Explanatory Value and Probabilistic Reasoning: An Empirical Study

Matteo Colombo* Marie Postma† Jan Sprenger‡

Friday 2nd May, 2014

Abstract

The relation between probabilistic and explanatory reasoning is a classical topic in philosophy of science. Most philosophical analyses are concerned with the compatibility of Inference to the Best Explanation (IBE) with probabilistic, Bayesian inference, and the impact of explanatory considerations on the assignment of subjective probabilities (Van Fraassen 1989; Okasha 2000; Lipton 2004). This paper reverses the question and asks how causal and explanatory considerations are affected by probabilistic information. We investigate (i) how probabilistic information determines the explanatory value of a hypothesis, and (ii) in which sense folk explanatory practice can be said to be rational. Our study identifies three main factors in reasoning about a (potentially) explanatory hypothesis: cognitive salience, rational acceptability and logical entailment. This corresponds well to the variety of philosophical accounts of explanation. Moreover, we

*Contact information: Tilburg Center for Logic and Philosophy of Science (TiLPS), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: m.colombo@uvt.nl

†Contact information: Tilburg Center for Communication and Cognition (TiCC), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: m.colombo@uvt.nl

‡Contact information: Tilburg Center for Logic and Philosophy of Science (TiLPS), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: j.sprenger@uvt.nl. Webpage: www.laeuferpaar.de

show that these factors are highly sensitive to manipulations of probabilistic information. This finding suggests that probabilistic reasoning is a crucial part of explanatory inferences, and it motivates new avenues of research in the debate about Inference to the Best Explanation and probabilistic measures of explanatory power.

1 Introduction

The relationship between probabilistic and explanatory reasoning is a central topic of the epistemology of science and of the psychology of explanation. It is first touched upon in the inductive-statistical model of explanation by Carl G. Hempel (1965) and it has been developed further by writers such as Wesley Salmon (1971/84) and Peter Railton (1979). Their models attempt to capture explanatory relationships in terms of statistical dependencies between explanans and explanandum. In the last decades, however, Inference of the Best Explanation (IBE) and its compatibility with probabilistic reasoning have become the focus of the philosophical debate (Harman 1965; Van Fraassen 1989; Lipton 2004). The crucial question is whether inferring to the hypothesis with the highest explanatory power, or the optimal combination of explanatory virtues, can be defended as a rational form of inference, if "rational" is explicated in the probabilistic, Bayesian way (e.g., Oaksford and Chater 2007; Hartmann and Sprenger 2010). More precisely, it is debated whether Inference to the Best Explanation (IBE) contradicts the Bayesian rule of conditionalization, or whether IBE can be interpreted as a distinctive rule of inference that is not only coherent with, but also supplements, Bayesian inference.

Van Fraassen (1989, ch. 7) contends that IBE either amounts to Bayesian conditionalization, thereby being redundant as a rule of inference, or deviates from it, thereby being probabilistically incoherent. If IBE deviates from Bayesian inference, by assigning a confirmatory role to explanatory considerations, then IBE is probabilistic incoherent, on the grounds of dynamic Dutch book arguments (Lewis 1999). In other words, the betting odds implied by the beliefs of an agent who assigns a probabilistic bonus to good explanations would allow for a system of bets where he or she can only lose. Therefore, if IBE is probabilistically incoherent, it is irrational to follow it. If, on the other hand, IBE always agrees with Bayesian inference, it is unclear whether this rule of inference has independent epistemic significance, or whether it is just redundant with respect to Conditionalization. The upshot is that IBE cannot be both a distinctive rule of inference and rational.

Recently, philosophers of science have tried to sketch a more sophisticated picture of the interplay between probabilistic and explanatory reasoning. In particular, the idea that explanatory virtues boost the posterior probability of scientific hypotheses beyond the value they receive from Bayesian conditionalization is rejected as too simplistic. For example, Lipton (2001, 2004) argues that IBE may give a good *descriptive* account of our inferential practices, even if it fails to accomplish the normative ideal of probabilistic coherence. Many popular reasoning schemes are at odds with the axioms of probability, but they can be a useful heuristics for probabilistic inferences in real life (e.g., Schupbach 2011a). Perhaps IBE can play such an intermediate role between people's actual reasoning and the way they would reason if they were perfectly rational. In that sense, IBE could possess independent significance as a heuristic for probabilistic inference.

