Abduction and Inference to the Best Explanation[1]

Valeriano IRANZO

ABSTRACT: Aliseda's *Abductive Reasoning* is focused on the logical problem of abduction. My paper, in contrast, deals with the epistemic problems raised by this sort of inference. I analyze the relation between abduction and inference to the best explanation (*IBE*). Firstly a heuristic and a normative interpretation of *IBE* are distinguished. The epistemic problem is particularly pressing for the latter interpretation, since it is devoid of content without specific epistemic criteria for separating acceptable explanations from those which are not. Then I discuss two different normative interpretations of *IBE*. I. Niiniliuoto favours a "probabilistic-confirmational" translation of explanatory merit while S. Psillos thinks that the insight of *IBE* is lost in a pure probabilistic format. My conclusion is that Aliseda's theory of abduction fits better with a heuristic account of *IBE*.

Keywords: abduction, inference to the best explanation, reliabilism, confirmation, explanatory value.

1. Aliseda's approach to abduction

According to Aliseda 2006 (p. 46) the logical form of abduction can be represented as follows:

$$\Theta, \alpha \rightarrow \varphi \qquad\qquad (1)$$

where Θ and α refer, respectively, to the background knowledge and to the explanation obtained; '$\rightarrow$' designates the explanatory inferential link, and φ alludes to a particular bit of evidence, preferably some novel or anomalous fact.

In contrast to deduction, abduction is a sort of non-monotonic —ampliative— inference. It is necessary, then, to specify the background knowledge involved, Θ, since, even though it did not contain any falsity, further additions could turn an acceptable inference into an unacceptable one. On the other side, since Aliseda's research adopts a very general standpoint on abductive reasoning, she avoids any commitment to a particular theory of explanation. The symbol '$\rightarrow$' could be understood, then, as syntactic derivability, semantic entailment, probabilistic dependence, …, but neither of these interpretations is forced upon us. Aliseda claims that the *logical* challenge raised by abduction is to determine the formal conditions for generating all those αs that satisfy (1).

Nevertheless, there are better and worse explanations. In principle, we are interested in those that are more accurate. But there may be several explanations that fulfil the formal schema (1). They may also be incompatible. So, which one believe? Should we suspend judgment over all them? Are there further factors that could be relevant?

For this reason a complete analysis of abduction must include both the *generation* of abductions, where the logical challenge arises, and the *selection* over them.

Aliseda takes notice of this double dimension of abduction (*ibid.*, p. 33). She even suggests the possibility of including an additional parameter in (1) to rank the explanations obtained according to their quality. In a section entitled "Selecting the Best Explanation" (*ibid.*, pp. 72-74) she points out that simplicity and some other factors as likelihood or predictive power should be taken into account. However, she argues that "What makes a hypothesis good or best has no easy answer … As we have not been able to shed a new light from logic upon these matters, we will ignore these dimensions here" (p. 73). Consequently, Aliseda's approach focuses on the *logical* problem of abduction, that is, on "the characterization of abductive explanations as products [...] and with the processes for constructing these." (*ibid.*, p. 46).

In this paper, in contrast, our main concern will be how we could justify the preference criteria for selection among rival explanations. It's worth emphasizing that this is an e*pistemological* problem, since the relevant criteria intend to distinguish the true explanations —or the most probable explanations— from the false ones.

2. *Inference to the Best Explanation (IBE)*

Peirce used the term 'abduction' to refer both to generation and selection of hypotheses. Certainly, the distinction is not clear-cut. Perceptual judgments in ordinary conditions were understood by Peirce as "extreme cases of abductive inferences" where the abductive suggestion "comes to us as a flash" (CP, 5.181), with no consideration of any other alternatives. Concerning such episodes the distinction between generation and selection seems rather artificial. However, finding suitable explanations in scientific contexts use to be more complicated so it makes sense to distinguish between generation and selection. Firstly, there is an explicit search for alternative potential explanations which fit with the evidence. Then a subsequent comparison among them is made concerning the degree they satisfy some explanatory desiderata.[2]

Admittedly, generation is not inconsistent with an unconscious selection process. Scientists do not intend to compare all the logical possibilities. In fact, they normally take seriously only a few of them. This makes easier the subsequent comparison from the batch of alternatives. Since the demands of experimental testing concerning the amount of time, money, number of people involved, may be considerable, the resulting ranking gives a useful advice about how to proceed at this stage. More than thirty years ago Gilbert Harman coined the expression '*inference to the best explanation*' (IBE) to refer to this sort of non-deductive inference (Harman, 1965). It's worth emphasizing, then, that the goal of this process is to select that hypothesis that is the best *from an explanatory point of view.*

Having said this, it should be added that *IBE* may be understood in two different ways: either as a *heuristic procedure* or as an *epistemic rule*. According to the former, *IBE* equates to a search strategy for abductive solutions, that is, for *potential* explanations

---

[2] This "double-filter" model is defended in Lipton 2004, p. 148 and ff.

