides are higher during a full moon.'

Once an explicandum is sufficiently clear, philosophers proceed to engineer it into a model that strikes the right balance between similarity to the explicandum in conforming to prior usage, simplicity, precision, and fruitfulness in serving a variety of epistemic and practical purposes (Carnap 1950). In this process of explication, seldom is a model of explanation held accountable to people's relevant explanatory judgments. The standard approach is instead to attend the judgments made by philosophers on a battery of scenarios that involve putative explanations.

Unfortunately, philosophers' judgments are idiosyncratic, and can differ significantly from the explanatory judgments made by non-philosophers in everyday circumstances. If the method of explication is to track some properties of common usage of 'explanation' or of some hypothesised concept of *explanation* in human psychology, then philosophical explication ought to be calibrated on data about actual explanatory practice from cognitive science, but also from the history and sociology of science; in particular, it ought to be calibrated on ordinary people's explanatory judgments, not just on philosophers' intuitions. For it is ordinary judgments that most forthrightly express common usage of 'explanation.'

[2] The assumption that scientific explanation is substantially continuous with ordinary explanation motivates one to take both of two routes to developing a model of explanation. The route from an a priori model of explanation to taking into account instances of actual scientific explanations; and the route from taking into account actual scientific explanations to taking into account general psychological models of explanation.

Three more specific reasons suggest that philosophical explication ought to be calibrated on ordinary people's explanatory judgments. First, these judgments have a valuable role to play in *explication preparation*. For example, they can provide philosophers with evidence helpful to identify the explicatum's central features and its relation with other concepts; or they can point to sources of bias that affect judgment about alleged cases of genuine explanation (Shepherd & Justus 2014, Section 3; Waskan et al. 2014).

Second, psychological results can help philosophers assess to what extent a model of explanation is instrumentally rational, and hence fruitful (Lombrozo 2011; see Woodward 2012 on *causation*). This assessment will be made on a case-by-case fashion, after the actual function (or functions) served by the explicatum concept is identified in specific contexts. It may also involve a diagnosis of how the function of a certain concept of *explanation* can be biased, or improved (Bishop & Trout 2005).

Third and finally, philosophical explication ought to be calibrated on ordinary people's explanatory judgments because most philosophical models of explanation have commitments about how people actually think about explanation. Even Carl Hempel, who adopted an anti-psychologistic stance, wrote that "[l]ike any other explication, the construal here put forward has to be justified by… [showing that it] does justice to such accounts as *are generally agreed* to be instances of scientific explanation" (Hempel 1965, 489). Obviously, what cases "are generally agreed to be instances of scientific explanation" and to what extent the explication accords with such cases are descriptive questions that should be settled empirically.

Just to give some other examples. In introducing his theory of causal explanation, Woodward (2003) explains that "a significant portion [of his project] does involve a description of ordinary and scientific usage and judgment" (7). Lipton (2004) is explicit that his main aim is to put forward Inference to the Best Explanation "as a partial answer to the descriptive problem … of giving a principled account of the way we actually go about making non-demonstrative inferences" (142). Strevens (2008) focuses his account of explanation so that his aim is "purely descriptive:" the aim is "to say what kinds of explanations we give and why we give them" (37).

The precise ways in which philosophical models take facts about our psychology as relevant, and the sorts of commitments they make about how people think about explanation are not straightforward. However, all models appear to make some empirical claims about actual explanatory practice, and about explanatory judgment in particular, but they do not engage seriously with empirical studies of explanation.

3. Pluralism about *explanation*

If empirical evidence about explanatory cognition is taken seriously, then the multiplicity of models of explanation available in philosophy of science would be easily accounted for by the plurality of concepts of *explanation* in human psychology, and any universalist aspiration of a model of explanation should be abandoned. To substantiate that there is a plurality of concepts of *explanation*, I proceed in three steps. First, I discuss empirical evidence that indicates that the folk can distinguish *explanation* from other related epistemic concepts like *confirmation* and *logical implication*, which suggests that people possess some distinct concept of *explanation*. Then, after outlining three familiar types of explications of *explanation*, I show that all three find empirical support within psychology. Finally, I provide evidence that explanation has multiple psychological functions that cannot be reduced to any one model.