In an even more reconciliatory mood, Lipton suggests that explanatory value may help to determine the prior probability and/or the likelihood of the explanatory hypothesis on the phenomenon of interest (see also Okasha 2000; McGrew 2003; Weisberg 2009). In this view, IBE should not be construed just as an inference to the explanation with the highest posterior probability—then, there would be no need for a theory of *explanatory* inference—, but explanatory judgments should provide constraints on the ingredients that enter the Bayesian cooking recipe $p(H|E) = p(H)\, p(E|H)\ /\ p(E)$ (Bayes' Theorem). Instead of being a *heuristic* for probabilistic inference, explanatory value *guides* such inferences. For example, explanatory virtues such as general scope, simplicity or unification will positively affect the prior of a hypothesis H, while others, such as making the phenomenon E cognitively intelligible, will raise the likelihood of H on E, $p(E|H)$ (Lipton 2004, 113). IBE would then not only be compatible with, but actively supplement Bayesian Conditionalization. Henderson (2014) even proposes that IBE *emerges* from Bayesianism rather than being a constraint on Bayesian inference.

In Lipton's account, a special role is assigned to "lovely" explanations:

these are explanations "that would, if correct, be the most explanatory or provide the most understanding" (Lipton 2004, 59–60). Explanatory value is thus tightly connected to inspiring understanding in an individual reasoner (cf. De Regt and Dieks 2005). Accounts of explanation that stress the cognitive value of good explanations also tend to stress the link between understanding, causality and mechanisms—a topic that has recently received much attention with respect to explanation in the special sciences, including biology, psychology, economics, and so on (Machamer, Craver and Darden 2000; Woodward 2003; Strevens 2008). Notably, even the literature on probabilistic causation stresses that successful explanations presuppose a causal link between explanans and explanandum (Halpern and Pearl 2005).

Two questions which these debates have not fully elucidated are (i) how probabilistic considerations are in themselves explanatorily relevant and (ii) whether a powerful explanation should make the explanandum more expected. This latter idea actually goes back to Charles Saunders Peirce, who identified the explanatory power of a hypothesis with its ability to render an otherwise "surprising fact" as a "matter of course" (Peirce 1931-1935, Section 5.189). If this is correct, then a powerful explanation should at least raise the subjective expectedness of the explanandum, which suggests an analysis of explanatory power in probabilistic terms, such as conducted recently by McGrew (2003), Schupbach and Sprenger (2011) and Crupi and Tentori (2012). See Popper (1934/2002) and Good (1960) for early precursors.

In the light of the philosophical debates just outlined, the present paper contributes to the literature by investigating experimentally whether and under which circumstances judgments of explanatory power are associated with probabilistic characteristics of the potential explanation. This investigation gives us a nuanced and empirically informed assessment of the hypotheses that IBE is coherent with actual probabilistic reasoning, and that explanatory power is bound up with understanding.

In particular, we constructed various vignettes where experimental participants are given information about the priors and likelihoods of a potential explanatory hypothesis and are asked to make judgments on its explanatory power, posterior probability and acceptability, as well as on its logical, causal and cognitive relation to the explanandum. Causal information was kept sparse in order to be able to isolate the impact of probabilistic factors on explanatory judgment. By eliciting these judgments, we sought to elucidate how these concepts cluster together, identifying a small set of factors that accounted for most of the variation in the participants' judgments. Overall, our results show that explanatory judgments are largely sensitive to subtle changes in the probabilistic setting, underwriting the efficacy of our experimental manipulations.

The rest of the paper is structured as follows. Section 2 briefly surveys the relevant empirical literature. Section 3 presents our experiment and Section 4 our results. Finally, Section 5 puts the results into a broader philosophical perspective and discusses, inter alia, the implications for quantitative approaches to explanatory power and broader consequences for theories of explanatory reasoning.

