Explanation, Understanding, and Belief Revision*

Andrés Páez
Universidad de los Andes

Contents

1 Introduction

For the longest time, philosophers of science avoided making reference to the notion of understanding in their accounts of explanation. Although Hempel, Salmon, and other philosophers who wrote about the subject in the 20^{th} century recognized that understanding is one of the main goals of science, at the same time they feared that any mention of the epistemic states of the individuals involved would compromise the objectivity of explanation. Understanding is a pragmatic notion, they argued, and

*Previous versions of this paper were presented at the *Biennial Meeting of the Philosophy of Science Association* in Vancouver, at the *XVI Jornadas de Epistemología e Historia de la Ciencia* in Córdoba (Argentina), and at the *III Conference of the Latin American Association for Analytic Philosophy* in Buenos Aires. I am grateful to the audiences in these venues for their useful comments and questions.

although a subject worthy of psychological investigation, pragmatics should be kept at a safe distance from the universal, epistemological features of explanation. Although this attitude towards the notion of understanding has changed in the last decade[1], there are still many misgivings about using pragmatic notions in the analysis of one of the central concepts in the philosophy of science[2].

My main purpose of this paper is to defend the idea that there is a sense in which it is meaningful and useful to talk about understanding in an objective sense[3], and that to characterize this notion it is necessary to formulate an account of scientific explanation that makes reference to the doxastic states and epistemic goals of the participants in a cognitive enterprise. It is important to clarify at the outset that my goal is not to offer a general analysis of the notion of understanding, and that my approach is restricted to the understanding of singular facts in well-defined scientific contexts.

The essay is divided as follows. In the next section I introduce three theses about scientific explanation that will serve as the basis for the rest of the discussion. The first thesis, which is defended in sections 3 and 4, states that determining the potential explanations of a fact is essentially a non-pragmatic matter. This thesis is meant to allay the fears of those who see the introduction of pragmatic factors as the beginning of the road towards an unbounded relativism. Since the objective basis of explanation will be probabilistic, at the beginning of the paper I include a detailed discussion about the way in which probability will be used in my account of explanation. The second thesis, which is presented in section 5, states that it is possible to determine the epistemic value of most potential explanations of a fact, and that such value can be established in a non-arbitrary way despite being the result of the evaluation of individual researchers. Finally, towards the end of the essay I explain the third thesis, which establishes the criteria for the acceptance of an explanation in the corpus of beliefs of those researchers involved. These criteria are based on their joint assessment of the credibility and epistemic value of potential explanations.

2 Three Theses about Explanation

It has often been said that explanation is an interest-relative notion. Different inquiring agents impose different demands on the information they regard as explanatorily valuable. The interest-relativity of explanation has been accounted for in several ways: some authors have proposed a contrastive analysis of the explanandum (van Fraassen,

[1] See, for example, de Regt (2009), de Regt, Leonelli & Eigner (2009), Fey (2014), Grimm (2008), Khalifa (2012), Kvanvig (2009), and Strevens (2008, 2013).

[2] See, for example, Trout (2002, 2007) and Craver (2013) for more recent defenses of a purely ontic approach to explanation.

[3] Objective understanding in this sense will turn out to be the opposite of what de Regt (2009, p. 585) calls "the objectivist view of the relation between explanation and understanding," which he attributes to Hempel and Trout.

1980; Lipton, 2004) or a detailed description of the illocutionary context of an explanatory speech act (Achinstein, 1983). In my view, the interest-relativity of explanation has a much deeper origin. It derives from the interest-relativity of inquiry in general. Different agents use information for different purposes, and their acceptance of new information is directed by their cognitive and practical interests and goals. Far from being a superficial characteristic of inquiry, I believe that this is a fundamental trait of the acquisition of knowledge in general. The cost and effort that goes into obtaining new information makes the beliefs[4] that an inquiring agent has accepted a valuable asset that must be treated with care. Gratuitous losses must be prevented and the agent's acceptance of new information always involves the risk of bringing error into his system of beliefs. The risk must always be compensated by an epistemic incentive that outweighs the cost.

One of the biggest epistemic incentives of all is to obtain understanding of a fact.[5] But if understanding a given fact fulfills no purpose in the eyes of an inquiring agent, he will be more reluctant to incur the risks involved in accepting an explanation of it. On the other hand, if understanding a fact fulfills the cognitive interests and goals of the agent, but the information explains too much, it might be too good to be true. The acceptance of an explanation thus requires a delicate balance between two conflicting cognitive goals: the acquisition of valuable explanatory information and the avoidance of error.

The account of explanation that I present in this paper takes into account the difference between information and informational value, between the informational content of an explanation and the epistemic value of that content. When an agent seeks to expand his beliefs, his interest is restricted to information that promotes his cognitive goals or that is relevant to the problems he is trying to solve. In Catherine Elgin's words, "truth does not always enhance understanding. An irrelevant truth is epistemically inert" (1996, p. 124). I will argue that the goal of an inquiring agent is not just to find factually accurate explanations; it is to find explanations that are both factually accurate and epistemically valuable. This idea is captured by the following three theses:

1. Whether a piece of information is a *potential explanation* of the fact that P is mostly a non-pragmatic matter.

[4] In this paper beliefs should be understood as an agent's epistemic commitments, in the sense of Levi (1980). Some authors, such as Cohen (1989), use the term 'acceptance' for such attitudes, reserving the term 'belief' for involuntary epistemic states, akin to feelings. There is an extensive literature on the distinction between acceptance and belief (e.g., Engel, 2000; Cresto, 2009, among many others), but I cannot discuss the issue in this essay.