3.1. Explanation is distinguished from other related concepts

Colombo et al. (2015) investigated under what circumstances reasoners ascribe explanatory value to a hypothesis. Specifically, they asked whether experimental participants distinguish the explanatory value of a hypothesis from its posterior probability. In their study, each participant was selectively given information about a battery of situations, where one hypothesis was given as a possible explanation of observed data. Across situations, Colombo and colleagues manipulated statistical relevance relations or causal relations (or both) between the hypothesis and the data. Participants were then asked to make a series of explanatory judgments along seven dimensions, including judgments about the explanatory value of the hypothesis, about its posterior probability and acceptability, as well as about its logical, causal and cognitive relation to the data.

Prior to the analysis of the effects of their manipulations, Colombo and colleagues explored the interdependencies of the seven dependent variables in the response questionnaire, determining whether the participants could distinguish those dimensions. A principal component analysis showed that the seven response variables could be decomposed into three constructs that could explain together most of the variation in the data: *Cognitive Salience* (primarily loaded with the response variables Causality, Explanatory Value and Understanding), *Rational Acceptability* (Posterior Probability, Confirmation and Truth) and *Entailment* (Logical Implication).

Furthermore, Colombo and colleagues found that explanatory value is a complex psychological phenomenon affected by both causal priming, and by information about logical and statistical relevance relations. However, these effects were triggered by different circumstances. Where no obvious causal mechanism appeared to be in place, judgments of explanatory value were selectively sensitive to relations of statistical relevance and of logical entailment between explanans and explanandum. In particular, the likelihood of the explanans drove judgments of explanatory value. In situations where a potentially great number of alternative explanations competed, the prior credibility of specific causal hypothesis drove the explanatory value of the hypothesis, regardless of its logical and statistical relevance relations to the target explanandum.

Two conclusions underwritten by Colombo et al.'s (2015) results are that reasoners can neatly distinguish between confirmatory and explanatory value of a hypothesis (see also Brem & Rips 2000), which suggests that people possess some distinct concept of *explanation*, and that explanatory judgment is subtly sensitive to variations in the probabilistic and causal information available across situations.

3.2. Psychological evidence supports three explications of explanation

Starting with the pioneering work of Hempel (1965), philosophers of science have articulated three main models of explanation. Hempel's (1965) explication imposes logical constraints on what constitutes an explanation. According to his *deductive-nomological* model, explanations are logical arguments demonstrating how what is being explained follows deductively from some general laws and empirical conditions. While the deductive-nomological model has been dismissed by many philosophers on the basis of intuitive counterexamples, Friedman's (1974) and Kitcher's (1981) explications also emphasize the logical character of explanation. According to their *unificationist* model, explanations show the explanandum to be an instance of a general argument-pattern that can be comprehensively applied to many different phenomena. Following Salmon's (1984) lead, several contemporary models provide a third type of explication, which imposes causal constraints on what constitutes an explanation (Machamer et al. 2000; Woodward 2003; Strevens 2008). According to

the *causal-mechanical* model, explanations reveal organized component parts and causes that bring about explananda phenomena.

The debate has often been conducted as if there is one true model of explanation, and a decision should be made about which one of these three explications is the best. However, it is plausible that each one of these models is correct in capturing some salient aspect of explanatory practice. The concepts of *explanation* corresponding to the three models singled out above all appear to have a place in human psychology, tracking distinct representations and processes.

The deductive-nomological and the unificationist models, which both impose logical constraints on explanation, accord with the fact that in some tasks experimental participants tend to subsume particular instances of what is to be explained to broader principles and regularities; they do not look for causal information. Specifically, Williams & Lombrozo (2010) hypothesized that if sometimes explanation prompts one to find unifying or subsuming regularities instead of causal mechanisms, then, under some circumstances, experimental participants who are asked to produce explanations should be more likely to look for encompassing regularities that subsume a large proportion of target cases under a certain category.