2 Empirical Research on Explanatory Reasoning

Empirical research in cognitive and developmental psychology has started to uncover some of the properties and mechanisms of explanatory reasoning. In particular, explanation has been shown to be closely related to confirmation; explanatory considerations can contribute to making some hypotheses more credible. Koehler (1991), for example, reviewed much of the work on 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 has also been demonstrated to influence how probabilities are assigned to one proposition in the light of another. 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 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 (on how coherence considerations can impact explanation see, e.g., Thagard (1989)).

Relatively little empirical research has focused on the link between explanatory value and understanding. Available results points to the cognitive dangers of putting too much weight on the sense of understanding that certain explanations can induce. For example, there is evidence that people tend to overestimate the depth of their own explanatory understanding. Rozenblit and Keil (2002) called this phenomenon "illusion of explanatory depth", and they demonstrated that this illusion is significantly stronger for explanatory knowledge relative to other knowledge domains (see also Keil 2006). Naturalistically minded philosophers of explanation such as J.D. Trout (2002, 2007) argue that a sense of understanding is indeed frequently deceptive, induced by overconfidence and hindsight bias rather than by explanatory power, and in any case not a good indicator for a valuable explanation.

These empirical results are broadly in line with Lipton's (2004) account of the relationship between IBE and Bayesian inference: explanatory considerations guide the assignments of subjective probabilities to propositions, and inferences to lovely explanations are, in some sense, bound up with understanding. However, like its counterpart in philosophy of science, the relevant psychological literature has mainly been concerned with the question of how probability judgments are informed by explanatory considerations, neglecting the question of how probabilistic and logical information can themselves influence explanatory judgements (see also Keil and Wilson 2000; Keil 2006; Lombrozo 2011, 2012). An answer to this latter question will be relevant to better elucidate the relationships between probabilistic and explanatory reasoning, explanatory value and understanding, as well as to assess whether or not IBE is probabilistically coherent. Furthermore, even if explanation, confirmation and acceptance are closely related in IBE, it is far from obvious whether they correspond to distinct concepts in our psychology that tap onto different reasoning capacities. What is more, it is far from obvious whether these three concepts in our psychology—if they exist—agree with their philosophical explications (Crupi et al. 2007; Schupbach 2011b; Crupi 2012).

3 Experiment and Methods

Our experiment consisted of an online questionnaire, conducted via the LimeSurvey environment. The participants for our study were undergraduate students of Tilburg University from the School of Economics and Management and the School of Social and Behavioral Sciences. They were recruited via emails from a teacher of one of their classes. Incentives were provided in terms of points for the final exam and a prize lottery.

The respondents of the survey were 744 students, of which 671 completed the questionnaire (383 male, M_{age} = 21.5 (SD = 2.3)). They were randomly assigned to one of the 12 versions of an experimental vignette. Each participant received exactly one vignette.

Design and Material

Participants were presented with an experimental vignette where two possible events were related to two possible explanations for that event:

> **Vignette 1:** *There are two urns on the table. Urn A contains 67% white and 33% black balls, Urn B contains only white balls. One of these urns is selected. You don't know which urn is selected, but you know that the chance that Urn A is selected is 25%, and that the chance that Urn B is selected is 75%. From the selected urn a white ball is taken at random.*
> *Please now consider the hypothesis that Urn A has been chosen.*

The participants were then asked to assess the following seven items (the construct names in italics were not provided to the participants) on a Likert scale ranking from 1 ("do not agree at all") to 7 ("fully agree"):

- (*Logical Implication*) The hypothesis logically implies that a white ball has been taken out.

- (*Causality*) The hypothesis specifies the cause that a white ball has been taken out.

- (*Confirmation*) The hypothesis is confirmed by a white ball has been taken out.

- (*Posterior Probability*) The hypothesis is probable given that a white ball has been taken out.

- (*Explanatory Power*) The hypothesis explains that a white ball has been taken out.

- (*Understanding*) The hypothesis provides understanding why a white ball has been taken out.

- (*Truth*) The hypothesis is true.