[5] Understanding laws and regularities is, of course, an equivalent or even greater epistemic incentive. The account presented here is restricted to the explanation of singular facts because the well-known objections against the explanation of laws require an entirely different analysis, one that most likely will not be probabilistic.

2. It is possible to determine the *objective epistemic value* of a subset of all the potential explanations of the fact that P.

3. In trying to understand the fact that P, an inquiring agent should only accept the potential explanations with positive objective epistemic value.

In the rest of the paper I discuss and defend each of these three theses.

3 The Objective Basis of Explanation

In this section and the next I defend the first of the three theses stated above, namely, that determining the potential explanations of a given fact is mostly a non-pragmatic matter. My basic contention is that an explanation of a singular fact should provide the information required to integrate the explanandum into an agent's cognitive system. An explanation should provide some of the factors that contributed to make P a fact, and some of the obstacles that could have, but did not prevent it from being one. Without such information, P will describe a brute fact, isolated from the rest of the agent's beliefs about the world. Probability sentences are the connecting tissue of an agent's corpus of beliefs. The influence of the preventing and contributing factors is captured by probability sentences of the form $p(P|Q) > p(P| \sim Q)$ and $p(P|Q) < p(P| \sim Q)$ that indicate that the fact that Q is statistically relevant to the explanandum[6].

The notion of statistical or probabilistic relevance has been used by many authors in the analysis of explanation. The best-known examples are Hempel's (1965) I-S model, Salmon's (1971, 1984) S-R model, Railton's (1978) D-N-P model, and Fetzer's (1974) causal-relevance model. All of these accounts consider precise probability values to be an essential part of an explanation. In contrast, I will argue that reference to probability values is largely unnecessary. Probability values have descriptive, predictive, and evidential value, but not explanatory value.

3.1 Probability Values

Probability values are thought to be important for two different reasons. If a statistical explanation is conceived of as an inductive argument, as it was in Hempel's original Inductive-Statistical model, the degree of expectation that a body of evidence confers upon a given event must be very high. Thus the value of the inductive probability must be kept in check to make sure it does not fall below a certain threshold as inquiry proceeds. On the other hand, if a statistical explanation is understood as an objective

[6] Many authors have used probabilities to model the epistemic states of researchers (e.g. Boutilier, 1995; van Fraassen, 1995; Halpern, 2003; van Benthem, 2003; Arló-Costa & Parikh, 2005). My account uses probability sentences to model the doxastic basis of an explanation, but an agent's epistemic states should not be understood to be probabilistic.

account of the stochastic process involved, as it is in Salmon's and Railton's models, it is crucial to avoid the attribution of false probability values to the probabilistic laws.

In response to criticism by Jeffrey (1971), Hempel (2001) gave up the high probability requirement, together with the claim that the explanans of an I-S explanation should show that the phenomenon described by the explanandum sentence was to be expected. Without this claim, however, the first reason to attribute any importance to probability values disappears. If the explanans is not supposed to justify our expectations that the explanandum will occur, there is no need to make sure that the value of the probability remains over a certain threshold.

Before we can evaluate the second reason why probability values are deemed to be explanatory, we must take a closer look at the logical structure of statistical explanations. One of the features of probability theory is that it does not have a weakening principle. A sound inductive argument that strongly supports its conclusion can be transformed into one that strongly undermines its conclusion with the insertion of additional true premises. An individual event can be referred to different reference classes, and the probability of the property associated with the event can vary considerably from one class to another. Hence, a body of evidence may confer a high degree of expectation upon a given event, while another body of evidence may confer a very low degree of expectation upon the same event. This is the problem that Hempel called the *ambiguity* of I-S explanation.

Hempel's partial solution to the problem is the requirement of maximal specificity. The requirement states that an acceptable statistical explanation should be based "on a statistical probability statement pertaining to the narrowest reference class of which, according to our total information, the particular occurrence under consideration is a member" (1965, p. 398). The requirement does not completely eliminate the ambiguity because the narrowest reference class can only be determined in the light of our current knowledge. It does not guarantee that there are no unknown statistical generalizations that can be used to construct a rival argument. In fact, Hempel claimed that "*the concept of statistical explanation for particular events is essentially relative to a given knowledge situation as represented by a set K of accepted sentences*" (p. 402, emphasis kept).

Salmon (1971) showed that the requirement of maximal specificity failed to rule out counterexamples in which irrelevant information finds its way into the explanation. But his main reason to reject Hempel's solution to the problem was his strong conviction that the appropriate reference class for a statistical explanation is one that is *objectively* homogeneous, not one that is *epistemically* homogeneous.

The notion of an objective homogenous reference class amounts to this: For any given reference class A, and for any given property C, there is, *in principle*, a partition of that class into two subclasses $A \wedge C$ and $A \wedge \sim C$. A property C is statistically relevant to a property B within A if and only if $p(B|A \wedge C) \neq p(B|A)$. Using von Mises's concept of place selection, Salmon defines a *homogeneous reference class* as follows:

> If every property $[C_1, C_2, C_3, \ldots, C_n]$ that determines a place selection is statistically irrelevant to B in A, I shall say that A is a homogeneous reference class for B. A reference class is homogeneous if there is no way, *even in principle*, to effect a statistically relevant partition without already knowing which elements have the attribute in question and which do not (1971, p. 43).

Salmon then replaces Hempel's requirement of maximal specificity for the *reference class rule*: "Choose the broadest homogeneous reference class to which the single event belongs" (p. 43). This characterization of statistical explanations is supposed to avoid any epistemic relativity because any statement of the form $p(G|F) = r$ that meets the homogeneity condition must be regarded as a fundamental statistical law of nature. Its reference class cannot be further specified, not because we do not know how to make a further relevant partition, but because *in principle* it is impossible to make a further relevant partition.

