

## Abduction

*First published Wed Mar 9, 2011; substantive revision Tue May 18, 2021*

In the philosophical literature, the term “abduction” is used in two related but different senses. In both senses, the term refers to some form of explanatory reasoning. However, in the historically first sense, it refers to the place of explanatory reasoning in *generating hypotheses*, while in the sense in which it is used most frequently in the modern literature it refers to the place of explanatory reasoning in *justifying hypotheses*. In the latter sense, abduction is also often called “Inference to the Best Explanation.”

This entry is exclusively concerned with abduction in the modern sense, although there is a supplement on abduction in the historical sense, which had its origin in the work of Charles Sanders Peirce—see the

[Supplement: Peirce on Abduction](#).

See also the entry on [scientific discovery](#), in particular the section on discovery as abduction.

Most philosophers agree that abduction (in the sense of Inference to the Best Explanation) is a type of inference that is frequently employed, in some form or other, both in everyday and in scientific reasoning. However, the exact form as well as the normative status of abduction are still matters of controversy. This entry contrasts abduction with other types of inference; points at prominent uses of it, both in and outside philosophy; considers various more or less precise statements of it; discusses its normative status; and highlights possible connections between abduction and Bayesian confirmation theory.

- [1. Abduction: The General Idea](#)
  - [1.1 Deduction, induction, abduction](#)
  - [1.2 The ubiquity of abduction](#)
- [2. Explicating Abduction](#)
- [3. The Status of Abduction](#)
  - [3.1 Criticisms](#)
  - [3.2 Defenses](#)
- [4. Abduction versus Bayesian Confirmation Theory](#)
- [Bibliography](#)
- [Academic Tools](#)
- [Other Internet Resources](#)
- [Related Entries](#)

---

### 1. Abduction: The General Idea

You happen to know that Tim and Harry have recently had a terrible row that ended their friendship. Now someone tells you that she just saw Tim and Harry jogging together. The best explanation for this that you can think of is that they made up. You conclude that they are friends again.

One morning you enter the kitchen to find a plate and cup on the table, with breadcrumbs and a pat of butter on it, and surrounded by a jar of jam, a pack of sugar, and an empty carton of milk. You conclude that one of your

house-mates got up at night to make him- or herself a midnight snack and was too tired to clear the table. This, you think, best explains the scene you are facing. To be sure, it might be that someone burgled the house and took the time to have a bite while on the job, or a house-mate might have arranged the things on the table without having a midnight snack but just to make you believe that someone had a midnight snack. But these hypotheses strike you as providing much more contrived explanations of the data than the one you infer to.

Walking along the beach, you see what looks like a picture of Winston Churchill in the sand. It could be that, as in the opening pages of Hilary Putnam's book *Reason, Truth, and History*, (1981), what you see is actually the trace of an ant crawling on the beach. The much simpler, and therefore (you think) much better, explanation is that someone intentionally drew a picture of Churchill in the sand. That, in any case, is what you come away believing.

In these examples, the conclusions do not follow logically from the premises. For instance, it does not follow logically that Tim and Harry are friends again from the premises that they had a terrible row which ended their friendship and that they have just been seen jogging together; it does not even follow, we may suppose, from all the information you have about Tim and Harry. Nor do you have any useful statistical data about friendships, terrible rows, and joggers that might warrant an inference from the information that you have about Tim and Harry to the conclusion that they are friends again, or even to the conclusion that, probably (or with a certain probability), they are friends again. What leads you to the conclusion, and what according to a considerable number of philosophers may also warrant this conclusion, is precisely the fact that Tim and Harry's being friends again would, if true, *best explain* the fact that they have just been seen jogging together. (The proviso that a hypothesis be true if it is to explain anything is taken as read from here on.) Similar remarks apply to the other two examples. The type of inference exhibited here is called *abduction* or, somewhat more commonly nowadays, *Inference to the Best Explanation*.

## 1.1 Deduction, induction, abduction

Abduction is normally thought of as being **one of three major types of inference**, the other two being deduction and induction. The distinction between deduction, on the one hand, and induction and abduction, on the other hand, corresponds to the distinction between **necessary** and **non-necessary** inferences. In deductive inferences, what is inferred is *necessarily* true if the premises from which it is inferred are true; that is, the truth of the premises *guarantees* the truth of the conclusion. A familiar type of example is inferences instantiating the schema

All *As* are *Bs*.  
*a* is an *A*.  
Hence, *a* is a *B*.

But not all inferences are of this variety. Consider, for instance, the inference of "John is rich" from "John lives in Chelsea" and "Most people living in Chelsea are rich." Here, the truth of the first sentence is not guaranteed (but only made likely) by the joint truth of the second and third sentences. Differently put, it is not necessarily the case that if the premises are true, then so is the conclusion: it is logically compatible with the truth of the premises that John is a member of the minority of non-rich inhabitants of Chelsea. The case is similar regarding your inference to the conclusion that Tim and Harry are friends again on the basis of the information that they have been seen jogging together. Perhaps Tim and Harry are former business partners who still had some financial matters to discuss, however much they would have liked to avoid this, and decided to combine this with their daily exercise; this is compatible with their being firmly decided never to make up.

It is standard practice to group non-necessary inferences into *inductive* and *abductive* ones. Inductive inferences form a somewhat **heterogeneous class**, but for present purposes they may be characterized as those inferences that are based purely on **statistical data**, such as observed **frequencies** of occurrences of a particular feature in a given population. An example of such an inference would be this:

96 per cent of the Flemish college students speak both Dutch and French.

Louise is a Flemish college student.

Hence, Louise speaks both Dutch and French.

However, the relevant statistical information may also be more vaguely given, as in the premise, “Most people living in Chelsea are rich.” (There is much discussion about whether the conclusion of an inductive argument can be stated in purely qualitative terms or whether it should be a quantitative one—for instance, that it holds with a probability of .96 that Louise speaks both Dutch and French—or whether it can *sometimes* be stated in qualitative terms—for instance, if the probability that it is true is high enough—and sometimes not. On these and other issues related to induction, see Kyburg 1990 (Ch. 4). It should also be mentioned that Harman (1965) conceives induction as a special type of abduction. See also Weintraub 2013 for discussion.)

The mere fact that an inference is based on statistical data is not enough to classify it as an inductive one. You may have observed many gray elephants and no non-gray ones, and infer from this that all elephants are gray, *because that would provide the best explanation for why you have observed so many gray elephants and no non-gray ones*. This would be an instance of an abductive inference. It suggests that the best way to distinguish between induction and abduction is this: both are *ampliative*, meaning that the conclusion goes beyond what is (logically) contained in the premises (which is why they are non-necessary inferences), but in abduction there is an implicit or explicit appeal to explanatory considerations, whereas in induction there is not; in induction, there is *only* an appeal to observed frequencies or statistics. (I emphasize “only,” because in abduction there may also be an appeal to frequencies or statistics, as the example about the elephants exhibits.)

A noteworthy feature of abduction, which it shares with induction but not with deduction, is that it violates *monotonicity*, meaning that it may be possible to infer abductively certain conclusions from a subset of a set  $S$  of premises which cannot be inferred abductively from  $S$  as a whole. For instance, adding the premise that Tim and Harry are former business partners who still have some financial matters to discuss, to the premises that they had a terrible row some time ago and that they were just seen jogging together may no longer warrant you to infer that they are friends again, even if—let us suppose—the last two premises alone do warrant that inference. The reason is that what counts as the best explanation of Tim and Harry’s jogging together in light of the original premises may no longer do so once the information has been added that they are former business partners with financial matters to discuss.