To test this hypothesis, Williams & Lombrozo (2010) used a categorization-learning task where some participants were prompted to *explain* their categorizations, others to *describe* category members, others to *think aloud*, and yet others to engage in *free-study*. Compared to control conditions, participants that *explained* were more likely to seek more unifying, subsuming regularities, which supports the idea that explanation sometimes prompts one to relate what is being explained to encompassing argumentative-patterns, general principles or theories, rather than to causal mechanisms (see also Keil 2006, 229-30; Legare 2014).

The psychological literature on causal reasoning is impressive (Lagnado et al. 2007). Much interest has been attracted by causal induction, where experimental participants are asked to infer the presence of an unknown causal link from information about potential mechanisms (e.g., Cheng 1997; Griffiths & Tenenbaum, 2005). Evidence more relevant for assessing the psychological adequacy of the causal-mechanical model of explanation indicates that causal and mechanistic properties are often the most salient when participants are asked to explain some target phenomenon (Waldmann 1996; Lombrozo & Carey 2006). Asking experimental participants to explain some phenomenon often prompts them to cite causal and mechanistic properties relevant to the production of the phenomenon (Walker et al. 2014).

So, while explanation sometimes engages deductive reasoning and theory-like representations in human psychology, explanation often recruits inductive reasoning and information about causal mechanisms too. Each explication highlights constraints on what is being explained that correspond to distinct processes and representations in human psychology and that have unique cognitive consequences across different explanatory contexts.

3.3. Multiple cognitive functions of explanation

Explanation has a variety of cognitive functions. It constrains underdetermined problems and grounds reliable generalization; it facilitates learning and discovery, and plays a central role in confirmation and inference (Lombrozo 2006). With respect to confirmation, explanatory considerations can contribute to make some hypotheses more credible; and, more generally, they guide the assignments of subjective probabilities to propositions. Koehler (1991) reviewed a set of findings concerning how explanation influences subjective probabilities, and argued that merely focusing on a hypothesis as if it were the true explanation of some observed data is sufficient to

boost the subjective probability assigned to that hypothesis. Explanation can influence how probabilities are assigned to one proposition in the light of another. For instance, Sloman (1994) found that a proposition boosted the probability assigned to another proposition if they shared an explanation. Furthermore, there is evidence that epistemic virtues such as the simplicity, coherence, or the breadth of a potentially explanatory hypothesis can influence its perceived probability. Lombrozo (2007), for example, found that experimental participants rely on the simplicity of a potentially explanatory hypothesis as a cue commensurate to base-rate information in the face of probabilistic uncertainty.

With respect to learning, generating explanations to oneself (self-explaining) or to others can facilitate the integration of new information into existing bodies of knowledge, can lead to deeper processing, and promote greater understanding. Tutors often learn more than their students by explaining and responding to explanatory questions (Roscoe & Chi 2008). Children's normal developing understanding of minds relates to how often their mothers explain people's behaviour in terms of beliefs and desires (Peterson & Slaughter 2003). Performance on a variety of reasoning tasks, including logic, probabilistic and categorization tasks, can be enhanced by being prompted to explain (Fonseca & Chi 2011).

These and other results demonstrate that explanation has several distinct functions across several cognitive domains. This plurality of functions mirrors a plurality of explications. Each explication coheres with only some results about the relation between explanation and cognitive function, which suggests that these multiple functions cannot be reduced to any one model of explanation (on the relation between learning and two different explications of explanation, see Williams & Lombrozo 2010; and Legare & Lombrozo 2014).

Taken together, the plurality observed in philosophy and in the psychology of explanatory judgment is best accounted for by pluralism. If pluralism is true, and our psychology features a plurality of concepts of explanation, then we should observe widespread plurality in results from both the philosophy and the psychology of explanation. Since this plurality is just what is observed, pluralism is vindicated.[3]

4. How to assess a model of explanation in the face of pluralism

If pluralism is true, how can we justify different philosophical models of explanation or different norms of explanation?

