2

What Is Abduction?

In this chapter, I introduce explanatory reasoning from three different angles. First, I give some real-world examples that are offered as clear and uncontroversial instances of explanatory reasoning. To forestall misunderstanding, I stress that I do not present these examples by themselves to show or even to suggest that explanatory reasoning is a form of reasoning in its own right that cannot be adequately modeled, for example, in Bayesian terms. The book as a whole *is* intended to show that the kind of reasoning that these examples exhibit is not, or at least not in all cases, best modeled in those terms; that we can model them by explicitly factoring in explanatory considerations, even in ways inconsistent with the tenets of Bayesianism; and that there is nothing necessarily wrong with the resulting model. The examples to be considered shortly are merely meant to give the reader—especially the reader new to the topic of abduction—an impression of what abductive reasoning actually is.

Next, I look at some of the extant theory concerning this mode of reasoning, in particular what philosophers have been saying about it. There is a welter of proposals concerning the shape that explanatory reasoning might take. Section 1.2.1 has made it clear that, in my opinion, such different proposals can peacefully coexist.

Finally, I describe the main source of interest in abduction, at least in philosophy, which is the belief that, as previously mentioned, abduction may contribute to resolving cases of underdetermination, that is, cases in which observation alone, or observation plus logic, is not enough to settle a question or even to arrive at a reasonably grounded answer to the question, because a number of rival answers appear equally reasonable. Various philosophers have argued that abduction can be invoked to break such deadlocks. Abduction is used to this effect in chapter 8, where I respond to the Cartesian skeptic.

2.1 Abduction in the Wild

2.1.1 Newton and the Raid on the Medway

In 1667, under the leadership of Michiel Adriaanszoon de Ruyter the Dutch fleet invaded the Thames and inflicted what is still the greatest ever defeat on the British navy.[1] At the time of this Raid on the Medway, as the invasion is also called, Isaac Newton was a Fellow at Trinity College, Cambridge, where the rumble of the battle could be heard. In his *Memoirs of Sir Isaac Newton's Life*, William Stukeley reports about it as follows:

> Their guns were heard as far as Cambridg, and the cause was well known; but the event was only cognisable to Sir Isaac's sagacity, who boldly pronounc'd that they had beaten us. The news soon confirm'd it, and the curious would not be easy whilst Sir Isaac satisfy'd them of the mode of his intelligence, which was this; by carefully attending to the sound, he found it grew louder and louder, consequently came nearer; from whence he rightly infer'd that the Dutch were victors. (Stukeley, cited in Westfall, 1980, p. 194)

According to Richard Westfall (1980), the "mode of his intelligence" referred to here struck the other Fellows in Cambridge with awe. How *did* Newton arrive at the conclusion that the Dutch had won? What made it *right*—if it *was*—for him to infer from the sound of the gun fire getting louder that the Dutch had beaten the British fleet?

2.1.2 Discovering Neptune

Uranus was discovered in 1781, and by the mid-nineteenth century it had become clear that its orbit deviated from the orbit predicted by Newton's theory of gravity. Astronomers realized that this finding could betoken a straightforward falsification of the theory. But that thought was anathema to most at the time, given the theory's otherwise hugely impressive track record. Another possibility was that the predictions derived from Newton's theory were based on faulty calculations. New calculations by the French mathematician Urbain Le Verrier showed that earlier ones indeed contained some mistakes. On the other hand, those new calculations still did not result

1. I have found that this fact is virtually unknown outside the Netherlands. In the Netherlands, by contrast, almost everyone knows the story of the invasion, probably because it is the only success in the whole military history of our country.

in an orbit that accorded with the available data. Le Verrier considered various possible explanations of why Uranus's orbit was inconsistent with predictions based on Newton's theory; for instance, that Uranus was itself orbited by a giant moon that steered it off course, or that it was hit by a comet, or that it was hampered by something like an ether. In the end, he came to think that the wobbles in Uranus's orbit were attributable to the gravitational pull of another, so far unknown, planet (Levenson, 2015, p. 36). He also thought that if it existed, he could calculate the position of the planet at that time. He first sent his calculations to the astronomers of the Parisian observatory. When his countrymen were slow to act, however, and had after several weeks not made any attempt to verify Le Verrier's prediction, he turned for help to the German astronomer Johann Galle of the observatory in Berlin. Galle wasted no time. As soon as he had received Le Verrier's message, he directed his telescope to the location in the sky where according to Le Verrier's calculations the planet should be visible, and within hours he discovered what soon became known as "Neptune."

2.1.3 Non-discovering Vulcan

Not long after the discovery of Neptune, Le Verrier concerned himself with an at the time unexplained irregularity (from the viewpoint of Newtonian gravitation) in the orbit of yet another planet, Mercury. The discrepancy was much smaller than in the case of Uranus's orbit but large enough to worry astronomers of the time. In 1859, Le Verrier proposed that this discrepancy was to be explained by the presence of a small, as yet unknown, planet, a proposal that seemed entirely reasonable in light of how Neptune had been discovered. Soon reports arrived of observations of this supposed new planet—which had been dubbed "Vulcan"—and examining records of past observations revealed many that could also be classified as sightings of such a planet. On the other hand, those observations were used to calculate the orbit of Vulcan, from which predictions were derived about where and when one should be able to observe it again, and although there were sporadic reports—typically by amateur enthusiasts—of sightings that seemed to confirm those predictions, more rigorously organized attempts to spot Vulcan remained unsuccessful. This did not, however, seal the fate of Vulcan. Le Verrier considered the possibility that Vulcan was so small that it would be hard to observe from earth or that it could be observed only under especially favorable conditions.