## 1.2 The ubiquity of abduction

The type of inference exemplified in the cases described at the beginning of this entry will strike most as entirely familiar. Philosophers as well as psychologists tend to agree that abduction is frequently employed in everyday reasoning. Sometimes our reliance on abductive reasoning is quite obvious and explicit. But in some daily practices, it may be so routine and automatic that it easily goes unnoticed. A case in point may be our trust in other people’s testimony, which has been said to rest on abductive reasoning; see Harman 1965, Adler 1994, Fricker 1994, and Lipton 1998 for defenses of this claim. For instance, according to Jonathan Adler (1994, 274f), “[t]he best explanation for why the informant asserts that  $P$  is normally that ... he believes it for duly responsible reasons and ... he intends that I shall believe it too,” which is why we are normally justified in trusting the informant’s testimony. This may well be correct, even though in coming to trust a person’s testimony one does not normally seem to be aware of any abductive reasoning going on in one’s mind. Similar remarks may apply to what some hold to be a further, possibly even more fundamental, role of abduction in linguistic practice, to wit, its role in determining what a speaker means by an utterance. Specifically, it has been argued that decoding utterances is a matter of inferring the best explanation of why someone said what he or she said in the context in which the utterance was made. Even more specifically, authors working in the field of pragmatics have suggested that hearers invoke the Gricean maxims of conversation to help them work out the best explanation of a speaker’s utterance whenever the semantic content of the utterance is insufficiently informative for the purposes of the conversation, or is too informative, or off-topic, or implausible, or otherwise odd or inappropriate; see, for instance, Bach and Harnish 1979 (92f), Dascal 1979 (167), and Hobbs 2004. As in cases of reliance on speaker testimony, the requisite abductive reasoning would normally seem to take place at a subconscious level.

Abductive reasoning is not limited to everyday contexts. Quite the contrary: philosophers of science have argued that abduction is a cornerstone of scientific methodology; see, for instance, Boyd 1981, 1984, Harré 1986, 1988, Lipton 1991, 2004, and Psillos 1999. According to Timothy Williamson (2007), “[t]he abductive methodology is the best science provides” and Ernan McMullin (1992) even goes so far to call abduction “the inference that makes science.” To illustrate the use of abduction in science, we consider two examples.

At the beginning of the nineteenth century, it was discovered that the orbit of Uranus, one of the seven planets known at the time, departed from the orbit as predicted on the basis of Isaac Newton’s theory of universal gravitation and the auxiliary assumption that there were no further planets in the solar system. One possible explanation was, of course, that Newton’s theory is false. Given its great empirical successes for (then) more than two centuries, that did not appear to be a very good explanation. Two astronomers, John Couch Adams and Urbain Leverrier, instead suggested (independently of each other but almost simultaneously) that there was an eighth, as yet undiscovered planet in the solar system; that, they thought, provided the best explanation of Uranus’ deviating orbit. Not much later, this planet, which is now known as “Neptune,” was discovered.

The second example concerns what is now commonly regarded to have been the discovery of the electron by the English physicist Joseph John Thomson. Thomson had conducted experiments on cathode rays in order to determine whether they are streams of charged particles. He concluded that they are indeed, reasoning as follows:

As the cathode rays carry a charge of negative electricity, are deflected by an electrostatic force as if they were negatively electrified, and are acted on by a magnetic force in just the way in which this force would act on a negatively electrified body moving along the path of these rays, I can see no escape from the conclusion that they are charges of negative electricity carried by particles of matter. (Thomson, cited in Achinstein 2001, 17)

The conclusion that cathode rays consist of negatively charged particles does not follow logically from the reported experimental results, nor could Thomson draw on any relevant statistical data. That nevertheless he could “see no escape from the conclusion” is, we may safely assume, because the conclusion is the best—in this case presumably even the only plausible—explanation of his results that he could think of.

Many other examples of scientific uses of abduction have been discussed in the literature; see, for instance, Harré 1986, 1988 and Lipton 1991, 2004. Abduction is also said to be the predominant mode of reasoning in medical diagnosis: physicians tend to go for the hypothesis that best explains the patient’s symptoms (see Josephson and Josephson (eds.) 1994, 9–12; see also Dragulinescu 2016 on abductive reasoning in the context of medicine).

Last but not least, abduction plays a central role in some important philosophical debates. See Shalkowski 2010 on the place of abduction in metaphysics (also Bigelow 2010), Krzyżanowska, Wenmackers, and Douven 2014 and Douven 2016a for a possible role of abduction in the semantics of conditionals, and Williamson 2017 for an application of abduction in the philosophy of logic. Arguably, however, abduction plays its most notable philosophical role in epistemology and in the philosophy of science, where it is frequently invoked in objections to so-called underdetermination arguments. Underdetermination arguments generally start from the premise that a number of given hypotheses are empirically equivalent, which their authors take to mean that the evidence—indeed, any evidence we might ever come to possess—is unable to favor one of them over the others. From this, we are supposed to conclude that one can never be warranted in believing any particular one of the hypotheses. (This is rough, but it will do for present purposes; see Douven 2008 and Stanford 2009, for more detailed accounts of underdetermination arguments.) A famous instance of this type of argument is the Cartesian argument for global skepticism, according to which the hypothesis that reality is more or less the way we customarily deem it to be is empirically equivalent to a variety of so-called skeptical hypotheses (such as that we are beguiled by an evil demon, or that we are brains in a vat, connected to a supercomputer; see, e.g., Folina 2016). Similar arguments have been given in support of scientific antirealism, according to which it will never be warranted for us to choose between empirically equivalent rivals concerning what underlies the observable part of reality (van Fraassen 1980).

Responses to these arguments typically point to the fact that the notion of empirical equivalence at play unduly neglects explanatory considerations, for instance, by defining the notion strictly in terms of hypotheses' making the same predictions. Those responding then argue that even if some hypotheses make exactly the same predictions, one of them may still be a better explanation of the phenomena predicted. Thus, if explanatory considerations have a role in determining which inferences we are licensed to make—as according to defenders of abduction they have—then we might still be warranted in believing in the truth (or probable truth, or some such, depending—as will be seen below—on the version of abduction one assumes) of one of a number of hypotheses that all make the same predictions. Following Bertrand Russell (1912, Ch. 2), many epistemologists have invoked abduction in arguing against Cartesian skepticism, their key claim being that even though, by construction, the skeptical hypotheses make the same predictions as the hypothesis that reality is more or less the way we ordinarily take it to be, they are not equally good explanations of what they predict; in particular, the skeptical hypotheses have been said to be considerably less simple than the “ordinary world” hypothesis. See, among many others, Harman 1973 (Chs. 8 and 11), Goldman 1988 (205), Moser 1989 (161), and Vogel 1990, 2005; see Pargetter 1984 for an abductive response specifically to skepticism regarding other minds. Similarly, philosophers of science have argued that we are warranted to believe in Special Relativity Theory as opposed to Lorentz's version of the æther theory. For even though these theories make the same predictions, the former is explanatorily superior to the latter. (Most arguments that have been given for this claim come down to the contention that Special Relativity Theory is ontologically more parsimonious than its competitor, which postulates the existence of an æther. See Janssen 2002 for an excellent discussion of the various reasons philosophers of science have adduced for preferring Einstein's theory to Lorentz's.)

## 2. Explicating Abduction

Precise statements of what abduction amounts to are rare in the literature on abduction. (Peirce did propose an at least fairly precise statement; but, as explained in the supplement to this entry, it does not capture what most nowadays understand by abduction.) Its core idea is often said to be that **explanatory considerations have confirmation-theoretic import**, or that **explanatory success** is a (not necessarily unfailing) **mark of truth**. Clearly, however, these formulations are **slogans at best**, and it takes little effort to see that they can be cashed out in a great variety of prima facie plausible ways. Here we will consider a number of such possible explications, starting with what one might term the **“textbook version of abduction,”** which, as will be seen, is manifestly **defective**, and then going on to consider various possible refinements of it. What those versions have in common—unsurprisingly—is that they are all inference rules, requiring premises encompassing explanatory considerations and yielding a conclusion that makes some statement about the truth of a hypothesis. The differences concern the premises that are required, or what exactly we are allowed to infer from them (or both).