The choice of these seven items was motivated by the crucial role that concepts such as logical implication, causality and confirmation play in reasoning about candidate explanations, according to different philosophical accounts of explanatory value.

After filling in this questionnaire, the participants could explain the way they made their judgments, and we collected some demographic data.

This vignette was, in a between-subjects design, varied in three dimensions, corresponding to three main independent variables:

1. **Likelihood:** the choice of the evidence (white or black ball) which leads to a high or a low likelihood of the hypothesis on the evidence;

2. **Target Hypothesis**: the choice of the hypothesis (Urn A or Urn B) which is either deterministic or probabilistic;

3. **Prior Probability:** the prior probability of the hypothesis under consideration (.25, .5, or .75).

All possible $2 \times 2 \times 3 = 12$ combinations of the values of these variables were realized in the experiment. Since the above vignette is quite abstract, we also set up two other vignettes that are closer to cases of ordinary reasoning, and repeated the experiment for these vignettes. Example vignettes, that copy the probabilistic structure of Vignette 1 above, are given below.

> **Vignette 2:** *Again and again, Ruud has knee problems when playing football. The doctors give him two options: knee surgery or a conservative treatment. If Ruud chooses to go into surgery, he cannot play football for half a year; if he chooses the conservative treatment, there is a 33% chance that he can play again after one month; otherwise (with a chance of 67%) he has to rest longer. You don't know which option Ruud chooses, but you believe that the chance that he chooses surgery is 75%—and that the chance that he chooses the conservative treatment is 25%. A month later a joint friend tells you that Ruud is still unable to play football.*
> *Please now consider the hypothesis that Ruud has chosen for the conservative treatment.*

> **Vignette 3:** *Louise arrives by train in Twin City. Twin city has two districts: West Bank and East Bank. In West Bank, there is only one taxi company, namely Green Taxi Ltd., and all their cabs are green. Green Taxi Ltd. also owns 67% of all cabs in East Bank. The other cabs in East Bank are owned by The Red Taxi Inc., all their cabs are red. Louise does not know which part of the city the train is entering, but judging from her knowledge of Twin City she assumes that there is a 75% chance that she is in West Bank (and a chance of 25% that she is in East Bank). At some point, Louise sees a green cab from the train.*
> *Please now consider the hypothesis that Louise is in East Bank.*

Procedure

Participants completed the questionnaire on a university PC or their own computer in the digital environment of LimeSurvey installed on a local server. The use of LimeSurvey guaranteed that the data could be protected and provided with a time stamp and information about the IP address of the respondent. The experiment was self-paced and took approximately 10 minutes to complete. In total, the experiment thus contained 36 cells, corresponding to twelve different combinations of the values of the independent variables times three different scenarios.

4 Results

Prior to the analysis of the effects of vignette manipulation, we explored the interdependencies of the seven items in the response questionnaire. To recall, the participants were asked to judge several aspects of the hypothesis with respect to the evidence: logical implication, causal relevance, explanatory power, increase in understanding, confirmation, posterior probability and truth. By analyzing the interdependencies with the help of the Pearson zero-order correlation coefficient, we determined whether the participants clearly separate these seven concepts, or whether some of them can be identified with each other.

The correlations are presented in Table 1. The analysis revealed that all of the variables correlated at least with .3 with several other variables, but at most .63. This validates the hypothesis that the participants do not conflate cognate concepts (e.g., causality, explanatory power) with each other, which would be reflected in correlation coefficients of greater than .7. At the same time, the response variables were sufficiently related to each other to motivate a *factor analysis*: that is, a decomposition of the seven response variables into 2-4 constructs that explain together most of the variation in the data.

	1	2	3	4	5	6	7
1. *Logical Implication*	-	.38	.22	.32	.46	.30	.12
2. *Causality*		-	.45	.39	.56	.63	.37
3. *Confirmation*			-	.56	.35	.47	.63
4. *Post Probability*				-	.37	.51	.46
5. *Explan. Power*					-	.60	.28
6. *Understanding*						-	.36
7. *Truth*							-

Table 1: Zero-order correlations for 7 items (N = 671), all correlations with $p < .01$.