4.1. Reflective equilibrium and theories of explanation

Many philosophers of science seem to implicitly employ something akin to reflective equilibrium when they seek to prove justification for theories of explanation (Goodman 1983; Stich 1990, Ch. 4; Rawls 1971, Ch. 1). When applied to a model of explanation, seeking reflective equilibrium consists in working back and forth among judgments about particular historical or fictitious cases of explanation, about the norms that according to the model govern judgments of these cases, and

[3] Monism might still be true. One possibility is that there is only one concept of *explanation* in our psychology, but many different explanation properties, which satisfy this concept locally in different domains of inquiry. Another possibility is that there is only one concept of *explanation*, and this concept is a second-order functional concept that gets applied to a plurality of first-order functional concepts. While these possibilities are intriguing, they are difficult to assess on the basis of available evidence; and, given the subtleties they involve, it may not be straightforward to test them empirically.

about the theoretical, empirical and pragmatic considerations that are believed to bear on particular examples of explanation. The coherence of this system of judgments is increased by the presence of inferential and evidential relations between the judgments in the system; and the increase in degree of coherence is a function of the number and strength of such relations. Lack of inferential or evidential relations between these judgments will diminish the coherence of the system. When an acceptable degree of coherence is reached among all the judgments, reflective equilibrium is reached, and the philosophical model of explanation earns positive epistemic status.

Each of the judgments entering the process are open to substantial revision in the light of their mutual degree of coherence, in the light of independently justified theoretical desiderata, and of changes in the practical goals of scientific inquiry. Since no judgment is immune from revision over time, a given norm of explanation may receive more or less justification as new or different considerations enter the process altering the inferential or evidential relations within the system. No one norm of explanation is likely to rest justified in all historical settings. And no one model of explanation will be able to serve as unique explanatory ideal applicable to all contexts.

In order to seek reflective equilibrium for a model of explanation, the core questions are: What are the norms of explanation embedded in the model? What are the theoretical, empirical and pragmatic considerations that support (or fail to support) them? How important are these different considerations? To address these questions and begin illustrate the method of reflective equilibrium, I concentrate on the norm of *explanatory power*.

4.2. Justifying a norm of explanatory power: theoretical considerations

According to a norm of *explanatory power*, the goodness of an explanation should be a function of its power. The framework of the Bayesian theory of probabilistic inference can be used to make this norm more precise (e.g., Good 1960; McGrew 2003; Schupbach & Sprenger 2011; Crupi & Tentori 2012). The common pattern of analysis begins from a set of intuitively desirable adequacy conditions, and then determines a unique measure of explanatory power that satisfies such conditions. For example, Schupbach & Sprenger (2011) propose that the explanatory power that a hypothesis H has over an explanandum e should be measured as $(P(H|e) - P(H|\sim e)) / (P(H|e) + P(H|\sim e))$. For any hypothesis known to provide an explanation of some phenomenon, this measure reveals how powerful that explanation is.

Once the norm is made precise, we may begin to examine the degree to which it coheres with theoretical commitments and desiderata of different models of explanation. If we interpret a norm of explanatory power with Schupbach & Sprenger (2011, 108) as "a hypothesis's ability to decrease the degree to which we find the explanandum surprising," then this norm will nicely cohere with the Deductive-Nomological model, which can be understood as claiming that an explanatory hypothesis will necessarily increase the degree to which the explanandum is expected (Hempel 1965). Instead, if we interpret a norm of explanatory power with Ylikoski & Kuorikoski (2010), in terms of the ability of an explanatory hypothesis to provide answers to many what-if-things-had-been-different questions, then the norm will be mostly coherent with causal models of explanation that understand causality in contrastive-counterfactual terms (e.g., Woodward 2003).

However, we may not all agree that an analysis of a given norm of explanation is theoretically adequate. The sources of disagreement can be twofold. First, there may be disagreement about whether a norm of explanation should be analysed within a particular framework. For instance, while Schupbach & Sprenger (2011) conducted their analysis of explanatory power pursuing a Bayesian approach, Ylikoski & Kuorikoski (2010) provided an

informal, qualitative analysis pursuing a contrastive-counterfactual approach. The fact that different frameworks can be used to analyze a norm of explanation and that analyses within different frameworks cohere better with some model of explanation and not with others only indicates that there is no single model of explanation that has theoretical advantages over all others.