In textbooks on epistemology or the philosophy of science, one often encounters something like the following as a formulation of abduction:

ABD1     **Given evidence  $E$  and candidate explanations  $H_1, \dots, H_n$  of  $E$ ,** infer the truth of *that*  $H_i$  which best explains  $E$ .

An observation that is frequently made about this rule, and that points to a potential problem for it, is that **it presupposes the notions of candidate explanation** and best explanation, neither of which has a straightforward interpretation. While some still hope that the former can be spelled out in purely logical, or at least purely formal, terms, it is often said that the latter must appeal to the so-called theoretical virtues, like simplicity, generality, and coherence with well-established theories; the best explanation would then be the hypothesis which, on balance, does best with respect to these virtues. (See, for instance, Thagard 1978 and McMullin 1996.) The problem is that none of the said virtues is presently particularly well understood. (Giere, in Callebaut (ed.) 1993 (232), even makes the radical claim that the theoretical virtues lack real content and play no more than a rhetorical role in science. In view of recent formal work both on simplicity and on coherence—for instance, Forster and Sober 1994, Li and Vitanyi 1997, and Sober 2015, on simplicity and Bovens and Hartmann 2003 and Olsson 2005, on coherence—the first part of this claim has become hard to maintain; also, Schupbach and Sprenger (2011) present an account of explanatory goodness directly in probabilistic terms. Psychological

evidence casts doubt on the second part of the claim; see, for instance, Lombrozo 2007, on the role of simplicity in people's assessments of explanatory goodness and Koslowski *et al.* 2008, on the role of coherence with background knowledge in those assessments.)

Furthermore, many of those who think ABD1 is headed along the right lines believe that it is too strong. Some think that abduction warrants an inference only to **the probable truth** of the best explanation, others that it warrants an inference only to **the approximate truth** of the best explanation, and still others that it warrants an inference only to the *probable approximate* truth.

The real problem with ABD1 runs deeper than this, however. Because abduction is ampliative—as explained earlier—it will **not be a sound rule of inference in the strict logical sense**, however abduction is explicated exactly. It can still be *reliable* in that it mostly leads to a true conclusion whenever the premises are true. An obvious necessary condition for ABD1 to be reliable in this sense is that, *mostly*, when it is true that  $H$  best explains  $E$ , and  $E$  is true, then  $H$  is true as well (or  $H$  is approximately true, or probably true, or probably approximately true). But this would not be *enough* for ABD1 to be reliable. For ABD1 takes as its premise only that some hypothesis is **the best explanation** of the evidence *as compared to other hypotheses in a given set*. Thus, if the rule is to be reliable, it must hold that, at least typically, the best explanation relative to the set of hypotheses that we consider would also come out as being best **in comparison with any other hypotheses** that we might have conceived (but for lack of time or ingenuity, or for some other reason, did not conceive). In other words, it must hold that at least typically the *absolutely* best explanation of the evidence is to be found among the candidate explanations we have come up with, for else ABD1 may well lead us to believe “the best of a bad lot” (van Fraassen 1989, 143).

How reasonable is it to suppose that this extra requirement is usually fulfilled? Not at all, presumably. To believe otherwise, we must assume some sort of **privilege on our part** to the effect that when we consider possible explanations of the data, we are somehow predisposed to hit, **inter alia**, upon the absolutely best explanation of those data. After all, hardly ever will we have considered, or will it even be possible to consider, *all* potential explanations. As van Fraassen (1989, 144) points out, it is *a priori* rather implausible to hold that we are thus privileged.

In response to this, one might argue that the challenge to show that the best explanation is always or mostly among the hypotheses considered can be met without having to assume some form of privilege (see Schupbach 2014 for a different response, and see Dellsén 2017 for discussion). For given the hypotheses we have managed to come up with, we can always generate a set of hypotheses which jointly exhaust logical space. Suppose  $H_1, \dots, H_n$  are the candidate explanations we have so far been able to conceive. Then simply define  $H_{n+1} := \neg H_1 \wedge \dots \wedge \neg H_n$  and add this new hypothesis as a further candidate explanation to the ones we already have. Obviously, the set  $\{H_1, \dots, H_{n+1}\}$  is exhaustive, in that one of its elements must be true. Following this in itself simple procedure would seem enough to make sure that we never miss out on the absolutely best explanation. (See Lipton 1993, for a proposal along these lines.)

Alas, there is a catch. For even though there may be many hypotheses  $H_j$  that imply  $H_{n+1}$  and, had they been formulated, would have been evaluated as being a better explanation for the data than the best explanation among the candidate explanations we started out with,  $H_{n+1}$  itself will in general be hardly informative; in fact, in general **it will not even be clear what its empirical consequences are**. Suppose, for instance, we have as competing explanations Special Relativity Theory and Lorentz's version of the æther theory. Then, following the above proposal, we may add to our candidate explanations that neither of these two theories is true. But surely this further hypothesis will be ranked quite low *qua* explanation—if it will be ranked at all, which seems doubtful, given that it is wholly unclear what its empirical consequences are. This is not to say that the suggested procedure may never work. The point is that in general it will give little assurance that the best explanation is among the candidate explanations we consider.

A more promising response to the above “argument of the bad lot” begins with the observation that the argument capitalizes on a peculiar asymmetry or incongruence in ABD1. The rule gives license to an absolute conclusion—that a given hypothesis is true—on the basis of a comparative premise, namely, that that particular hypothesis

is the best explanation of the evidence relative to the other hypotheses available (see Kuipers 2000, 171). This incongruence is not avoided by replacing “truth” with “probable truth” or “approximate truth.” In order to avoid it, one has two general options.

The first option is to **modify the rule** so as to **have it require an absolute premise**. For instance, following Alan Musgrave (1988) or Peter Lipton (1993), one may require the hypothesis whose truth is inferred to be not only the best of the available potential explanations, but also to be **satisfactory (Musgrave) or good enough (Lipton)**, yielding the following variant of ABD1:

ABD2      Given evidence  $E$  and candidate explanations  $H_1, \dots, H_n$  of  $E$ , infer the truth of *that*  $H_i$  which explains  $E$  best, provided  $H_i$  **is satisfactory/good enough qua explanation.**

Needless to say, ABD2 needs supplementing by a criterion for the satisfactoriness of explanations, or their being good enough, which, however, we are still lacking.

Secondly, one can formulate a symmetric or congruous version of abduction by having it sanction, given a comparative premise, **only a comparative conclusion**; this option, too, can in turn be realized in more than one way. Here is one way to do it, which has been proposed and defended in the work of Theo Kuipers (e.g., Kuipers 1984, 1992, 2000).

ABD3      Given evidence  $E$  and candidate explanations  $H_1, \dots, H_n$  of  $E$ , if  $H_i$  explains  $E$  better than any of the other hypotheses, infer that  $H_i$  **is closer to the truth than any of the other hypotheses.**

Clearly, ABD3 requires an account of closeness to the truth, but many such accounts are on offer today (see, e.g., Niiniluoto 1998).

One noteworthy feature of the congruous versions of abduction considered here is that they do not rely on the assumption of an implausible privilege on the reasoner's part that, we saw, ABD1 implicitly relies on. Another is that if one can be certain that, however many candidate explanations for the data one may have missed, none equals the best of those one *has* thought of, then the congruous versions license exactly the same inference as ABD1 does (supposing that one would not be certain that no potential explanation is as good as the best explanation one has thought of if the latter is not even satisfactory or sufficiently good).