Factor Analysis

The factorability of the 7 items was examined with a Principle Component Analysis (PCA). The Kaiser-Meyer-Olkin measure of sampling adequacy was .82 and the Bartlett's test of sphericity was significant ($\chi^2(21)$ = 1790.77, $p < .0001$). The initial eigenvalues showed 51% of variance explained by the first factor, 16% explained by the second factor, and 10% explained by the third factor. A visual inspection of the scree plot revealed a 'leveling off' of eigenvalues after the three factors, therefore, a three factor solution using the oblique rotation was conducted, with the three factors explaining 77% of the variance. All items had primary loadings over .7, viz. Table 2, which presents the factor loading matrix (loadings under .30 suppressed). In the remainder, we will restrict our analysis to these three factors.

The names for these factors are derived from the clustering that Table 2 indicates. Factor 1, **Cognitive Salience**, clusters explanatory power together with cognitive values that are often seen as related, such as causal coherence and enhancement of understanding (De Regt and Dieks 2005; Strevens 2008). Factor 2, **Rational Acceptability**, captures those cognitive values that hang together with the acceptability of a hypothesis: its prob-

	1	2	3	Communality
Logical Implication			.94	.94
Causality	.86			.74
Confirmation		-.84		.77
Post Probability		-.72		.67
Explan. Power	.81			.73
Understanding	.87			.78
Truth		-.88		.75

Table 2: Factor loadings and communalities based on a principle component analysis with oblimin rotation for 7 items (N = 671).

ability, its confirmation by the evidence, and finally, its truth. The strong correlations between these values are not surprising: confirmation raises posterior probability, which is in turn an indicator of the truth of a theory (e.g., Howson and Urbach 2006). Finally, Factor 3 captures the logical relation between hypothesis and evidence. Since no other response variable is loaded on this factor, it figures as **Entailment**, showing the link to the response variable Logical Implication.[1]

Tests of Experimental Manipulation

We conducted three analyses of variance (ANOVAs) to test the effects of the independent variables, Target Hypothesis, Likelihood and Prior Probability, on Cognitive Salience, Rational Acceptability and Entailment, respectively.[2]

[1] The internal consistency for two of the three scales (the third scale only consisted of one item) was examined using Cronbach's alpha, resulting in alpha .82 for Factor 1 and .79 for Factor 2. Composite scores were calculated for each of the three factors using the mean of the items with primary loadings on each factor. The descriptive values for the newly constructed scales were M = 3.65, SD = 1.91 for Explanatory Value, M = 3.63, SD = 1.87 for Rational Acceptability, and M = 3.75, SD = 2.40 for Logical Implication.

[2] A prior analysis of the effect of the vignette on the three dependent variables revealed that Explanatory Value (but not Rational Acceptability and Logical Implication) was also affected by the vignette manipulation. For clarity of exposition, the statistics is

	Target Hypothesis	
Likelihood	*Probabilistic*	*Deterministic*
High	3.06 (.12)	4.67 (.12)
Low	4.68 (.12)	1.95 (.13)

Table 3: Estimated Marginal Means and SE of Explanatory Value by Target Hypothesis and Likelihood (N = 671).

First, we tested the effects of the experimental manipulation on Cognitive Salience. There was a main effect of Target Hypothesis, $F(1, 659) = 21.56$, $p < .001$,$\eta^2{}_p = .03$, and Likelihood, $F(1, 659) = 21.09$, $p < .001$,$\eta^2{}_p = .03$ on Cognitive Salience, with no main effect of Prior Probability, $F(2, 659) < 1$, $p = .82$. The significant main effects can only be interpreted in the light of the interaction effect between Target Hypothesis and Likelihood, $F(1, 659) = 323.65$, $p < .001$,$\eta^2{}_p = .33$—see Table 3 for the descriptives.

The cognitive salience of the hypothesis was high in two cases: (1) when the actually observed event was impossible under the alternative explanation (in other words, when the hypothesis provided the only available explanation), and (2) when the actually observed event was logically implied by the candidate explanation. This result confirms the philosophical thesis that both the epistemic status of the explanans and the expectedness of the explanandum under the explanans play an important role in explanatory reasoning (e.g., Okasha 2000; Lipton 2004; Schupbach and Sprenger 2011). There were no other significant interaction effects ($F(2, 659) < 1$, n.s.).