Salmon then defines a statistical explanation as follows. If we want to know why a member of the class A has the property B, the answer will be a S-R explanation that consists of: (i) the prior probability that a member of the class A will have the property $B : p(B|A) = r$, (ii) a partition into homogeneous cells with respect to the property in question: $A \wedge C_1$, $A \wedge C_2$, etc., (iii) the posterior probabilities of the property in cells of the partition $p(B|A \wedge C_1) = r_1$, $p(B|A \wedge C_2) = r_2$, etc., and (iv) a statement of the location of the individual in question in a particular cell of the partition: "a is a member of $A \wedge C_k$" (pp. 76-77).

Salmon explicitly requires the use of probability values in providing an explanation. The use of probability values stems from the fact that the S-R model is at bottom a covering-law model. Since any statement of the form $p(G|F) = r$ that meets the homogeneity condition must be regarded as a fundamental statistical law of nature, each of the probability sentences in the explanans of a S-R explanation is a law of nature. And since the factive condition on explanation demands that every element in an explanation must be true, the probability assigned to the explanandum by each of these probability sentences must be the right one.

To see how restrictive this requirement is, consider the following example provided by Humphreys:

> If a man dies from lung cancer, having been a heavy smoker, omitting from a probabilistic explanation any of the following minor relevant factors will result in a false probability claim: cosmic radiation from Alpha Centauri, particles from a chimney in Salem, Oregon, and a smoke-filled room he entered briefly at the Democratic convention eight years ago. It is good to be strict in matters of completeness, but not to the point of absurdity (1989, p. 111).

Humphreys argues that if one insists in providing the exact probability of the explanandum as part of the truth conditions of an explanation, it will be impossible to distinguish between a complete explanation and a true explanation. The omission of absurdly small probabilistically relevant factors, known or unknown, will result in a false explanation.

How can a true but incomplete statistical explanation be provided? Humphreys argues that instead of focusing on probability values, we should focus on causal relevance. An explanation should provide one or more of the factors that are causally relevant to an explanandum, and a factor is causally relevant if it changes the propensity for an outcome. His strategy has the advantage that it makes it possible to offer a true explanation of an event by providing a contributing or a counteracting cause even in cases where the other factors are not known and the true probability value cannot be calculated.

3.2 Epistemic Relativity

Although Humphreys' approach offers an appropriate formal basis for providing a statistical explanation, there is an obvious objection. As the many versions of Simpson's Paradox illustrate, one or more of the factors that the agent is unaware of can turn a contributing cause into a counteracting cause, or vice versa. Humphreys' response to this objection is puzzling. He says: "Of course, epistemically, we can never know for certain that such confounding factors do not exist, but that is an entirely separate matter, although regrettably relative frequentists have often failed to separate epistemic aspects of probabilistic causality from ontic aspects" (p. 114).

It seems to me that it is Humphreys who is guilty of not keeping epistemic and ontic matters in separate baskets. If Salmon's model is too demanding, as Humphreys maintains, it is because we can never know if we have met all the conditions that it imposes on explanation. But Humphreys' account suffers from a similar problem. In order for something to be a contributing or a counteracting cause in Humphreys' sense, there cannot be *any* further factor, known or unknown, that will invert the influence of these causes on the explanandum, or that will neutralize them altogether. Thus an agent who offers a causal statistical explanation will always have to relativize the explanation to a knowledge situation.

The accounts offered by Salmon and Humphreys avoid the epistemic relativity of statistical explanation by introducing a condition that effectively rules out the possibility that a bona fide statistical explanation will be defeated by a rival statistical claim. But the cost of avoiding the epistemic relativity of explanation is to render useless their accounts of explanation. It is hard to see how such a relativization can be eliminated if we want to provide a coherent picture of the role of explanation in inquiry. If we adopt the view that epistemic relativity is an unacceptable feature of explanation, we will be forced to conclude that there has never been a genuine scientific explanation in the history of science. Furthermore, we lose one of the main incentives for any scientific inquiry. Why would anyone want to incur the cost and effort involved in searching for explanations if the results cannot be assumed to be true in future decisions and deliberations? In Isaac Levi's words,

> If inquiry cannot be motivated by a concern to remove doubt, what is its rationale? If we cannot incorporate the solutions we come close to establishing into the evidence and background information for future investigations, why should we care that we come close? The truth of the well-established conjecture remains an open question and a legitimate issue for future investigation. Inquiry never settles anything and, hence, inquiry–even inquiry into a specific problem–never legitimately terminates because the matter is settled but only, so it seems, because the investigators are tired or bored or have run out of funds. No matter how minute a question might be, if inquiry into that question is free of costs, it should go on forever (1991, p. 2).

The reference to a specific epistemic context in the characterization of explanation is clearly a departure from tradition. Many philosophers have claimed that pragmatic elements have no place in the study of explanation. They recognize that there are interesting issues associated with the process of providing an explanation in an actual context, and their intention is not to belittle their importance. But the concept of explanation that they characterize is, in Hempel's words, "a concept which is abstracted, as it were, from the pragmatic one, and which does not require relativization with respect to questioning individuals any more than does the concept of mathematical proof" (1965, p. 426). The same general idea is defended by many other philosophers of science.

Michael Friedman has pointed out that there is a certain equivocation about the term 'pragmatic'. 'Pragmatic' can mean roughly the same as 'psychological', i.e., having to do with the thoughts, beliefs, attitudes, etc. of individuals. But 'pragmatic' can also be synonymous with 'subjective'. In the latter sense, a pragmatic notion must always be relativized to a particular individual. Friedman's claim is that "a concept can be pragmatic in the first sense without being pragmatic in the second." Further on he explains: "I don't see why there can't be an objective or rational sense of 'scientific understanding', a sense on which what is scientifically comprehensible is constant for a relatively large class of people" (1974, p. 8).