As mentioned, there is widespread agreement that people frequently rely on abductive reasoning. Which of the above rules *exactly* is it that people rely on? Or might it be still some further rule that they rely on? Or might they in some contexts rely on one version, and in others on another (Douven 2017, forthcoming)? Philosophical argumentation is unable to answer these questions. In recent years, experimental psychologists have started paying attention to the role humans give to explanatory considerations in reasoning. For instance, Tania Lombrozo and Nicholas Gwynne (2014) report experiments showing that *how* a property of a given class of things is explained to us—whether mechanistically, by reference to parts and processes, or functionally, by reference to functions and purposes—matters to how likely we are to generalise that property to other classes of things (see also Sloman 1994 and Williams and Lombrozo 2010). And Igor Douven and Jonah Schupbach (2015a), (2015b) present experimental evidence to the effect that people's probability updates tend to be influenced by explanatory considerations in ways that makes them deviate from strictly Bayesian updates (see below). Douven (2016b) shows that, in the aforementioned experiments, participants who gave more weight to explanatory considerations tended to be more accurate, as determined in terms of a standard scoring rule. (See Lombrozo 2012 and 2016 for useful overviews of recent experimental work relevant to explanation and inference.) Douven and Patricia Mirabile (2018) found some evidence indicating that people rely on something like ABD2, at least in some contexts, but for the most part, empirical work on the above-mentioned questions is lacking.

With respect to the normative question of which of the previously stated rules we *ought* to rely on (**if we ought to rely on any form of abduction**), where philosophical argumentation should be able to help, the situation is hardly any better. In view of the argument of the bad lot, ABD1 does not look very good. Other arguments

against abduction are claimed to be independent of the exact explication of the rule; below, these arguments will be found wanting. On the other hand, arguments that have been given in favor of abduction—some of which will also be discussed below—do not discern between specific versions. So, supposing people do indeed commonly rely on abduction, it must be considered an open question as to which version(s) of abduction they rely on. Equally, supposing it is rational for people to rely on abduction, it must be considered an open question as to which version, or perhaps versions, of abduction they ought to, or are at least permitted to, rely on.

### 3. The Status of Abduction

Even if it is true that we routinely rely on abductive reasoning, it may still be asked **whether this practice is rational**. For instance, experimental studies have shown that when people are able to think of an explanation for some possible event, they tend to overestimate the likelihood that this event will actually occur. (See Koehler 1991, for a survey of some of these studies; see also Brem and Rips 2000.) More telling still, Lombrozo (2007) shows that, in some situations, people tend to grossly overrate the probability of simpler explanations compared to more complicated ones. Although these studies are not directly concerned with abduction in any of the forms discussed so far, they nevertheless suggest that taking into account explanatory considerations in one's reasoning may not always be for the better. (It is to be noted that Lombrozo's experiments *are* directly concerned with some proposals that have been made for explicating abduction in a Bayesian framework; see Section 4.) However, the most pertinent remarks about the normative status of abduction are so far to be found in the philosophical literature. This section discusses the main criticisms that have been levelled against abduction, as well as the strongest arguments that have been given in its defense.

#### 3.1 Criticisms

We have already encountered the so-called argument of **the bad lot**, which, we saw, is valid as a criticism of ABD1 but powerless against various (what we called) congruous rules of abduction. We here consider two objections that are meant to be more general. The first even purports to challenge the core idea underlying abduction; the second is not quite as general, but it is still meant to undermine a broad class of candidate explications of abduction. Both objections are due to Bas van Fraassen.

The first objection has as a premise that it is part of the meaning of “explanation” that if one theory is more explanatory than another, the former must be more informative than the latter (see, e.g., van Fraassen 1983, Sect. 2). The alleged problem then is that it is “an elementary logical point that a more informative theory cannot be more likely to be true [and thus] attempts to describe inductive or evidential support through features that require information (such as ‘Inference to the Best Explanation’) must either contradict themselves or equivocate” (van Fraassen 1989, 192). The elementary logical point is supposed to be “most [obvious] ... in the paradigm case in which one theory is an extension of another: clearly the extension has more ways of being false” (van Fraassen 1985, 280).

It is important to note, however, that in any other kind of case than the “paradigm” one, the putative elementary point is not obvious at all. For instance, it is entirely unclear in what sense Special Relativity Theory “has more ways of being false” than Lorentz's version of the æther theory, given that they make the same predictions. And yet the former is generally regarded as being superior, *qua* explanation, to the latter. (If van Fraassen were to object that the former is not really more informative than the latter, or at any rate not more informative in the appropriate sense—whatever that is—then we should certainly refuse to grant the premise that in order to be more explanatory a theory must be more informative.)

The second objection, proffered in van Fraassen 1989 (Ch. 6), is levelled at probabilistic versions of abduction. The objection is that such rules must either amount to Bayes' rule, and thus be redundant, or be at variance with it but then, on the grounds of Lewis' dynamic Dutch book argument (as reported in Teller 1973), be probabilistically incoherent, meaning that they may lead one to assess as fair a number of bets which together ensure a financial loss, come what may; and, van Fraassen argues, it would be irrational to follow a rule that has this feature.

However, this objection fares no better than the first. For one thing, as Patrick Maher (1992) and Brian Skyrms (1993) have pointed out, a loss in one respect may be outweighed by a benefit in another. It might be, for instance, that some probabilistic version of abduction does much better, at least in our world, than Bayes' rule, in that, on average, it approaches the truth faster in the sense that it is faster in assigning a high probability (understood as probability above a certain threshold value) to the true hypothesis (see Douven 2013, 2020, and Douven and Wenmackers 2017; see Climenhaga 2017 for discussion). If it does, then following that rule instead of Bayes' rule may have advantages which perhaps are not so readily expressed in terms of money yet which should arguably be taken into account when deciding which rule to go by. It is, in short, not so clear whether following a probabilistically incoherent rule must be irrational.

For another thing, Douven (1999) argues that the question of whether a probabilistic rule is coherent is not one that can be settled independently of considering which other epistemic and decision-theoretic rules are deployed along with it; coherence should be understood as a property of packages of both epistemic and decision-theoretic rules, not of epistemic rules (such as probabilistic rules for belief change) in isolation. In the same paper, a coherent package of rules is described which includes a probabilistic version of abduction. (See Kvanvig 1994, Harman 1997, Leplin 1997, Niiniluoto 1999, and Okasha 2000, for different responses to van Fraassen's critique of probabilistic versions of abduction.)

### 3.2 Defenses

Hardly anyone nowadays would want to subscribe to a conception of truth that posits a necessary connection between explanatory force and truth—for instance, because it stipulates explanatory superiority to be necessary for truth. As a result, a priori defenses of abduction seem out of the question. Indeed, all defenses that have been given so far are of an empirical nature in that they appeal to data that supposedly support the claim that (in some form) abduction is a reliable rule of inference.

The best-known argument of this sort was developed by Richard Boyd in the 1980s (see Boyd 1981, 1984, 1985). It starts by underlining the theory-dependency of scientific methodology, which comprises methods for designing experiments, for assessing data, for choosing between rival hypotheses, and so on. For instance, in considering possible confounding factors from which an experimental setup has to be shielded, scientists draw heavily on already accepted theories. The argument next calls attention to the apparent reliability of this methodology, which, after all, has yielded, and continues to yield, impressively accurate theories. In particular, by relying on this methodology, scientists have for some time now been able to find ever more instrumentally adequate theories. Boyd then argues that the reliability of scientific methodology is best explained by assuming that the theories on which it relies are at least approximately true. From this and from the fact that these theories were mostly arrived at by abductive reasoning, he concludes that abduction must be a reliable rule of inference.

Critics have accused this argument of being circular. Specifically, it has been said that the argument rests on a premise—that scientific methodology is informed by approximately true background theories—which in turn rests on an inference to the best explanation for its plausibility. And the reliability of this type of inference is precisely what is at stake. (See, for instance, Laudan 1981 and Fine 1984.)