(Hanson, 1961). Selective search guided by explanatory considerations is preferred to blind search as a way of discovering the possibilities that would explain φ and fit with Θ in (1). In contrast, the epistemological interpretation of *IBE* intended by Harman does not only affirm that, when comparing and choosing among rival hypotheses, scientific reasoning is guided by explanatory criteria. It also defends the claim that this policy is truth-conducive: true or highly probable explanations are favoured while false, or low-probability ones, are discarded. The epistemological interpretation of *IBE* can be put schematically thus:

e is a collection of data, facts, observations, …

h explains e (if h were true, e would be a matter of course)

no other hypothesis explains e as well as h does

Therefore, h is true.

The dotted line represents a non-deductive connection between premises and conclusion. The normative principle is clear: among the array of hypotheses that account for the evidence we should *accept* as true the best of them *qua* explanation.[3] Simplicity, explanatory scope, fertility, precision, and compatibility with the background theories, are some of the properties usually invoked to this effect. Anyway, the crucial factor for acceptance is the explanatory goodness afforded by the hypotheses at issue.

3. A reliabilist justification for IBE?

As we remarked earlier, sometimes abductive inference does not leave room for any doubt except in a purely academic sense. A *reliabilist* justification for abduction perhaps could be grounded on these episodes. A few remarks will suffice to show how the reliabilist argument proceeds.

Not all of my perceptual judgements are true, but a great number of them are indeed true. Leaving aside exceptional circumstances, human perception is a reliable process insofar as the rate of true beliefs obtained through it is fairly high. Even though it can be false, any particular perceptual belief of mine is then automatically justified, according to a reliabilist standpoint, since it belongs to a class of beliefs that are mostly true (Goldman 1979). Now, think of those perceptual beliefs where the conclusion comes to us as a flash or, if you prefer less controversial examples, think about all those trivial facts whose explanation is obvious. The argument for *IBE* is analogous. Given that among the beliefs-explanations arrived at through *IBE* in everyday reasoning the rate of true beliefs is considerably high, *IBE* is a reliable process. Consequently, its output —belief in h or acceptance of h as true, when h is the best explanation— is justified.

Granted that ordinary perceptual beliefs are extreme cases of *IBE,* a reliabilist approach could work. In general, there are no special difficulties to assess the truth-value

---

[3] Weaker versions of *IBE* would conclude that h is highly probable, approximately true, closer to the truth than its competitors, …, but more on this below.

of everyday perceptual judgements. And the same is for the "obvious explanations". In both cases we have a wide basis of episodes to get a rough estimate. But things are different when dealing with theoretical scientific hypotheses because theoretical truth is much more elusive. Theoretical entities, mechanisms, processes, …, are devised to explain the empirical domain. Unfortunately, history of science is paved with discarded, and presumably false, theories. On account of this, straight extrapolation of truth-rate in everyday contexts to theoretical ones is, on my opinion, unacceptable. On the other side, the *IBE*-schema infers an absolute conclusion from a comparative premise. It is affirmed that the best explanation is true only because it is *better* than its rivals. But, wouldn't it be more appropriate to conclude that the best explanation is just closer to the truth —or that it has more verisimilitude— than the remaining alternatives? (see Kuipers 2004 and Niiniluoto 2005).

The normative principle 'accept the best explanation' puts the epistemological question —are explanatory criteria truth-conducive?— at the forefront. Advocates of the epistemic interpretation of *IBE* are bound to provide a rationale for their confidence on *IBE*'s normative import. And the foregoing comments about reliabilism suggest, in fact, that a sophisticated detour seems unavoidable to justify *IBE*'s, at least in respect of the scientific context.