	Target Hypothesis	
Likelihood	*Probabilistic*	*Deterministic*
High	3.07 (.10)	3.56 (.10)
Low	5.61 (.10)	1.98 (.11)

not included here because the size and direction of the significant main effects and the interaction effect remained the same when vignette was included as a factor.

Table 4: Estimated Marginal Means and SE of Rational Acceptability by Target Hypothesis and Likelihood (N = 671).

Second, with respect to Rational Acceptability, there was again a main effect of Target Hypothesis, $F(1, 659) = 234.77$, $p < .001$,$\eta^2{}_p = .26$, and Likelihood, $F(1, 659) = 21.45$, $p < .001$,$\eta^2{}_p = .03$, with no main effect of Prior Probability, $F(2, 659) = 1.68$, $p = .19$. There was an interaction effect between Target Hypothesis and Likelihood, $F(1, 659) = 401.09$, $p < .001$,$\eta^2{}_p = .38$, see Table 4 for the descriptives. The interaction effect is easy to explain: The candidate explanation is strongly accepted whenever the alternative hypothesis is incompatible with the observed evidence. Conversely, a candidate explanation scores low on this factor when it is itself incompatible with the observed evidence. The score is rather middling when the evidence is inconclusive and posteriors are non-extreme, in agreement with philosophical analyses of rational hypothesis acceptance. There were no other significant interaction effects ($F(2, 659) <= 1$, n.s.).

	Target Hypothesis	
Likelihood	*Probabilistic*	*Deterministic*
High	3.03 (.15)	5.92 (.15)
Low	3.72 (.15)	2.19 (.16)

Table 5: Estimated Marginal Means and SE of Entailment by Target Hypothesis and Likelihood(N = 671).

Third and last, we assessed the impact of the experimental manipulations on Entailment. There was a main effect of Target Hypothesis, $F(1, 659) = 20.11$, $p < .001$,$\eta^2{}_p = .03$, and Likelihood, $F(1, 659) = 97.93$, $p < .001$,$\eta^2{}_p = .13$, with no main effect of Prior Probability, $F(2, 659) = 1.23$, $p = .29$. There was an interaction effect between Target Hypothesis and

Likelihood, $F(1, 659) = 207.11$, $p < .001$,$\eta^2{}_p = .24$, see Table 5 for the descriptives. Briefly, the response values of Logical Implication respect the ranking given by the (objective) likelihoods $p(E|H)$, demonstrating the consistency of the participants with the probabilistic information in the vignettes. In particular, a logical implication between explanans and explanandum receives the highest value and logical incompatibility receives the lowest value. There were no other significant interaction effects.

5 Discussion

Overall, the results of our study shows that explanatory value is, as Lombrozo (2012, 270) reports, a complex psychological phenomenon related to several other cognitive processes such as deductive and inductive reasoning, causal reasoning, and reasoning about truth. The correlations between participants' judgments on individual items suggest that it may be complicated to tease apart the specific roles of these processes in determining explanatory judgments and making explanatory inferences.

Such a complexity might explain the disagreement about existing philosophical accounts of scientific explanation. Each of the three most prominent philosophical accounts of explanation attempts to explicate the concept of scientific explanation by focusing on a particular aspect of explanatory value. Hempel (1965) ties explanatory value to derivability, on the grounds that for a hypothesis to successfully explain an explanandum, the explanandum must be a logical consequence of the explanans. Schupbach and Sprenger (2011) and Crupi and Tentori (2012) generalize Hempel's account to making the explanandum more expected (instead of deriving it), and they use probabilistic relevance relations for spelling out explanatory value. According to causal-mechanical accounts such as Salmon (1971/84), Woodward (2003), and Strevens (2008), explanatory value lies in the identification of causal mechanisms. For unificationist models such as Friedman (1974) and Kitcher (1981), explanatory value is a matter of providing a unified account of a range of different phenomena that could yield understanding. If explanatory value is indeed a complex psychological phenomenon, and insofar as these accounts aim to capture *the* value of explanation, it is not surprising that disagreements about the correct account of explanation will persist, and that explanatory value cannot be mapped neatly on one of these accounts.