To this, Stathis Psillos (1999, Ch. 4) has responded by invoking a distinction credited to Richard Braithwaite, to wit, the distinction between premise-circularity and rule-circularity. An argument is premise-circular if its conclusion is amongst its premises. A rule-circular argument, by contrast, is an argument of which the conclusion asserts something about an inferential rule that is used in the very same argument. As Psillos urges, Boyd's argument is rule-circular, but not premise-circular, and rule-circular arguments, Psillos contends, *need not* be viciously circular (even though a premise-circular argument is always viciously circular). To be more precise, in his view, an argument for the reliability of a given rule *R* that essentially relies on *R* as an inferential principle is not vicious, provided that the use of *R* does not guarantee a positive conclusion about *R*'s reliability. Psillos claims that in Boyd's argument, this proviso is met. For while Boyd concludes that the background theories on which scientific methodology relies are approximately true on the basis of an abductive step, the use of abduction itself does not guarantee the truth of his conclusion. After all, granting the use of abduction does

nothing to ensure that the best explanation of the success of scientific methodology is the approximate truth of the relevant background theories. Thus, Psillos concludes, Boyd's argument still stands.

Even if the use of abduction in Boyd's argument might have led to the conclusion that abduction is *not* reliable, one may still have worries about the argument's being rule-circular. For suppose that some scientific community relied not on abduction but on a rule that we may dub "Inference to the Worst Explanation" (IWE), a rule that sanctions inferring to the *worst* explanation of the available data. We may safely assume that the use of this rule mostly would lead to the adoption of very unsuccessful theories. Nevertheless, the said community might justify its use of IWE by dint of the following reasoning: "Scientific theories tend to be hugely unsuccessful. These theories were arrived at by application of IWE. That IWE is a reliable rule of inference—that is, a rule of inference mostly leading from true premises to true conclusions—is surely the worst explanation of the fact that our theories are so unsuccessful. Hence, by application of IWE, we may conclude that IWE is a reliable rule of inference." While this would be an utterly absurd conclusion, the argument leading up to it cannot be convicted of being viciously circular anymore than Boyd's argument for the reliability of abduction can (if Psillos is right). It would appear, then, that there must be something else amiss with rule-circularity.

It is fair to note that for Psillos, the fact that a rule-circular argument does not guarantee a positive conclusion about the rule at issue is not sufficient for such an argument to be valid. A further necessary condition is "that one should not have reason to doubt the reliability of the rule—that there is nothing currently available which can make one distrust the rule" (Psillos 1999, 85). And there is plenty of reason to doubt the reliability of IWE; in fact, the above argument *supposes* that it is unreliable. Two questions arise, however. First, why should we accept the additional condition? Second, do we really have *no* reason to doubt the reliability of abduction? Certainly *some* of the abductive inferences we make lead us to accept *falsehoods*. How many falsehoods may we accept on the basis of abduction before we can legitimately begin to distrust this rule? No clear answers have been given to these questions.

Be this as it may, even if rule-circularity is neither vicious nor otherwise problematic, one may still wonder how Boyd's argument is to convert a critic of abduction, given that it relies on abduction. But Psillos makes it clear that the point of philosophical argumentation is not always, and in any case need not be, to convince an opponent of one's position. Sometimes the point is, more modestly, to assure or reassure oneself that the position one endorses, or is tempted to endorse, is correct. In the case at hand, we need not think of Boyd's argument as an attempt to convince the opponent of abduction of its reliability. Rather, it may be thought of as justifying the rule from within the perspective of someone who is already sympathetic towards abduction; see Psillos 1999 (89).

There have also been attempts to argue for abduction in a more straightforward fashion, to wit, via enumerative induction. The common idea of these attempts is that every newly recorded successful application of abduction—like the discovery of Neptune, whose existence had been postulated on explanatory grounds (see Section 1.2)—adds further support to the hypothesis that abduction is a reliable rule of inference, in the way in which every newly observed black raven adds some support to the hypothesis that all ravens are black. Because it does not involve abductive reasoning, this type of argument is more likely to also appeal to disbelievers in abduction. See Harré 1986, 1988, Bird 1998 (160), Kitcher 2001, and Douven 2002 for suggestions along these lines.

## 4. Abduction versus Bayesian Confirmation Theory

In the past decade, Bayesian confirmation theory has firmly established itself as the dominant view on **confirmation**; currently one cannot very well discuss a confirmation-theoretic issue without making clear whether, and if so why, **one's position on that issue deviates from standard Bayesian thinking**. Abduction, in whichever version, assigns a confirmation-theoretic role to **explanation**: **explanatory considerations contribute to making some hypotheses more credible, and others less so**. By contrast, Bayesian confirmation theory makes no reference at all to the concept of explanation. Does this imply that abduction is at loggerheads with the prevailing doctrine in confirmation theory? Several authors have recently argued that not only is **abduction compatible with Bayesianism**, it is **a much-needed supplement to it**. The so far fullest defense of this view has been given by Lipton (2004, Ch. 7); as he puts it, Bayesians should also be "explanationists" (his name for the

advocates of abduction). (For other defenses, see Okasha 2000, McGrew 2003, Weisberg 2009, and Poston 2014, Ch. 7; for discussion, see Roche and Sober 2013, 2014, and McCain and Poston 2014.)

This requires some clarification. For what could it mean for a Bayesian to be an explanationist? In order to apply Bayes' rule and determine the probability for  $H$  after learning  $E$ , the Bayesian agent will have to determine the probability of  $H$  conditional on  $E$ . For that, he needs to assign unconditional probabilities to  $H$  and  $E$  as well as a probability to  $E$  given  $H$ ; the former two are mostly called "prior probabilities" (or just "priors") of, respectively,  $H$  and  $E$ , the latter the "likelihood" of  $H$  on  $E$ . (This is the official Bayesian story. Not all of those who sympathize with Bayesianism adhere to that story. For instance, according to some it is more reasonable to think that conditional probabilities are basic and that we derive unconditional probabilities from them; see Hájek 2003, and references therein.) How is the Bayesian to determine these values? As is well known, probability theory gives us more probabilities once we have some; it does not give us probabilities from scratch. Of course, when  $H$  implies  $E$  or the negation of  $E$ , or when  $H$  is a statistical hypothesis that bestows a certain chance on  $E$ , then the likelihood follows "analytically." (This claim assumes some version of Lewis' (1980) Principal Principle, and it is controversial whether or not this principle is analytic; hence the scare quotes.) But this is not always the case, and even if it were, there would still be the question of how to determine the priors. This is where, according to Lipton, abduction comes in. In his proposal, Bayesians ought to determine their prior probabilities and, if applicable, likelihoods on the basis of explanatory considerations.

Exactly how are explanatory considerations to guide one's choice of priors? The answer to this question is not as simple as one might at first think. Suppose you are considering what priors to assign to a collection of rival hypotheses and you wish to follow Lipton's suggestion. How are you to do this? An obvious—though still somewhat vague—answer may seem to go like this: Whatever exact priors you are going to assign, you should assign a higher one to the hypothesis that explains the available data best than to any of its rivals (provided there is a best explanation). Note, though, that your neighbor, who is a Bayesian but thinks confirmation has nothing to do with explanation, may well assign a prior to the best explanation that is even higher than the one you assign to that hypothesis. In fact, his priors for best explanations may even be consistently higher than yours, not because in his view explanation is somehow related to confirmation—it is not, he thinks—but, well, just because. In this context, "just because" is a perfectly legitimate reason, because any reason for fixing one's priors counts as legitimate by Bayesian standards. According to mainstream Bayesian epistemology, priors (and sometimes likelihoods) are up for grabs, meaning that one assignment of priors is as good as another, provided both are coherent (that is, they obey the axioms of probability theory). Lipton's recommendation to the Bayesian to be an explanationist is meant to be entirely general. But what should your neighbor do differently if he wants to follow the recommendation? Should he give the same prior to any best explanation that you, his explanationist neighbor, give to it, that is, *lower* his priors for best explanations? Or rather should he give even *higher* priors to best explanations than those he already gives?