4. *Explanatory value, probability and acceptance*

Niiniluoto has explored the link between explanatory goodness and *degree of confirmation*. From this point of view we are forced to translate the former notion into probabilistic terms. Niiniluoto thinks that Hempelian "systematic power" (*syst*) is a suitable translation: $syst(h, e) = p(\sim h/\sim e)$. If *syst* is used as a truth-dependent epistemic utility (Niiniluoto 1999, 187), maximization of *syst* recommends the acceptance of the hypothesis with the highest value for a well-known measure of degree of *incremental* confirmation: $conf(h, e) = p(h/e) - p(h)$ (Niiniluoto 2004, 71). Therefore, if the set of rival hypotheses is a partition, that is, if h_1, h_2, ..., h_n, are mutually exclusive and jointly exhaustive, and granted that h_1 is the best explanation in the partition, then:

(i) $conf(h_1, e) > conf(h_{i \neq 1}, e)$

that is, the best explanation is also the hypothesis that enjoys the highest degree of confirmation. Niiniluoto concludes that the Bayesian framework is appropriate to solve the "selection problem" of abduction: "These results establish *a probabilistic link between explanatory power and truth*: posterior probability $p(h/e)$ is the rational degree of belief in the truth of *h* on the basis of *e*, and thereby confirmation (as increase of probability by new evidence) means that we rationally become more certain of the truth of *h* than before" (Niiniluoto 2004, 76).

A straightforward result of Niiniluoto's approach is that *IBE* can only give us a comparative conclusion, i.e., that the best explanation is more confirmed than its competitors. However, at this point it is worth noticing that *IBE* affirms that explanatory considerations may be so overwhelming that the best explanation must be *accepted*.

And that means that we must *believe* that it is true.[4] However, no partisan of *IBE* would endorse the application of the rule 'accept the best explanation' with no further qualifications. For instance, if the best explanation were not, after all, good enough, the most reasonable option would be to suspend judgment and wait for new evidence. As P. Lipton has put it: "Inference to the Best Explanation might be more accurately if less memorably called 'Inference to the Best Explanation if the Best is Sufficiently Good'" (Lipton 2004, 154; see also chap. 4). Then, even though Niiniluto's proposal guarantees that the more (less) explanatory merit, the higher (lower) value for *conf*, acceptance requires an absolute threshold for $p(h/e)$.

Demanding that $p(h/e)$ is more probable than its negation —$p(h/e) > p(\sim h/e)$—, is, on my view, an appropriate constraint here. If the set of rival hypotheses is a partition, granted that h_1 is the best explanation, then $p(h_1/e) > p(h_2 \vee h_3 \vee ... \vee h_n/e)$. The rationale for this is that although $p(h_1/e)$ may be higher than each of the remaining conditional probabilities —$p(h_2/e),...,p(h_n/e)$—, if $p(h_1/e)$ is low, we should not accept h_1. In sum, acceptance of h_1 demands a further necessary condition:

(ii) $p(h_1/e) > \frac{1}{2}$

Then, provided that *syst* encapsulates the intuitive notion of explanatory goodness, (i) and (ii) establish, on my view, a probabilistic link between explanatory value and *acceptance*. Acceptance, on its turn, has full epistemic import, since it depends on probability and degree of confirmation, and not on pragmatic or instrumental considerations.

Nonetheless, it is controversial that (i) and (ii) offer an adequate interpretation of the normative principle embodied in *IBE* —'accept the best explanation!'. Some advocates of *IBE* object that "something would have gone amiss if we thought that the best explanation was *not* reasonably acceptable before it was subjected to Bayesian confirmation" (Psillos 2004, 89-90). According to Psillos, an essential aspect of *IBE* is that explanatory merit determines the probability of the hypothesis. Then, if accepting the best explanation entirely depends on confirmational support, *IBE* loses much of its excitement because "what is particularly challenging with *IBE* is the suggestion that the fact that a hypothesis is the *best* explanation (…) ipso facto warrants the judgement that it is likely" (Psillos 2002, 617). Consequently, it cannot occur that a good explanation is both unlikely and unacceptable.

Apparently, Psillos's objection misses the point. The possibility raised by him is ruled out by conditions (i) and (ii). Thus, according to the foregoing discussion higher initial plausibility —that is, higher prior probability— and higher explanatory likelihood together are *sufficient* for higher *conf* (Niiniluoto 2004, 71). Given that explanatory merit has to do both with prior probabilities and with likelihoods,[5] a good explanation

---

[4] Obviously, I do not subscribe van Fraassen's distinction between belief and acceptance (see his 1985, sec. 6).