The traditional avoidance of any pragmatic element in a theory of explanation can thus be evaluated in two different ways. If one takes 'pragmatic' to mean the same as 'subjective', the insistence in providing a non-pragmatic analysis of explanation, i.e., an analysis that does not depend on the idiosyncrasies of the individuals involved, is perfectly justified. But if 'pragmatic' is interpreted in Friedman's first sense, there is no reason why an analysis of the concept of explanation should not make reference to the epistemic states of the individuals involved in a cognitive project.

I believe that we should take Friedman's suggestion seriously and explore the possibility of characterizing, in logically precise terms, a notion of explanation that is both objective and pragmatic, that does not depend on the idiosyncrasies of the individuals involved but that regards their epistemic states, their shared commitments, and their cognitive interests and goals as a fundamental part of the analysis. The concept of explanation will still be an "abstraction", in Hempel's sense, but an abstraction based on the decisions that take place when a group of inquiring agents rationally accept explanatory information. The resulting concept will be a hybrid, a combination of the formal, semantic, and pragmatic dimensions of explanation.

4 Potential Explanations

The epistemological framework for the account of explanation that I will present is Isaac Levi's version of the belief-doubt model first proposed by Peirce (1877)[7]. According to the belief-doubt model, an inquiring agent presupposes that everything he is currently committed to fully believing is true. This does not mean that truth or falsity is relative to what the agent believes. But the agent's *judgments* of truth are relative to what he believes. If the agent is concerned with establishing true explanations of phenomena, his decision to accept an explanation can only be made relative to the judgments of truth available to him.

To claim that an inquiring agent presupposes that everything he is currently committed to fully believing is true is not to say that he cannot change his mind. Certainty or full belief does not entail incorrigibility. Levi explains the claim thus: "To regard some proposition as certainly true and as settled is to rule out its falsity as a serious possibility for the time being. . . . But from this it does not follow that good reasons will not become available in the future for a change of mind and for calling into question what is currently considered to be true" (1991, p. 3). Peirce puts it more graphically: "The scientific spirit requires a man to be at all times ready to dump his whole cartload of beliefs, the moment experience is against them" (1931, p. 55).

An inquiring agent has no doubt that all the sentences in his corpus of beliefs are true. Nonetheless, he does not regard all of the facts stated by these sentences as being equally well understood. The degree to which an agent understands the fact expressed by a sentence P will depend on how well integrated P is to the agent's cognitive system. It will not depend on how much support it has or on how epistemically entrenched

[7] Although my account of explanation uses Levi's belief revision theory as theoretical framework, it must be pointed out that Levi does not agree with my approach (personal communication). The main reason is that Levi believes that all explanations with probabilistic premises presuppose a D-N explanation stated in dispositional terms. Furthermore, Levi associates statistical explanations with the elimination of surprise and an increase in the expectation of the occurrence of the explanandum (Levi, 1988, 1995). The account of explanation that I present here does not entail those two consequences.

it is. On the one hand, if a sentence has been accepted in his corpus of beliefs, it is judged to be true and no further argument is necessary. On the other hand, poorly understood phenomena can be highly epistemically entrenched, and completely useless facts can be very well understood.

According to the belief-doubt model, an inquiring agent's judgments of truth are always relative to what he is currently committed to fully believing. Thus, an agent's decision to accept an explanation can only be made relative to the judgments of truth available to him. Naturally such decisions will lack any sort of *objectivity*. An agent who wants to claim objectivity for the explanations that he accepts must first make sure that the explanation is consistent with K, the set of beliefs that represents the shared agreement between the members of a community of experts. More technically, the states of belief of the individual experts can be partially ordered in a manner satisfying the requirements of a Boolean algebra. In consequence, it will be possible to form the meet of their individual states, i.e., the strongest common consequence of all their states of belief (Levi, 1991, p. 13).

Let P be a sentence in K. A set of sentences E is a *potential explanation* of the fact stated by P relative to K just in case the following conditions are fulfilled:

(i) $K \cup E$ is consistent.

(ii) $E \not\subset K$.

(iii) There is a sentence Q such that $Q \in E$.

(iv) Either $p(P|Q) > p(P|\sim Q) \in E$ or $p(P|Q) < p(P|\sim Q) \in E$.

(v) There is no $R \in K$ such that $p(P|Q\&R) = p(P|\sim Q\&R)$.

(vi) P and Q are logically independent.

(vii) Nothing else is an element of E.

The first condition states that a potential explanation must be consistent with the corpus of beliefs in which the explanandum is accepted. The second condition states that the potential explanation cannot be already accepted in K. The third condition says that the potential explanation must include a singular sentence Q that describes a potentially relevant factor. The fourth condition states that Q is positively or negatively relevant to the fact that P. The fifth condition guarantees that P and Q will not be spuriously correlated, as far as we know. Condition (vi) guarantees that P will not explain itself. It also prevents the inclusion of trivial cases in which $p(P|Q) = 1$ because $P \vdash Q$. A potential explanation is thus a set containing a singular sentence that describes a fact, and a probability sentence that states the potential statistical relevance of that fact to the explanandum.

Using this definition of a potential explanation, we can now characterize the notion of an *explanation space*. An explanation space can be understood as the set of sentences that contains all the potential explanations of P, regardless of whether the inquirers are aware of them or not.