Perhaps Lipton's proposal is not intended to address those who already assign highest priors to best explanations, even if they do so on grounds that have nothing to do with explanation. The idea might be that, as long as one does assign highest priors to those hypotheses, everything is fine, or at least finer than if one does not do so, regardless of one's reasons for assigning those priors. The answer to the question of how explanatory considerations are to guide one's choice of priors would then presumably be that one ought to assign a higher prior to the best explanation than to its rivals, if this is not what one already does. If it is, one should just keep doing what one is doing.

(As an aside, it should be noticed that, according to standard Bayesian usage, the term "priors" does not necessarily refer to the degrees of belief a person assigns before the receipt of *any* data. If there are already data in, then, clearly, one may assign higher priors to hypotheses that best explain the then-available data. However, one can sensibly speak of "best explanations" even before any data are known. For example, one hypothesis may be judged to be a better explanation than any of its rivals because the former requires less complicated mathematics, or because it is stated in terms of familiar concepts only, which is not true of the others. More generally, such judgments may be based on what Kosso (1992, 30) calls *internal features* of hypotheses or theories, that is, features that "can be evaluated without having to observe the world.")

A more interesting answer to the above question of how explanation is to guide one's choice of priors has been given by Jonathan Weisberg (2009). We said that mainstream Bayesians regard one assignment of prior probabilities as being as good as any other. So-called objective Bayesians do not do so, however. These Bayesians think priors must obey principles beyond the probability axioms in order to be admissible. Objective Bayesians are divided among themselves over exactly which further principles are to be obeyed, but at least for a while they agreed that the Principle of Indifference is among them. Roughly stated, this principle counsels that, absent a reason to the contrary, we give equal priors to competing hypotheses. As is well known, however, in its original form the Principle of Indifference may lead to inconsistent assignments of probabilities and so can hardly be advertised as a principle of rationality. The problem is that there are typically various ways to partition logical space that appear plausible given the problem at hand, and that not all of them lead to the same prior probability assignment, even assuming the Principle of Indifference. Weisberg's proposal amounts to the claim that explanatory considerations may favor some of those partitions over others. Perhaps we will not always end up with a unique partition to which the Principle of Indifference is to be applied, but it would already be progress if we ended up with only a handful of partitions. For we could then still arrive in a motivated way at our prior probabilities, by proceeding in two steps, namely, by first applying the Principle of Indifference to the partitions separately, thereby possibly obtaining different assignments of priors, and by then taking a weighted average of the thus obtained priors, where the weights, too, are to depend on explanatory considerations. The result would again be a probability function—the uniquely correct prior probability function, according to Weisberg.

The proposal is intriguing as far as it goes but, as Weisberg admits, in its current form, it does not go very far. For one thing, it is unclear how exactly explanatory considerations are to determine the weights required for the second step of the proposal. For another, it may be idle to hope that taking explanatory considerations into account will in general leave us with a manageable set of partitions, or that, even if it does, this will not be due merely to the fact that we are overlooking a great many *prima facie* plausible ways of partitioning logical space to begin with. (The latter point echoes the argument of the bad lot, of course.)

Another suggestion about the connection between abduction and Bayesian reasoning—to be found in Okasha 2000, McGrew 2003, Lipton 2004 (Ch. 7), and Dellsén 2018—is that the explanatory considerations may serve as a heuristic to determine, even if only roughly, priors and likelihoods in cases in which we would otherwise be clueless and could do no better than guessing. This suggestion is sensitive to the well-recognized fact that we are not always able to assign a prior to every hypothesis of interest, or to say how probable a given piece of evidence is conditional on a given hypothesis. Consideration of that hypothesis' explanatory power might then help us to figure out, if perhaps only within certain bounds, what prior to assign to it, or what likelihood to assign to it on the given evidence.

Bayesians, especially the more modest ones, might want to retort that the Bayesian procedure is to be followed if, and only if, either (a) priors and likelihoods can be determined with some precision and objectivity, or (b) likelihoods can be determined with some precision and priors can be expected to “wash out” as more and more evidence accumulates, or (c) priors and likelihoods can both be expected to wash out. In the remaining cases—they might say—we should simply refrain from applying Bayesian reasoning. *A fortiori*, then, there is no need for an abduction-enhanced Bayesianism in these cases. And some incontrovertible mathematical results indicate that, in the cases that fall under (a), (b), or (c), our probabilities will converge to the truth anyhow. Consequently, in those cases there is no need for the kind of abductive heuristics that the above-mentioned authors suggest, either. (Weisberg 2009, Sect. 3.2, raises similar concerns.)

Psillos (2000) proposes yet another way in which abduction might supplement Bayesian confirmation theory, one that is very much in the spirit of Peirce's conception of abduction. The idea is that abduction may assist us in selecting plausible candidates for testing, where the actual testing then is to follow Bayesian lines. However, Psillos concedes (2004) that this proposal assigns a role to abduction that will strike committed explanationists as being too limited.

Finally, a possibility that has so far not been considered in the literature is that abduction and Bayesianism do not so much work in tandem—as they do on the above proposals—as operate in different modes of reasoning; the Bayesian and the explanationist are characters that feature in different plays, so to speak. It is widely

accepted that sometimes we speak and think about our beliefs in a categorical manner, while at other times we speak and think about them in a graded way. It is far from clear how these different ways of speaking and thinking about beliefs—the epistemology of belief and the epistemology of degrees of belief, to use Richard Foley’s (1992) terminology—are related to one another. In fact, it is an open question whether there is any straightforward connection between the two, or even whether there is a connection at all. Be that as it may, given that the distinction is undeniable, it is a plausible suggestion that, just as there are different ways of talking and thinking about beliefs, there are different ways of talking and thinking about the *revision* of beliefs. In particular, abduction could well have its home in the epistemology of belief, and be called upon whenever we reason about our beliefs in a categorical mode, while at the same time Bayes’ rule could have its home in the epistemology of degrees of belief. Hard-nosed Bayesians may insist that whatever reasoning goes on in the categorical mode must eventually be justifiable in Bayesian terms, but this presupposes the existence of bridge principles connecting the epistemology of belief with the epistemology of degrees of belief—and, as mentioned, whether such principles exist is presently unclear.