Second, within the same framework, there may be disagreement about particular theoretical desiderata that the analysis should satisfy. One of the conditions of adequacy proposed by Schupbach & Sprenger (2011), for example, postulates that explanatory power should not depend on the prior plausibility of the explanans *H*, if either *H* or ~*H* implies the explanandum *e*. While this condition is not shared by alternative analyses, critics and advocates alike rely just on intuition to argue about it (Crupi & Tentori 2012). Relying on empirical evidence is one way to resolve this type of disagreement.

4.3. Justifying a norm of explanatory power: empirical considerations

If a norm of explanation is systematically breached, then the norm fails to capture actual explanatory practice, and this will undermine the justification for models relying on it. For instance, people may not follow a certain norm of explanation when they make explanatory judgments. From present day viewpoint, judgments of the explanatory power of historical scientific test-cases may not be predicted by the norm. Furthermore, an expression like 'explanatory power,' as it is generally used in current, everyday English, may have a meaning significantly different from the meaning assumed by the philosophical analysis; and some of the conditions put forward by the philosophical analysis may not fit people's explanatory judgments.

Conversely, if the norm fits some salient aspects of people's explanatory reasoning, if it captures real judgments of explanatory power on a battery of test cases, or if it shares salient aspects of the meaning of the expression 'explanatory power' as used in everyday English, then the norm will enjoy some empirical support, and the theory it underlies will gain in justification.

One source of evidence comes from the history of science. Henderson (2014) examined how a norm of explanatory power can account for the widely-shared belief that the Copernican theory is better than the Ptolemaic theory in explaining the retrograde motion of the planets. Henderson convincingly argued that the Copernican theory provides a more powerful explanation of the phenomenon because its explanation is less dependent on auxiliary hypotheses than the Ptolemaic theory. According to Henderson, this type of consideration is plausibly reflected in the Bayesian likelihoods of the two theories. For some independently motivated prior probabilities for the two theories, the Bayesian likelihoods favour the more powerful of the two theories, which is what a norm of explanatory power like the one put forward by Schupbach & Sprenger (2011) would advise.

Another source of evidence comes from experimental studies about how people's explanatory judgments are sensitive to manipulations of probabilistic information. For example Schupbach & Sprenger's (2011) favoured analysis received independent empirical support by Schupbach (2011) (see also Douven & Schupbach 2015). In his study, Schupbach asked experimental participants to make judgments of both explanatory power and conditional probabilities in an actual urn problem. He then compared five candidate probabilistic analyses of explanatory power in terms of their fit to participant's judgments, and found that his and Sprenger's measure provided the best fit (but see Glymour 2015, Sec. 4, for a criticism).

Linguistic evidence is a third source of empirical evidence. One strategy is to examine how the expression 'explanatory power' is used in a variety of linguistic corpora over time. Such an

examination will provide evidence about the (psycho)linguistic adequacy of a given analysis. It will also help us identify how different usage of the terms depend on particular features of the explanatory context, including the kind of causal information available, shared background knowledge, and the epistemic and practical interests of explainers and their audience (e.g., Van Fraassen 1980, Ch. 5).

A fourth type of evidence comes from studies of the complex, distributed, social dynamics shaping the activity of explaining and the usage of the term 'explanation' in both scientific and non-scientific contexts. Since explanation is an eminently social phenomenon, the adequacy of a norm of explanation should be assessed also in the light of information about how the production and evaluation of explanations are affected by factors like trust, authority, and one's socio-economic status (e.g., Longino 2002). Special attention should be paid to how scientific institutions like peer review, journal rankings and funding agencies contribute to determine what is taken to be explanatory (e.g., Knorr-Cetina 1981; Latour & Woolgar 1986).

From the finding that a norm of explanation is empirically inadequate in some respects, many philosophers will shrug, concluding that people's judgments are simply mistaken and that the psychological, linguistic, or sociological evidence should not impact the epistemic status of the norm, or of any model where the norm is embedded. People may not understand what the expression 'explanatory power' really means and how it should be used. Folks ought to follow the norm, but they fail to do so.