(E_P) For every sentence P in K, there is a set $\{E_1, E_2, \ldots, E_k\}$ such that E_i is an element of the set iff it is a potential explanation of P. The set, denoted E_P, is the *explanation space* of P.

The explanation space will contain logically equivalent and empirically equivalent potential explanations. On the one hand, if $E_1 = \{Q, p(P|Q) > p(P| \sim Q)\}$ and $E_2 = \{R, p(P|R) > p(P|\tilde{R})\}$, where Q and R are logically equivalent, then E_1 and E_2 are logically equivalent potential explanations. If an agent accepts E_1, she is thereby committed to E_2. On the other hand, if Q and R contain coextensive singular terms or predicates that occupy the same places in Q and R, E_1 and E_2 will be empirically equivalent potential explanations. However, the explanatory value and the credibility of E_1 and E_2 will not be assessed in the same way unless the agents who assess them are aware that the singular terms or predicates are coextensive[8].

5 The Epistemic Value of Explanation

Consistency with K, the set of beliefs that represents the shared agreement between the members of a learning community, is not enough to guarantee the objectivity of an explanation. The objectivity of our conjectures lies, as Popper correctly points out, "in the fact that they can be intersubjectively tested" (1959, p. 44). The intersubjective test that an explanation must pass is the evaluation of its credibility and of its explanatory value in the eyes of the experts.

Suppose a group of inquirers–a community of experts in the field–wants to consider the adoption of an explanation. To do so, they must first adopt a belief state K representing the shared agreement between them. Such a belief state will be the strongest common consequence of all their states of belief. Obviously, such a state will contain more than just singular sentences representing facts and probability sentences. It will also include sentences that state which are the most relevant problems in the field, what type of experiments and observations are considered more reliable, in addition to basic methodological and reasoning principles.

Once the members of the community of experts have accepted a common corpus K, they must take it as the basis for establishing a set of potential explanations of the problem at hand, For example, suppose a group of inquirers are trying to establish why P. They must initially agree on a set of ground facts and low-level hypotheses.

[8] Condition (iv) also introduces an element of epistemic relativity because the non-existence of a screening off factor can only be guaranteed relative to K.

Statistical data and the chronology of the explanandum will be easy to agree upon. The explanation of some aspects of the phenomenon can be non-controversially accepted, while the explanation of others will be a matter of heated debate. After the inquirers have agreed on a common corpus of beliefs K, they can put together a *set of explanatory options*, denoted O_P, which will include all the factors consistent with K that might explain P and that have been identified by the inquirers. At this stage of inquiry it does not matter whether the potential explanations are uncontroversial or completely outlandish, as long as they are somehow relevant to the problem at hand and consistent with K, that is, if they fulfill the requirements to be in E_P.

It is possible for a group of agents to share the same information and yet disagree about the degree of belief or credal probability that they assign to the information in the set of explanatory options. Since the agents do not want to beg the question by assigning the highest marks to their favorite explanations, they must adopt a common credal probability measure. A common strategy to eliminate the conflict between different credal probability distributions is to represent the shared agreement as the weighted average of the distributions in conflict. The resulting credal probability function C determines the objective risk of error incurred in accepting a potential explanation in O_P. Let E_i be the conjunction of the elements of a potential explanation E_i in O_P, i.e., the conjunction of a singular sentence and a probability sentence. For every potential explanation E_i, the risk of error is $1 - C(E_i)$.

On the other hand, different inquirers will disagree in their assessment of the importance of the explanations contained in the set of explanatory options. Despite these differences, there must be a minimal objective criterion to measure the explanatory value of any potential explanation. That criterion is the new information carried by the potential explanation, which, following Levi, I identify with its logical strength. The set of potential expansions of a belief set K can be partially ordered by a classical consequence relation. The set is a Boolean algebra in which the minimum is K and the maximum is the inconsistent state. If a probability function M is defined over this set, and if the only element that has probability zero is the inconsistent state, potential expansions of K will strictly increase in probability with a decrease in logical strength. When the M-function is defined over the set of potential explanations of interest to the inquirer, we obtain a measure of the informational content of the potential explanations in O_P. The measure of the informational content of a potential explanation E_i, denoted Cont(E_i), is $1 - M(E_i)$.

The informational content of a potential explanation is the first objective criterion that should be used in assessing the explanatory value of the elements of O_P. The evaluation of their explanatory value is subject to the following weak monotonicity requirement (WMR):

> (WMR) If a potential explanation E_1 in O_P carries at least as much information as another potential explanation E_2 in O_P, E_1 carries at least as much explanatory value as E_2.

Not all potential explanations of the fact that P are comparable in terms of logical content. Since the community of experts wants to consider all the explanations available to them, they might invoke further criteria in order to complete the quasi-ordering imposed by the weak monotonicity requirement. In order to assess the explanatory value of the remaining elements of O_P, they can evaluate if they have certain properties that are considered explanatorily virtuous.

There are several explanatory virtues mentioned in the philosophical literature. Friedman (1974) and Kitcher (1989), for example, argue that explanations improve our understanding through the unification of our knowledge. Explanations that reduce the number of independent assumptions we have to make about the world are to be preferred to those that do not. This suggests that the potential explanations in O_P could be ordered according to some set of rules that determines their unifying power.

The problem is that neither Friedman nor Kitcher have provided an account that can be applied to explanations generally. Friedman's original argument was intended as an account of the explanation of scientific laws. Friedman argued, for example, that the kinetic theory of gases is explanatory because it unified different laws and properties of gases that were previously disconnected. Friedman's only attempt to formalize and generalize the idea of explanation as unification was incisively criticized by Kitcher (1976) and Salmon (1989).