## Bibliography

- Achinstein, P., 2001. *The Book of Evidence*, Oxford: Oxford University Press.
- Adler, J., 1994. “Testimony, Trust, Knowing,” *Journal of Philosophy*, 91: 264–275.
- Bach, K. and Harnish, R., 1979. *Linguistic Communication and Speech Acts*, Cambridge MA: MIT Press.
- Bird, A., 1998. *Philosophy of Science*, London: UCL Press.
- Bigelow, J., 2010. “Quine, Mereology, and Inference to the Best Explanation,” *Logique et Analyse*, 212: 465–482.
- Bovens, L. and Hartmann, S., 2003. “Solving the Riddle of Coherence,” *Mind*, 112: 601–633.
- Boyd, R., 1981. “Scientific Realism and Naturalistic Epistemology,” in P. Asquith and R. Giere (eds.), *PSA 1980*, (vol. II), East Lansing MI: Philosophy of Science Association, pp. 613–662.
- , 1984. “The Current Status of Scientific Realism,” in J. Leplin (ed.), *Scientific Realism*, Berkeley CA: University of California Press, pp. 41–82.
- , 1985. “Lex Orandi est Lex Credendi,” in P. Churchland and C. Hooker (eds.), *Images of Science*, Chicago IL: University of Chicago Press, pp. 3–34.
- Brem, S. and Rips, L. J., 2000. “Explanation and Evidence in Informal Argument,” *Cognitive Science*, 24: 573–604.
- Callebaut, W. (ed.), 1993. *Taking the Naturalistic Turn*, Chicago IL: University of Chicago Press.
- Campos, D., 2011. “On the Distinction Between Peirce’s Abduction and Lipton’s Inference to the Best Explanation,” *Synthese*, 180: 419–442.
- Climenhaga, N., forthcoming. “Inference to the Best Explanation Made Incoherent,” *Journal of Philosophy*, [preprint available online](#).
- Dascal, M., 1979. “Conversational Relevance,” in A. Margalit (ed.), *Meaning and Use*, Dordrecht: Reidel, pp. 153–174.
- Dellsén, F., 2017. “Reactionary Responses to the Bad Lot Objection,” *Studies in History and Philosophy of Science*, 61: 32–40.
- , 2018. “The Heuristic Conception of Inference to the Best Explanation,” *Philosophical Studies*, 175: 1745–1766.
- Douven, I., 1999. “Inference to the Best Explanation Made Coherent,” *Philosophy of Science*, 66: S424–S435.
- , 2002. “Testing Inference to the Best Explanation,” *Synthese*, 130: 355–377.
- , 2008. “Underdetermination,” in S. Psillos and M. Curd (eds.), *The Routledge Companion to the Philosophy of Science*, London: Routledge, pp. 292–301.
- , 2013. “Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation,” *Philosophical Quarterly*, 63: 428–444.
- , 2016a. *The Epistemology of Indicative Conditionals*, Cambridge: Cambridge University Press.
- , 2016b. “Explanation, Updating, and Accuracy,” *Journal of Cognitive Psychology*, 28: 1004–1012.
- , 2017. “What Is Inference to the Best Explanation? And Why Should We Care?” in T. Poston and K. McCain (eds.), *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press, pp. 4–22.

- , 2020. “The Ecological Rationality of Explanatory Reasoning,” *Studies in History and Philosophy of Science*, 79: 1–14.
- , forthcoming. *The Art of Abduction*, Cambridge MA: MIT Press.
- Douven, I. and Mirabile, P., 2018. “Best, Second-best, and Good-enough Explanations: How They Matter to Reasoning,” *Journal of Experimental Psychology: Language, Memory, and Cognition*, 44: 1792–1813.
- Douven, I. and Schupbach, J., 2015a. “The Role of Explanatory Considerations in Updating,” *Cognition*, 142: 299–311.
- , 2015b. “Probabilistic Alternatives to Bayesianism: The Case of Explanationism,” *Frontiers in Psychology*, 6: 459. doi:10.3389/fpsyg.2015.00459
- Douven, I. and Wenmackers, S., 2017. “Inference to the Best Explanation versus Bayes’s Rule in a Social Setting,” *British Journal for the Philosophy of Science*, 68: 535–570.
- Dragulinescu, S., 2016. “Inference to the Best Explanation and Mechanisms in Medicine,” *Theoretical Medicine and Bioethics*, 37(3): 211–232.
- Fann, K. T., 1970. *Peirce’s Theory of Abduction*, The Hague: Martinus Nijhoff.
- Fine, A., 1984. “The Natural Ontological Attitude,” in J. Leplin (ed.), *Scientific Realism*, Berkeley CA: University of California Press, pp. 83–107.
- Foley, R., 1992. “The Epistemology of Belief and the Epistemology of Degrees of Belief,” *American Philosophical Quarterly*, 29: 111–124.
- Folina, J., 2016. “Realism, Skepticism, and the Brain in a Vat,” in S. Goldberg (ed.), *The Brain in a Vat*, Cambridge: Cambridge University Press, pp. 155–173.
- Forster, M. and Sober, E., 1994. “How to Tell when Simpler, More Unified, or Less *Ad Hoc*, Theories will Provide More Accurate Predictions,” *British Journal for the Philosophy of Science*, 45: 1–36.
- Frankfurt, H., 1958. “Peirce’s Notion of Abduction,” *Journal of Philosophy*, 55: 593–596.
- Fricker, E., 1994. “Against Gullibility,” in B. K. Matilal and A. Chakrabarti (eds.), *Knowing from Words*, Dordrecht: Kluwer, pp. 125–161.
- Goldman, A., 1988. *Empirical Knowledge*, Berkeley CA: University of California Press.
- Hájek, A., 2003. “What Conditional Probability Could Not Be,” *Synthese*, 137: 273–323.
- Harman, G., 1965. “The Inference to the Best Explanation,” *Philosophical Review*, 74: 88–95.
- , 1973. *Thought*, Princeton NJ: Princeton University Press.
- , 1997. “Pragmatism and Reasons for Belief,” in C. Kulp (ed.), *Realism/Antirealism and Epistemology*, Totowa NJ: Rowman and Littlefield, pp. 123–147.
- Harré, R., 1986. *Varieties of Realism*, Oxford: Blackwell.
- , 1988. “Realism and Ontology,” *Philosophia Naturalis*, 25: 386–398.
- Hobbs, J. R., 2004. “Abduction in Natural Language Understanding,” in L. Horn and G. Ward (eds.), *The Handbook of Pragmatics*, Oxford: Blackwell, pp. 724–741.
- Janssen, M., 2002. “Reconsidering a Scientific Revolution: The Case of Einstein *versus*, Lorentz,” *Physics in Perspective*, 4: 421–446.
- Josephson, J. R. and Josephson, S. G. (eds.), 1994. *Abductive Inference*, Cambridge: Cambridge University Press.
- Kitcher, P., 2001. “Real Realism: The Galilean Strategy,” *Philosophical Review*, 110: 151–197.
- Koehler, D. J., 1991. “Explanation, Imagination, and Confidence in Judgment,” *Psychological Bulletin*, 110: 499–519.
- Koslowski, B., Marasia, J., Chelenza, M., and Dublin, R., 2008. “Information Becomes Evidence when an Explanation Can Incorporate it into a Causal Framework,” *Cognitive Development*, 23: 472–487.
- Kosso, P., 1992. *Reading the Book of Nature*, Cambridge: Cambridge University Press.
- Krzyżanowska, K, Wenmackers, S., and Douven, I., 2014. “Rethinking Gibbard’s Riverboat Argument,” *Studia Logica*, 102: 771–792.
- Kuipers, T., 1984. “Approaching the Truth with the Rule of Success,” *Philosophia, Naturalis*, 21: 244–253.
- , 1992. “Naive and Refined Truth Approximation,” *Synthese*, 93: 299–341.
- , 2000. *From Instrumentalism to Constructive Realism*, Dordrecht: Kluwer.
- Kvanvig, J., 1994. “A Critique of van Fraassen’s Voluntaristic Epistemology,” *Synthese*, 98: 325–348.
- Kyburg Jr., H., 1990. *Science and Reason*, Oxford: Oxford University Press.
- Laudan, L., 1981. “A Confutation of Convergent Realism,” *Philosophy of Science*, 48: 19–49.