In the light of this diversity, a rapprochement between competing theories of scientific explanation seems desirable. The goal of a philosophical account of explanation should be to capture the complexity of explanatory value. Precisely characterizing the heterogeneity of explanatory judgment, isolating the factors and processes that contribute to the assessment of a potentially explanatory hypothesis, is thus an important direction for future research in psychology as well as in philosophy and epistemology of science.

Three more specific conclusions can be drawn from the results of our study.

First, our factor analysis separated the items "causality", "understanding" and "explanatory power" ($\rightarrow$ Cognitive Salience) from the items "confirmation", "posterior probability", "truth" ($\rightarrow$ Rational Acceptability) and "logical implication" ($\rightarrow$ Entailment). This corresponds to three types of factors that figure in philosophical accounts of explanatory reasoning: cognitive, logical and probabilistic relations. On the one hand, this finding substantiates previous results that indicated the existence of a tight connection between explanatory power, causality and a sense of understanding (Lipton 2004; Keil 2006; Lombrozo 2007; Trout 2007). On the other hand, the finding indicates that folk reasoning about potential explanations can neatly distinguish between the concepts of Rational Acceptability and Cognitive Salience of a hypothesis. This suggests that the explanatory value attributed to a hypothesis need not indicate its acceptability, spelling trouble for IBE.

Second, participants' judgments on the three main factors we identified—Cognitive Salience, Rational Acceptability and Logical Implication—were strongly affected by changes in explicit probabilistic information. Hence, explanatory reasoning is not only bound up with causal mechanisms: it is heavily affected by logical and probabilistic relations between hypothesis and evidence. In other words, our results suggests that probabilistic reasoning might be a crucial ingredient of explanatory inferences—at least in situations, such as our vignettes, where there is a limited set of competing explanations. Future research may investigate whether this result stays robust if we extend the type of vignettes to scenarios where more explicit causal mechanisms are given, and where a potentially unlimited number of hypotheses compete for being the best explanation. In the light of the strong link between causal and explanatory judgments, it also seems promising to explore a dual model of explanatory reasoning, e.g., a model where a judgment on explanatory power combines the plausibility of a causal mechanism with probabilistic relevance for the phenomenon of interest.

Third, the compatibilist research program about IBE and probabilistic inference is not supported by our findings. The concepts of posterior probability, degree of confirmation and truth, that all relate to the rational acceptability of the explanans, cluster together, but they are quite remote from explanatory power. Attributions of explanatory value to a hypothesis often diverge from the acceptability of the hypothesis. This poses an important challenge for those who defend explications of IBE according to which explanatory value is a reliable guide to truth. Broader applications in philosophy of science abound. For example, arguments for scientific realism often postulate a link between explanation and truth, where the truth of our best scientific theories is supported by the observation that it is the best explanation for the success of science (Boyd 1983; Psillos 1999).

All this does not rule out that Inference to the Best Explanation may, under favorable circumstances, be a good heuristic for Bayesian inference, and that participants use it in some of these cases. In fact, our study is silent on this latter issue. It just demonstrates that sound probabilistic inference and explanatory power are quite different concepts, and that subjects recognize this difference.

To sum up, our results demonstrate that probabilistic reasoning is an important part of explanatory inferences. Overall, we see that for a proper understanding of explanatory reasoning, it is essential to investigate the precise relationship between explanatory power and other cognitive values, as well as the subtle relationship between explanatory value and probabilistic inference. These topics have been relatively neglected in the psychological and philosophical literature. With our study, we hope to make a small step forward toward a genuinely pluralistic and naturalized philosophy of explanation.

References

Boyd, R. N. (1983). On the Current Status of the Issue of Scientific Realism. *Erkenntnis* 19: 45–90.

Crupi, V. (2012). An argument for not equating confirmation and explanatory power. *The Reasoner* 6: 39–40.