But Kitcher's account is no more helpful that Friedman's. According to Kitcher, the explanatory worth of candidates cannot be assessed individually. In his view, a successful explanation earns that name because it belongs to the explanatory store, a set that contains those derivations that collectively provide the best systematization of our beliefs. 'Science supplies us with explanations whose worth cannot be appreciated by considering them one-by-one but only by seeing how they form part of a systematic picture of the order of nature" (1989, p. 430). The idea that a virtuous explanation should have the potential to unify our beliefs is uncontroversial, but no one, to my knowledge, has provided a general account of explanation as unification that is not restricted to the case of scientific laws or scientific explanatory exemplars.

Mellor (1995) provides an account of explanatory value that is better suited for our purposes.

Mellor approaches explanation via his theory of causation. The theory requires every cause to raise the chances of its effects. That is, a fact C causes a fact E iff $ch_C(E) > ch_{\sim C}(E)$. When causes are used in the explanation of a given fact, Mellor argues that the explanans must necessitate its explanandum, or at least raise its probability as much as possible, thereby reducing its chance of not existing. Thus, he concludes, "the more C raises E's chance the better it explains it" (p. 77). If we were to accept Mellor's idea, it would be possible to order the potential explanations in O_P according to the difference between $ch_C(E)$ and $ch_{\sim C}(E)$.

The main problem with Mellor's proposal is that when we examine a genuinely stochastic process, the value of the chance that the cause confers on the explanandum will be irrelevant. As Jeffrey has convincingly argued, the information required to explain E is the same information used to explain $\sim E$, regardless of the value of the chance. Furthermore, if E is sometimes randomly caused by C and sometimes randomly caused by C^*, and $ch_C(E) > ch_{C*}(E)$, there is no reason to think that C is a better explanation than C^*.

Mellor will respond to the objection by claiming that chances measure possibilities. "The less possible $\sim E$ is, i.e. the less $ch(\sim E)$ is and hence the greatest $ch(E)$ is, the closer the fact E is to being necessary. This is the sense in which a cause C may explain E better or worse, depending on how close it comes to making E necessary, i.e. on how much it raises $ch(E)$" (p. 77). Independently of whether we can make sense of such concepts as *almost necessary* or *nearly impossible*, it is not clear how such notions would enhance our notion of explanation. Probabilities are important in statistical contexts because knowing that C raises the chance of E allows us to know what makes E possible, and because the chance that C gives E allows us to adjust our expectations of E's occurrence. But it seems to me that mixing chances and possibilities adds nothing to our understanding of why E is a fact.[9]

A third candidate for judging the epistemic value of an explanation is Whewell's (1837) notion of *consilience*. Consilience is intended to serve as a measure of how much a theory explains, and it can therefore be used to compare the explanatory value of two different hypotheses. One hypothesis has more explanatory value than another if the former explains more of the evidence than the latter. Thagard (1978) provides compelling evidence that this idea is often used by scientists in support of their theories. For example, Fresnel defended the wave theory of light by saying that it explained the facts of reflection and refraction at least as well as did the particle theory, and that there were other facts involving diffraction and polarization that only the wave theory could explain. Translated into my account, this means that if E_i raises the probability of more facts connected to the explanandum than E_j, then E_i is a better explanation than E_j.

The problem with consilience is that, once again, the account works well in the explanation of laws, but it will not work in the explanation of singular facts. Whether a given fact explains more aspects connected to the explanandum than another fact is hard to say. We would have to define what a fact "connected to the explanandum" is, and it is doubtful that a non-pragmatic formalization of this notion can be found. Besides, sometimes a theory can explain too much. Lavoisier accused the phlogiston theory of this particular crime.

Are there any other criteria that will allow us to assess the explanatory value of the potential explanations in O_P? We still have not examined the values that are usually mentioned in the context of theory choice: simplicity, accuracy, fruitfulness, and so

[9] In Páez (2013) I offer an exhaustive analysis of the relation between causation and explanation in Mellor's work.

on[10]. But such an analysis is unnecessary. If the criteria are such that the community of experts can agree on their importance and on how they should be applied in particular cases, they can be added to the belief state K that represents their shared agreement. The agents will then be able to complete, to some degree, the quasi-ordering generated by the monotonicity condition with respect to the M-function. But to expect a complete agreement in the way that all the agents engaged in common inquiry assess the explanatory value of different potential explanation is to expect a heterogeneous group of inquirers to agree on what aspects of reality they find interesting or useful.

If a common decision is required nonetheless, the community of experts can adopt the following compromise. The agents must first identify the elements of the set O_P that can be completely ordered because they are comparable in terms of strength or because they can be compared using the criteria to evaluate explanatory value that they have incorporated to K. The agents can then agree to disagree about the explanatory value of the remaining elements of O_P. Let O_P^* be a set of explanatory options such that $O_P^* \subseteq O_P$ and such that the M-value of each element of the set is determined. Combining the credal probability function C with the M-function defined over the elements of O_P^* we obtain a value that the community of experts can use to select the best explanation of P. I will call this result the *objective epistemic value* of a potential explanation[11]:

(OEV) $V(E_i) = \alpha C(E_i) + (1-\alpha)\text{Cont}(E_i)$

The agents' interest in valuable information should not outweigh the desideratum to avoid error; thus $\alpha \geq 0.5$. And since the information they seek should not be worthless, $1 > \alpha$.