Regardless of many philosophers' reaction to the psychological evidence, it is implausible that no empirical consideration should be brought to bear on the epistemic status of a norm of explanation. Even if we believed that an adequate theory of explanation should not be informed by evidence about the pieces of explanatory information humans tend to produce or accept (Hempel 1965; Salmon 1989; Craver 2014), it does not follow that we should ignore *any* piece of empirical evidence. For example, while Craver (2014) claims that "individual explanatory judgments are not data to be honored by a normative theory that seeks to specify when such judgments go right and when they go wrong," he argues for particular norms of explanation on the basis of a putative lexical ambiguity in the term 'explanation.' But if differences in the use of the term 'explanation' are relevant to an account of the norms of explanation, then judgments about the theoretical adequacy of the norm should be held accountable at least to results from psycholinguistic and sociological studies.

4.4. Justifying a norm of explanatory power: pragmatic considerations

As a source of control that could provide independent justificatory power to the method of reflective equilibrium, we should weigh in a third set of judgments concerning the pragmatic adequacy of the norms. We should ask to what extent following a norm in its relevant context of application helps us satisfy goals we care about. Hempel appears to subscribe to this idea, when he writes: "Man wants not only to survive in the world, but also to improve his strategic position in it. This makes it important for him to find reliable ways of foreseeing changes in his environment and, if possible, controlling them to his advantage" (Hempel 1965, 333).

The goals that may be considered in forming particular judgments about the pragmatic adequacy of a norm are diverse, at different levels of generality, and both individual and social. These goals are diverse because diverse are the cognitive functions of explanation, and diverse are the practical interests, values, technical possibilities, and contexts where some individual or community make use of a certain model of explanation. These goals are at different levels of

generality: prediction and control are relatively general pragmatic goals, while the goal of discovering functional-mechanical features of a specific kind of system is less general. The goals may be shared by a community, or they may be entertained only by some individuals.

Suppose that the aim is to make reliable predictions about phenomena of interest. The question we should ask is: How does a specific norm of explanatory power help in the attainment of this aim? By working back and forth between questions about the pragmatic adequacy of particular norms and about the contextual features that contribute to the success of the norms, we relate theoretical judgments and empirical results, thereby systematizing our thinking about explanation.

Consider again Schupbach & Sprenger's (2011) analysis of the norm of explanatory power. This norm is pragmatically adequate when its application helps us reliably, precisely, or simply to identify what particular explanatory hypotheses are good predictors of phenomena of interest. The norm will also gain in pragmatic adequacy, if its application has many significant, useful cognitive consequences. For instance, explanations that comply with a norm of explanatory power as understood by Schupbach & Sprenger (2011) can favour people's learning. What one learns about the environment is constrained by the explanation's ability to decrease the degree to which one finds the explanandum surprising (Lombrozo 2012, 265-8; Wittwer & Renkl 2008).

5. Conclusion

This paper has argued that there is a plurality of concepts of *explanation*. This plurality is reflected in the multiplicity of philosophical models of explanation, whose normative status can be assessed through a method of reflective equilibrium. If what I argued in this paper is correct, then a way forward for the philosophy and psychology of explanation is to abandon the idea that there is one true model of explanation, and to pursue an experimental approach to the philosophy of explanation.

Acknowledgements

I am grateful to Liz Irvine, Raoul Gervais, Jan Sprenger, and Naftali Weinberger for helpful comments on previous versions of this paper and for engaging discussion. A special thank you goes to three anonymous referees of this journal for their time and their exceptionally constructive comments. The work on this project was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program New Frameworks of Rationality ([SPP 1516]).

References

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity*. Princeton, NJ: Princeton University Press.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive science*, *24*(4), 573-604.

Carnap R. (1950). *Logical Foundation of Probability*. London: Routledge and Keegan Paul.

Colombo, Bucher, Postma-Nilsenovà, & Sprenger (2015). Explanatory Value and Probabilistic Reasoning: An Empirical Study. Available at: http://philsci-archive.pitt.edu/11392/ (Last accessed August 26th 2015).

Craver, C.F. (2014). The ontic conception of scientific explanation. In A.Hütteman & M.Kaiser (Eds.), *Explanation in the special sciences: Explanation in the biological and historical sciences* (pp. 27–52). Dordrecht: Springer.