Crupi, V. & Tentori, K. (2012). A second look at the logic of explanatory power (with two novel representation theorems). *Philosophy of Science* 79, 365–385.

Crupi, V., Tentori, K., and González, M. (2007). On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science* 74: 229–252.

De Regt, H., and Dieks, D. (2005). A Contextual Approach to Scientific Understanding. *Synthese* 144: 137–170.

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

Good, I. J. (1960). Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments. *Journal of the Royal Statistical Society B* 22: 319–331.

Halpern, J., and Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part II: Explanations. *British Journal of Philosophy of Science* 56: 889–911.

Hartmann, S., and Sprenger, J. (2010). Bayesian Epistemology. In D. Pritchard (ed.): *Routledge Companion to Epistemology*, 609–620. London: Routledge.

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

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Henderson, L. (2014). Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science*, Online First doi: 10.1093/bjps/axt020.

Howson, C. and Urbach, P. (2006): *Scientific Reasoning: The Bayesian Approach*, 3rd edition. La Salle: Open Court.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology* 57: 227-254.

Keil, F. C., & Wilson, R. A. (2000). *Explanation and Cognition*. Cambridge, MA: MIT Press.

Kitcher, P. (1981). Explanatory Unification. *Philosophy of Science* 48: 507–531.

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

Lewis, D. (1999): *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lipton, P. (2001). What Good is an Explanation? In G. Hon & S. Rackover (eds.), *Explanation: Theoretical Approaches*, 43–59. Dordrecht: Kluwer.

Lipton, P. (2004). *Inference to the Best Explanation* (second edition). London: Routledge.

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

Lombrozo, T. (2011). The instrumental value of explanations. *Philosophy Compass* 6: 539–551.

Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. G. Morrison (eds.): *Oxford Handbook of Thinking and Reasoning*, 260–276. Oxford, UK: Oxford University.

Machamer, P. K., Darden, L., and Craver, C.F. (2000). Thinking about mechanisms. *Philosophy of Science* 67: 1-25

McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science* 54: 553–567.

Oaksford, M., and Chater, N. (2007). *Bayesian Rationality*. Oxford: Oxford University Press.

Okasha, S. (2000). Van Fraassen's Critique of Inference to the Best Explanation. *Studies in History and Philosophy of Science* 31: 691–710.

Peirce, C. S. (1931-1935). *The Collected Papers of Charles Sanders Peirce. Vol. I-VI*. Eds. C. Hartshorne and P. Weiss. Cambridge/MA: Harvard University Press.

Popper, K. R. (1934/2002). Logik der Forschung. Berlin: Akademie Verlag. Translated as *The Logic of Scientific Discovery*. London: Routledge.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Railton, P. (1989). Explanation and metaphysical controversy. In P. Kitcher & W. C. Salmon (eds.): *Scientific explanation: Minnesota studies in the philosophy of science*, Vol. 8, 220–252. Minneapolis: University of Minnesota Press.

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

Salmon, W. (1971/1984). Statistical Explanation. Reprinted in Salmon (1984): *Scientific Explanation and the Causal Structure of the World*, 29–87. Princeton: Princeton University Press.

Schupbach J. (2011a). *Studies in the Logic of Explanation*. PhD Thesis, University of Pittsburgh.

Schupbach J. (2011b). Comparing Probabilistic Measures of Explanatory Power. *Philosophy of Science*, 78: 813–829.

Schupbach, J., and Sprenger J. (2011). The Logic of Explanatory Power. *Philosophy of Science* 78: 105–127.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition* 52: 1–21.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Cambridge/MA: Harvard Unviersity Press.

Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Sciences* 12: 435–502.

Trout, J. D. (2002). Scientific Explanation and the Sense of Understanding. *Philosophy of Science* 69: 212–233.

Trout, J. D. (2007). The Psychology of Scientific Explanation. *Philosophy Compass* 2: 564–591.

Van Fraassen, B. C. (1989). *Laws and Symmetry*. New York: Oxford University Press.

Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese* 167: 125–143.

Woodward, J. (2003). *Making Things Happen*. New York: Oxford University Press.