Now, some researchers will be bolder than others in privileging content over credibility, while others will be more cautious and adopt the opposite attitude. Let q be a common boldness index, which is the average of their individual boldness indices. If $q = (1-\alpha)/\alpha$, we obtain the following affine transformation of OEV:

(OEV) $V(E_i) = C(E_i) - qM(E_i)$

The experts should reject a potential explanation in O_P^* if OEV is negative, remain uncommitted if it is 0, and accept it if it is positive. Any potential explanation in O_P^* with positive objective epistemic value is an *objective explanation* of P in K. The disjunction of all such objective explanations is *the* objective explanation of P in K:

(OE_P) *The objective explanation* of P in K, denoted OE_P, is the disjunction of all the potential explanations in O_P^* with positive objective epistemic value.

One of the consequences of taking the functions C and M –which represent the average credibility and the agreed upon explanatory value, respectively– as a basis

[10] There is a vast literature on the epistemic and social values used in science. The compilations by Machamer and Wolters (2004) and Kinkaid, Dupré and Wylie (2007) offer a contemporary perspective on the topic.

[11] This strategy is similar to the one followed by Levi (1991) to characterize the maximization of the expected epistemic utility obtained by expanding a corpus of beliefs.

for the analysis of the potential explanations in O^*_P is that each individual agent was forced to sacrifice his personal evaluation of credibility and value in order to accept the verdict of the community of experts. Suppose an agent has accepted a potential explanation of P based on his individual assessment of its credibility and explanatory value. Now suppose that he submits his "subjective" explanation to the community of experts, and the explanation is judged to be maximally credible and maximally valuable by the community, thus becoming an objective explanation. Does the agent understand *more* now that his explanation has been certified by others? It seems to me that he does not. But if the agent does not obtain more understanding from this recognition, why should anyone seek objectivity for an explanation that he or she already believes?

Part of the answer is that the belief-doubt model is not a recipe for dogmatism. A seldom-noted fact about inquiry is that most newly suggested explanatory hypotheses do not survive the test of intersubjective scrutiny. If the agent is aware of this fact– and he should be if he is a responsible inquirer-it would be imprudent for him to give his full assent to an explanatory hypothesis that contradicts firmly established theories and findings without obtaining at least a partial intersubjective assessment of its merit. An agent does not need to fully believe that an explanation is true to obtain the understanding that the explanation provides. Any inquirer can explore the consequences of a hypothesis by assuming, for the sake of argument, that it is true. If the hypothesis is judged to have positive objective epistemic value by a community of experts, the inquirer will then be fully justified in giving it his full assent.

But the question remains. If the agent does not obtain new understanding in the approval that he receives from his peers, why should he seek their approval? What prevents an agent from individually assessing the credibility and explanatory value of a potential explanation, and deciding to fully believe it if his individual understanding is thereby increased? In other words, why should objectivity matter? The answer is that objectivity itself is a property of information that some agents find valuable and some do not. An agent who decides to be a member of a learning community does so because he is convinced that his beliefs will be more valuable if they are objective. Other agents will find that objectivity adds no value to their corpus of beliefs. Just as there is a difference between objective and subjective explanation, there is an analogous distinction between objective and subjective understanding. The latter is the type of understanding that Hempel (1965) correctly believed should be shunned at all costs from an account of scientific explanation. But the reason it should be shunned is not that it is an inferior type of understanding. The reason is that the members of a scientific community are among the many agents who find objectivity valuable. Therefore, an account of scientific explanation should avoid any reference to an evaluative process in which the agent shows no concern for the views of others.

6 Conclusion

The belief-doubt model provides an adequate basis for an account of explanation that takes into consideration the epistemic value of the information that we acquire through inquiry. By including the shared commitments and the cognitive interests and goals of the individuals engaged in a cognitive enterprise, we obtain a notion of explanation that is objective by any reasonable standard of objectivity, and that clarifies the connection between explanation and understanding. The main reason why I have adopted the belief-doubt model is that an account of explanation that takes into consideration the epistemic value of the information that we acquire through inquiry leads to a natural resolution of the conflict between the purely pragmatic approach to explanation defended by Achinstein and van Fraassen, for example, and the more common approach in which pragmatic considerations are not assigned any serious role.

References

[Achinstein(1983)] P. Achinstein. *The nature of explanation*. New York: Oxford University Press, 1983.

[Aliseda(2006)] A. Aliseda. *Abductive reasoning*. Dordrecht: Springer, 2006.

[Arló-Costa and Parikh(2005)] H. Arló-Costa and R. Parikh. Conditional probability and defeasible inference. *Journal of Philosophical Logic*, 34:97–119, 2005.

[Boutilier (1995)] C. Boutilier. On the revision of probabilistic belief states. *Notre Dame Journal of Formal Logic*, 36:158–183, 1995.

[Boutilier and Becher(1995)] C. Boutilier and V. Becher. Abduction as belief revision. *Artificial Intelligence*, 77:43–94, 1995.

[Cohen(1989)] L. J. Cohen. Belief and acceptance. *Mind*, 98:367–389, 1989.

[Craver(2013)] C. Craver. The ontic conception of scientific explanation. In A. Hütteman & M. Kaiser (Eds.), *Explanation in the biological and historical sciences*. Dordrecht: Springer, 2013.

[Cresto(2010)] E. Cresto. Belief and contextual acceptance. *Synthese*, 177:41–66, 2010.

[de Regt, Leonelli, and Eigner (2009)] H. de Regt, S. Leonelli, and K. (Eds.) Eigner. *Scientific understanding: Philosophical perspectives*. Pittsburgh: University of Pittsburgh Press, 2009.

[Elgin(1996)] C. Z. Elgin. *Considered judgment*. Princeton: Princeton University Press, 1996.

[Engel(2000)] P. Engel. (Ed.) *Believing and accepting*. Dordrecht: Springer, 2000.