Craver, C.F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Clarendon Press.
Crupi, V., and Tentori, K. (2012). A second look at the logic of explanatory power (with two novel representation theorems). *Philosophy of Science*, 79, 365-385.
Douven, I. & Schupbach, J. (2015). The role of explanatory considerations in updating. *Cognition*, 142, 299-311.
Fonseca, B. A., & Chi, M. T. (2011). Instruction based on self-explanation. In Mayer, R. & Alexander, P. (Eds.), *Handbook of research on learning and instruction* (pp. 296-321). Routledge.
Glymour, C. (2015). Probability and the Explanatory Virtues. *The British Journal for the Philosophy of Science*, 66(3), 591-604.
Good, I.J. (1960). Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments, *Journal of the Royal Statistical Society, Series B (Methodological)*, 22, 319-331.
Goodman, N. (1983). *Fact, fiction, and forecast. 4th ed.* Cambridge, MA: Harvard University Press.
Hempel, C. (1965). Aspects of Scientific Explanation, in Carl Hempel's *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: Free Press, pp. 376-386.
Hempel, C., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15, 135-175.
Henderson, L. (2014). Bayesianism and inference to the best explanation. *The British Journal for the Philosophy of Science*, *65*(4), 687-715.
Kaplan, D.M., & Craver, C.F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science* 78 (4): 601–27.
Keil, F.C. (2006). Explanation and Understanding. *Annual Review of Psychology.* 57, 227-254.
Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher and W. Salmon (Eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science,* Vol. 13. (pp. 410–505). Minneapolis: University of Minnesota Press.
Knorr-Cetina, K. (1981). *The Manufacture of Knowledge*, Oxford: Pergamon Press.
Latour, B. & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*, 2d ed., Princeton: Princeton University Press.
Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, 8, 101–106.
Legare, C. H. & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology, 126,* 198-212.
Lombrozo, T. (2012). Explanation and abductive inference. In K.J. Holyoak and R.G. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning* (pp. 260-276), Oxford, UK: Oxford University Press.
Lombrozo, T. (2011). The instrumental value of explanations. *Philosophy Compass*, 6, 539–551.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology* 55, 232-257.
Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, 99, 167–204.
Longino, H.E. (2002). *The Fate of Knowledge*, Princeton: Princeton University Press.
Machamer P., Darden L., & Craver C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1-25.

Mancosu, P. (2001). Mathematical explanation: Problems and prospects. *Topoi*, *20*(1), 97-117.
McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning, *British Journal for the Philosophy of Science*, 54, 553-567.
Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, *18*(3), 399-429.
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.
Roscoe, R.D., & Chi, M.T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*(4), 534-574.
Salmon, W.C. (1989). Four decades of scientific explanation. In P. Kitcher and W. Salmon (Eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, Vol. 13. (pp. 3–219). Minneapolis: University of Minnesota Press.
Salmon W. (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NY: Princeton University Press.
Schupbach, J. (2011). Comparing Probabilistic Measures of Explanatory Power, *Philosophy of Science*, 78, 813-829.
Schupbach, J. & Sprenger, J. (2011). The Logic of Explanatory Power, *Philosophy of Science*,78, 105-127.
Shepherd, J., & Justus, J. (2014). X-Phi and Carnapian Explication. *Erkenntnis*, *80*(2), 381-402.
Steiner, M. (1978). Mathematical Explanation. *Philosophical Studies*, 34, 135-151.
Stich, S. (1990). *The Fragmentation of Reason*, Cambridge, MA: MIT Press.
Strevens, M. (2008). *Depth: An account of scientific explanation*. Cambridge, MA: Harvard University Press.
Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.
Walker, C.M., Lombrozo, T., Legare, C., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, 133, 343-357.
Waskan, J., Harmon, I., Horne, Z., Spino, J., & Clevenger, J. (2014). Explanatory anti-psychologism overturned by lay and scientific case classifications. *Synthese*, 191(5), 1013-1035.
Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalisation: Evidence from category learning. *Cognitive Science*, 34, 776-806.
Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*, 49-64.
Woodward, J. (2014). "Scientific Explanation," *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/>.
Woodward, J. (2012). Causation, Interactions Between Philosophical Theories and Psychological Research. *Philosophy of Science* 79 :961-972
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.
Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, 148, 201-219.