[5] Prior probability can be seen as the translation of the initial plausibility of a hypothesis into the Bayesian framework. Taking for granted a definition of explanatory merit that includes theoretical elegance, simplicity..., it follows that the initial plausibility of hypotheses has to do with their explanatory merit. Thus, in principle, good explanations should enjoy higher prior probabilities than bad ones. But 'ex-

should enjoy a high value for *conf*. Besides, the best explanation is not only more probable than its rivals but probable enough to deserve acceptance, as (ii) states. Thus, correspondence between qualitative intuitive criteria of explanatory goodness and quantitative probability is guaranteed.

But Psillos's misgivings against a probabilistic account of *IBE* run deeper. He stresses that explanatory considerations determine which hypothesis is rational to believe *with no concern about its confirmational consequences*: "if a hypothesis has been chosen as the best explanation, then it has fared best in an explanatory-quality test with its competing rivals. So unless there is reason to think that it is superseded by an even better explanation, or unless there is reason to believe that the recalcitrant evidence points to one of the rivals as a better explanation, to stick with the best explanatory hypothesis is entirely reasonable." (Psillos 2002, 622). To illustrate his position Psillos discusses the base-rate fallacy. Let us suppose that: (i) h_1 is the best explanation; (ii) $p(e/h_1)$ is high; (iii) $p(h_1/e)$ is low —because $p(h_1)$ is still lower. Psillos asserts that it is not difficult to find examples where those three conditions are fulfilled and we should accept h_1, that is, a hypothesis that is very unlikely given e (see Psillos 2004, 85-87).

It could be pointed out that perhaps what we take to be the best explanation is not really the best. Psillos replies that "if the search for other potential explanations has been thorough, and if the present information does not justify a further exploration of the logical space of potentially explanatory hypotheses, there is no *specific* reason to doubt that the current best explanation is simply the best explanation" (Psillos 2002, 622). From this point of view, the acceptance of the best explanation mainly depends on the characteristics of the process whereby it has been selected, and not on its degree of confirmation. But, according to Psillos, the acceptance of the best explanation is *prima facie* reasonable, since it enhances the explanatory coherence of our total belief corpus.

This seems a radical departure from Niiniluoto's standpoint. For him, *IBE* is epistemically sound because it provides us with the most confirmed hypothesis. For Psillos, in contrast, favouring the best explanation equates to favouring the hypothesis that fits better with our background beliefs. Not surprisingly, Psillos asserts that good explanations give us explanatory coherence. But the epistemological challenge still remains, albeit in a reformulated way: why should we think that those explanations that afford explanatory coherence are true, or highly likely? We may agree with Psillos that the best explanation is the most *plausible* explanatory alternative. The epistemological question here is whether the plausible and the probable (or the true) go in the

---

planatory value' is an umbrella term that covers a multitude of different properties. Some of them are related to the initial plausibility of hypotheses, but some others are concerned with the appropriateness of the hypothesis *qua* explanation of the evidence at issue —notice that a general feature of good explanations is that they make the *explanandum* a "natural" consequence of the *explanans*. Then, since likelihoods express the probability of the evidence e in case the hypothesis is true, differences in explanatory goodness among rival hypotheses *in respect of the evidence e*, should somehow be reflected in their respective likelihoods —$p(e/h_1)$, $p(e/h_2)$, ..., $p(e/h_n)$. That is the reason why explanatory merit has to do with both prior probabilities and likelihoods.

same direction. I think that talking about *prima facie* justification is not very helpful here.

Let us take stock. In section 2 I pointed at two main interpretations of *IBE*: the heuristic —*IBE* as an alternative to facilitate search for explanations— and the normative —*IBE* as a truth-conducive rule. Two different accounts of the normative interpretation were discussed in section 4 —Niiniluoto's and Psillos's. They agree that favouring the best explanation is the best epistemic move. Niiniluoto tries to forge an analytic link between explanatory value and truth by means of confirmation. I suggested a way of closing the gap between confirmation and acceptance by adding a condition that alludes to an absolute threshold. Psillos, in contrast, insists that this interpretation sacrifices the most genuine claim of *IBE*.

I will not try to solve this disagreement, but some remarks are in order here. The normative interpretation of *IBE* is inspired by standard situations where "the abductive suggestion comes to us as a flash", in Peirce's words. Admittedly, these situations are very common in daily life. They even give some chances to a reliabilist justification of abduction, as we argued in section 2. But the situation may be completely different when we move to the "theory-choice" scenery. Empirical research on probabilistic reasoning has shown that we repeatedly violate the axioms of probability theory: the most plausible conclusion for the agent —the one which she infers— is not the most probable given the available information to her (Kahneman, Slovic & Tversky, 1982). It may be true that the most plausible conclusion is that one that stands out for its explanatory merits. But this leaves open the possibility that intuitions underlying everyday criteria of explanatory goodness collide with standard scientific comparison of rival hypotheses.