[Faye(2014)] J. Faye. *The nature of scientific thinking: On interpretation, explanation, and understanding*. New York: Palgrave Macmillan, 2014.

[Fetzer(1974)] J. H. Fetzer. A single case propensity theory of explanation. *Synthese*, 28:171–198, 1974.

[Friedman(1974)] M. Friedman. Explanation and scientific understanding. *Journal of Philosophy*, 71:5–19, 1974.

[Grimm(2008)] S. R. Grimm. Explanatory inquiry and the need for explanation. *British Journal for the Philosophy of Science*, 59:481–497, 2008.

[Halpern(2003)] J. Y. Halpern. *Reasoning about uncertainty*. Cambridge: MIT Press, 2003.

[Hempel(1965)] C. G. Hempel. *Aspects of scientific explanation*. New York: The Free Press, 1965.

[Hempel(2001)] C. G. Hempel: Postscript 1976. More recent ideas on the problem of statistical explanation. In J. H. Fetzer (Ed.), *The philosophy of Carl G. Hempel. Studies in science, explanation, and rationality*. New York: Oxford University Press, 2001.

[Humphreys(1989)] P. Humphreys. *The chances of explanation. Causal explanation in the social, medical, and physical sciences*. Princeton: Princeton University Press, 1989.

[Jeffrey(1971)] R. Jeffrey. Statistical explanation vs. statistical inference. In W. C. Salmon (Ed.), *Statistical explanation and statistical relevance*. Pittsburgh: Pittsburgh University Press, 1971.

[Khalifa(2012)] K. Khalifa. Inaugurating understanding or repackaging explanation? *Philosophy of Science*, 79:15–37, 2012.

[Kinkaid, Dupré, and Wylie (2007)] H. Kinkaid, J. Dupré, and A. Wylie. (Eds.) *Value-free science? Ideals and illusions*. New York: Oxford, 2007.

[Kitcher(1976)] P. Kitcher. Explanation, conjunction, and unification. *Journal of Philosophy*, 73:207–212, 1976.

[Kitcher(1989)] P. Kitcher. Explanatory unification and the causal structure of the world. En P. Kitcher & W. Salmon (Eds.), Scientific explanation. Minnesota studies in the philosophy of science, Volume XIII. Minneapolis: University of Minnesota Press, 1989.

[Kvanvig(2009)] J. L. Kvanvig. The value of understanding. In A. Haddock, A. Millar & D. Pritchard (Eds.), Epistemic value, Oxford: Oxford Unviersity Press, 95–111, 2009.

[Levi(1988)] I. Levi. Four themes in statistical explanation. In W. L. Harper & B. Skyrms (Eds.), Causation in decision, belief change and statistics, 1988.

[Levi(1991)] I. Levi. *The fixation of belief and its undoing*. New York: Cambridge University Press, 1991.

[Levi(1995)] I. Levi. *Review of The Facts of Causation* by D.H. Mellor. *Theoria*, 61: 283–288, 1995.

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

[Machamer and Wolters(2004)] P. Machamer and G. Wolters. (Eds). *Science, values, and objectivity*. Pittsburgh: University of Pittsburgh Press, 2004.

[Mellor(1995)] H. D. Mellor. *The facts of causation*. London: Routledge, 1995.

[Páez(2009)] A. Páez. Artificial explanations: The epistemological interpretation of explanation in AI. *Synthese*, 170:131–146, 2009.

[Páez(2013)] A. Páez. Probability-lowering causes and the connotations of causation. *Ideas y Valores*, 151:43–55, 2013.

[Peirce(1877)] C. S. Peirce. The fixation of belief. *Popular Science Monthly*, 12: 1–15, 1877.

[Peirce(1931)] C. S. Peirce. Lessons from the history of science. In C. Hartshorne, & P. Weiss (Eds.), Collected Papers of Charles S. Peirce, Cambridge: Belknap Press, 1931.

[Popper(1959)] K. Popper. *The logic of scientific discovery*. London: Hutchinson, 1959.

[Railton(1978)] P. Railton. A deductive-nomological model of probabilistic explanation. *Philosophy of Science*, 45:206–226, 1978.

[Salmon(1971)] W. C. Salmon. Statistical explanation. In W. C. Salmon (Ed.), *Statistical explanation and statistical relevance*. Pittsburgh: Pittsburgh University Press, 1971.

[Salmon(1984)] W. C. Salmon. *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press, 1984.

[Salmon(1989)] W. C. Salmon. *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press, 1989.

[Strevens(2008)] M. Strevens. *Depth: An account of scientifc explanation*. Cambridge: Harvard University Press, 2008.

[Strevens(2013)] M. Strevens. No understanding without explanation. *Studies in History and Philosophy of Science*, 44:510–515, 2013.

[Thagard(1978)] P. Thagard. The best explanation. criteria for theory choice. *Journal of Philosophy*, 75:76–92, 1978.

[Trout(2002)] J. D. Trout. Scientific explanation and the sense of understanding. *Philosophy of Science*, 69:212–233, 2002.

[Trout(2007)] J. D. Trout. The psychology of explanation. *Philosophy Compass*, 2: 564–596, 2007.

[van Benthem(2003)] J. van Benthem. Conditional probability meets update logic. *Journal of Philosophical Logic*, 12:409–421, 2003.

[van Fraassen(1980)] B. C. van Fraassen. *The scientific image*. Oxford: Clarendon Press, 1980.

[van Fraassen(1995)] B. C. van Fraassen. Fine-grained opinion, probability, and the logic of full belief. *Journal of Philosophical Logic*, 24:349–377, 1995.

[Whewell(1837)] W. Whewell. *History of the inductive sciences*. London: Parker, 1837.