- Lewis, D., 1980. "A Subjectivist's Guide to Objective Chance," in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Berkeley CA: University of California Press, pp. 263–293.
- Li, M. and Vitanyi, P., 1997. *An Introduction to Kolmogorov Complexity and its Applications*, New York: Springer.
- Lipton, P., 1991. *Inference to the Best Explanation*, London: Routledge.
- , 1993. "Is the Best Good Enough?" *Proceedings of the Aristotelian Society*, 93: 89–104.
- , 1998. "The Epistemology of Testimony," *Studies in History and Philosophy of Science*, 29: 1–31.
- , 2004. *Inference to the Best Explanation*, (2nd ed.), London: Routledge.
- Lombrozo, T., 2007. "Simplicity and Probability in Causal Explanation," *Cognitive Psychology*, 55: 232–257.
- , 2012. "Explanation and Abductive Inference," in K. Holyoak and R. Morrison (eds.), *Oxford Handbook of Thinking and Reasoning*, Oxford: Oxford University Press, pp. 260–276.
- , 2016. "Explanatory Preferences Shape Learning and Inference," *Trends in Cognitive Sciences*, 20: 748–759.
- Lombrozo, T. and Gwynne, N. Z., 2014. "Explanation and Inference: Mechanistic and Functional Explanations Guide Property Generalization," *Frontiers in Human Neuroscience*, 8. doi:10.3389/fnhum.2014.00700
- Maher, P., 1992. "Diachronic Rationality," *Philosophy of Science*, 59: 120–141.
- McAuliffe, W., 2015. "How Did Abduction Get Confused with Inference to the Best Explanation?" *Transactions of the Charles S. Peirce Society*, 51: 300–319.
- McCain, K. and Poston, T., 2014. "Why Explanatoriness is Evidentially Relevant," *Thought*, 3: 145–153.
- McGrew, T., 2003. "Confirmation, Heuristics, and Explanatory Reasoning," *British Journal for the Philosophy of Science*, 54: 553–567.
- McMullin, E., 1992. *The Inference that Makes Science*, Milwaukee WI: Marquette University Press.
- , 1996. "Epistemic Virtue and Theory Appraisal," in I. Douven and L. Horsten (eds.), *Realism in the Sciences*, Leuven: Leuven University Press, pp. 13–34.
- Moore, G. E., 1962. "Proof of an External World," in his *Philosophical Papers*, New York: Collier Books, pp. 126–149.
- Moser, P., 1989. *Knowledge and Evidence*, Cambridge: Cambridge University Press.
- Musgrave, A., 1988. "The Ultimate Argument for Scientific Realism," in R. Nola (ed.), *Relativism and Realism in Science*, Dordrecht: Kluwer, pp. 229–252.
- Niiniluoto, I., 1998. "Verisimilitude: The Third Period," *British Journal for the Philosophy of Science*, 49: 1–29.
- , 1999. "Defending Abduction," *Philosophy of Science*, 66: S436–S451.
- Okasha, S., 2000. "Van Fraassen's Critique of Inference to the Best Explanation," *Studies in History and Philosophy of Science*, 31: 691–710.
- Olsson, E., 2005. *Against Coherence*, Oxford: Oxford University Press.
- Pargetter, R., 1984. "The Scientific Inference to Other Minds," *Australasian Journal of Philosophy*, 62: 158–163.
- Peirce, C. S. [CP]. *Collected Papers of Charles Sanders Peirce*, edited by C. Hartshorne, P. Weiss, and A. Burks, 1931–1958, Cambridge MA: Harvard University Press.
- Poston, T., 2014. *Reason and Explanation*, Basingstoke: Palgrave Macmillan.
- Psillos, S., 1999. *Scientific Realism: How Science Tracks Truth*, London: Routledge.
- , 2000. "Abduction: Between Conceptual Richness and Computational Complexity," in A. K. Kakas and P. Flach (eds.), *Abduction and Induction: Essays on their Relation and Integration*, Dordrecht: Kluwer, pp. 59–74.
- , 2004. "Inference to the Best Explanation and Bayesianism," in F. Stadler (ed.), *Induction and Deduction in the Sciences*, Dordrecht: Kluwer, pp. 83–91.
- Putnam, H., 1981. *Reason, Truth and History*, Cambridge: Cambridge University Press.
- Roche, W. and Sober, E., 2013. "Explanatoriness is Evidentially Irrelevant, or Inference to the Best Explanation Meets Bayesian Confirmation Theory," *Analysis*, 73: 659–668
- , 2014. "Explanatoriness and Evidence: A Reply to McCain and Poston," *Thought*, 3: 193–199.
- Russell, B., 1912. *The Problems of Philosophy*, Oxford: Oxford University Press.
- Schupbach, J., 2014. "Is the Bad Lot Objection Just Misguided?" *Erkenntnis*, 79: 55–64.
- Schupbach, J. and Sprenger, J., 2011. "The Logic of Explanatory Power," *Philosophy of Science*, 78: 105–127.
- Schurz, G., 2008. "Patterns of Abduction," *Synthese*, 164: 201–234.


- Shalkowski, S., 2010. "IBE, GMR, and Metaphysical Projects," in B. Hale and A. Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*, Oxford: Oxford University Press, pp. 169–187.
- Skyrms, B., 1993. "A Mistake in Dynamic Coherence Arguments?" *Philosophy of Science*, 60: 320–328.
- Sloman, S., 1994. "When Explanations Compete: The Role of Explanatory Coherence on Judgments of Likelihood," *Cognition*, 52: 1–21.
- Sober, E., 2015. *Ockham's Razor: A User's Manual*, Cambridge: Cambridge University Press.
- Stanford, K., 2009. "Underdetermination of Scientific Theory," in *Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2009/entries/scientific-underdetermination/>.
- Teller, P., 1973. "Conditionalization and Observation," *Synthese*, 26: 218–258.
- Thagard, P., 1978. "The Best Explanation: Criteria for Theory Choice," *Journal of Philosophy*, 75: 76–92.
- van Fraassen, B., 1980. *The Scientific Image*, Oxford: Oxford University Press.
- , 1983. "Glymour on Evidence and Explanation," in J. Earman (ed.), *Testing Scientific Theories*, Minneapolis: University of Minnesota Press, pp. 165–176.
- , 1985. "Empiricism in the Philosophy of Science," in P. Churchland and C. Hooker (eds.), *Images of Science*, Chicago IL: University of Chicago Press, pp. 245–308.
- , 1989. *Laws and Symmetry*, Oxford: Oxford University Press.
- Vogel, J., 1990. "Cartesian Skepticism and Inference to the Best Explanation," *Journal of Philosophy*, 87: 658–666.
- , 2005. "The Refutation of Skepticism," in M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology*, Oxford: Blackwell Publishing, pp. 72–84.
- Weintraub, R., 2013. "Induction and Inference to the Best Explanation," *Philosophical Studies*, 166: 203–216.
- Weisberg, J., 2009. "Locating IBE in the Bayesian Framework," *Synthese*, 167: 125–143.
- Williams, J. and Lombrozo, T., 2010. "The Role of Explanation in Discovery and Generalization: Evidence from Category Learning," *Cognitive Science*, 34: 776–806.
- Williamson, T., 2017. "Semantic Paradoxes and Abductive Methodology," in B. Armour-Garb (ed.), *Reflections on the Liar*, Oxford: Oxford University Press, pp. 325–346.

## Academic Tools

 [How to cite this entry.](#)

 [Preview the PDF version of this entry](#) at the [Friends of the SEP Society](#).

 [Look up topics and thinkers related to this entry](#) at the Internet Philosophy Ontology Project (InPhO).

 [Enhanced bibliography for this entry](#) at [PhilPapers](#), with links to its database.

## Other Internet Resources

[Please contact the author with suggestions.]

## Related Entries

[epistemology: Bayesian](#) | [induction: problem of](#) | [Peirce, Charles Sanders](#) | [scientific explanation](#) | [scientific realism](#) | [simplicity](#) | [skepticism](#) | [underdetermination, of scientific theories](#)

Copyright © 2021 by

Igor Douven <[igor.douven@paris-sorbonne.fr](mailto:igor.douven@paris-sorbonne.fr)>

[Open access to the SEP is made possible by a world-wide funding initiative.](#)  
[Please Read How You Can Help Keep the Encyclopedia Free](#)

The Stanford Encyclopedia of Philosophy is [copyright © 2021](#) by [The Metaphysics Research Lab](#), Department of Philosophy, Stanford University

Library of Congress Catalog Data: ISSN 1095-5054