The rationality of human cognition is a big issue beyond the purview of this paper. Here it suffices for us to acknowledge that scientific inference is not just an extended application of common sense criteria. I do not claim that this is something new. Psillos tries to link explanatory merit to acceptance without intermediaries. My point is that the limits of his approach are, after all, the remarkable limits of everyday explanatory reasoning when applied to the scientific context.

5. Epilogue: consequences of Aliseda's theory of abduction on IBE

Aliseda distinguishes between the logical and the epistemological questions raised by abduction —generation and selection, respectively. Although she focuses on the first one, the "selection problem" is the central question in the contemporary debate on *IBE*. Following Peirce, Aliseda insists that economy and empirical testability are important constraints on abductive explanations. Acceptance of an explanatory hypothesis depends on the results of subsequent empirical tests. The abductive explanation is, thus, a suggestion more or less convincing "that has to be put to test before converting itself into a belief".[6] These views on the epistemological problem of abduction are hardly compatible to Psillos's standpoint on *IBE*. But they neither fit well with Ni-

---

[6] Aliseda 2006, 175. See also pages 35-7 and 170-77 for Peirce's views on abduction.

iniluoto's account. After all, Niiniluoto's vindication of *IBE* is supported by an analytic relation between explanatory goodness and confirmation. On account of this I conclude that Aliseda's theory of abduction should favour a purely heuristic reading of *IBE*.

REFERENCES

Aliseda, A. (2006). *Abductive Reasoning. Logical Investigations into Discovery and Explanation.* Berlin: Springer.

Goldman, A.I. (1979). "What Is Justified Belief?", in G.S. Pappas (ed.), *Justification and Knowledge.* Dordrecht: Reidel, pp. 1-23.

Hanson, N.R. (1961). "Is there a Logic of Discovery?", in H. Feigl and G. Maxwell (eds.), *Current Issues in the Philosophy of Science.* New York: Holt, Rinehart and Winston, pp. 20-35.

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

Hon, G., and S. Rakover (2001). *Explanation: Theoretical Approaches and Applications.* Dordrecht: Kluwer.

Kahneman, D., P. Slovic and B. Tversky (eds.) (1982). *Judgment under uncertainty: Heuristic and Biases.* Cambridge: Cambridge University Press.

Kuipers, Th. (2004). "Inference to the Best Theory rather than Inference to the Best Explanation", in F. Stadler (ed.), pp. 25-51.

Lipton, P. (2004, 2nd edition). *Inference to the Best Explanation.* London: Routledge.

Niiniluoto, I. (1999). *Critical Scientific Realism.* New York: Oxford University Press.

——— (2004). "Truth-seeking by abduction", in F. Stadler (ed.), pp. 57-82.

——— (2005). "Abduction and Truthlikeness", in R. Festa, A. Aliseda and J. Peijnenburg (eds.), *Confirmation, Empirical Progress and Truth Approximation.* Amsterdam: Rodopi, 255-75.

Psillos, S. (2002). "Simply the Best: A Case for Abduction", in F. Sadri, F. and A. Kakas (eds.), *Computational Logic: From Logic Programming into the Future*, LNAI 2408. Berlin: Springer, 605-25.

——— (2004). "Inference to the Best Explanation and Bayesianism", in F. Stadler (ed.), 83-91.

Stadler, F. (ed.) (2004). *Induction and Deduction in the Sciences.* Dordrecht: Kluwer.

van Fraassen, B. (1985). "Empiricism in the Philosophy of Science", in P.M. Churchland and C.A. Hooker (eds.), *Images of Science.* Chicago: The University of Chicago Press, 245-308.

Valeriano IRANZO is Senior Research Fellow (Programa Ramón y Cajal) at the department of Lógica y Filosofía de la Ciencia (Universitat de València, España). He has widely published on epistemology and philosophy of science, especially on theory of justification and scientific realism. His current philosophical interests are confirmation theory, foundations of probability and statistics, and the "theory-choice" problem.

ADDRESS: Departamento de Lógica y Filosofía de la Ciencia, Facultad de Filosofía y CC. Educación, Blasco Ibáñez 30, 46010 Valencia, España. E-mail: valeriano.iranzo@uv.es.