However, doubts kept lingering, precisely because if the planet was so small, its mass would probably also be too small to completely explain the phenomenon that got the whole inquiry started. Also, failed attempts to observe Vulcan kept piling up, and over time, absence of evidence came to be regarded as evidence of absence (Levenson, 2015, p. 108). The solution to the puzzle of why Mercury's orbit deviated from Newtonian predictions came only in 1915, when Albert Einstein presented his General Theory of Relativity, which explained the wobble in Mercury's orbit in terms of the effect mass has on the curvature of spacetime (Einstein, 1915).[2]

2.1.4 On the Origin of Species

Charles Darwin's *Origin of Species* (1859) is one of the most important works of science of all time, according to some *the* most important (see, e.g., Wilson, 2009, p. xv). In the work, Darwin argued famously for a theory of the diversity of life forms and of the evolution of species by postulating mechanisms of individual variation and selective retention, and against the hypothesis that species are the immutable results of a divine act of creation. A revolutionary theory at the time of its inception, Darwin's proposal nevertheless gained rapid acceptance in the scientific community. That may have occurred, at least in part, because the theory was able to explain a great variety of facts, some of which Darwin did not even contemplate at the time he wrote *Origin*. As he put it in the sixth edition of the work,

> It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing; but it is a method used in judging of the common events of life, and has often been used by the greatest natural philosophers. The undulatory theory of light has thus been arrived at; and the belief in the revolution of the earth on its own axis was until lately supported by hardly any direct evidence. (Darwin, 1876, p. 421)

The facts that Darwin refers to as having been specified previously come from a variety of domains, including embryology, morphology, paleontology, and biogeography, and most of those facts were seen as wholly unconnected, prior

2. Levenson (2015) gives a fascinating and highly detailed account of the fruitless hunt for Vulcan.

to the advancement of Darwin's theory. As various authors have pointed out (e.g., Ruse, 1975, 1995; Thagard, 1977), Darwin was heavily influenced by the philosopher of science and polymath William Whewell, who had made a strong case for the confirmatory power of what he called "consilience of inductions," which occurs if a theory is successfully proposed to explain a given body of data and is then found to also explain, without requiring any further adjustment, one or more other bodies of data, none of which had played a role in the conception of the theory. Such a consilience of inductions brings about a "unification of two or more hitherto disparate areas of understanding beneath one or a few high-level hypotheses or established laws" (Ruse, 1995, pp. 2–3) and thereby provides a theory with a "stamp of truth" (Whewell, 1847, II, p. 66). In the preceding passage, Darwin appears to concur, as did and continue to do many of the advocates of the theory of natural selection: that this theory can explain such a great variety of facts in such a satisfactory fashion seems to rule out its falsity.

2.1.5 The Discovery of the Electron

The electron was the first known subatomic particle. It was discovered in 1897 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. Improving an experimental setup that had earlier been used (unsuccessfully) by Heinrich Hertz, Thomson let cathode rays travel between two parallel metal plates. He then showed that when those plates were oppositely electrified, the cathode rays were deflected. (That Hertz's experiment had shown no effect was due, Thomson argued, to the fact that Hertz had failed to sufficiently evacuate the cathode tube.) From this result, Thomson concluded that cathode rays are indeed charged particles, 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, 1897, p. 302)

Today's physicists are still convinced that Thomson drew the right conclusion.

2.1.6 Gödel's Demons

Kurt Gödel was the greatest logician of the twentieth century and possibly the greatest logician ever. He proved the completeness of first-order logic (logical truth coincides with finite tinkering in a system of axioms and rules that most freshmen are comfortable working with in just a couple of weeks) and the incompleteness of arithmetic (no consistent system of axioms and rules captures the notion of arithmetic truth). He was a genius, but he also suffered from mental illness all his life (Wang, 1996, ch. 1). Whether or not due to his illness, he was a strong believer in ghosts and demons. In a conversation with Georg Kreisel, another brilliant logician, he asked whether Kreisel had noticed the peculiar asymmetry that there are many more unpleasant events taking place in the universe than pleasant ones. According to Gödel, this was best explained by assuming the existence of demons. Kreisel responded that he was not so sure and that it might be in our nature to notice unpleasant things more than we do pleasant things and that, more generally, we may have an inborn tendency to be unhappy with what surrounds us. Gödel agreed that we have this tendency but also thought that it, too, was best explained by the existence of demons (Kreisel, 1980).

2.1.7 A Head in Jackson

On June 11, 2017, CNN reported on its website that in Jackson, Mississippi, a head had been found on the front porch of a home and that a headless body had been found less than a mile away. At that point in time, according to CNN, a medical examiner was still trying to determine whether the head and body were from the same person. Two days later, another CNN report confirmed that they were indeed from the same person (and the victim was identified). That was what I had believed, without the shadow of a doubt, from the moment I first read the story on June 11. Perhaps you also read this story when it broke and drew the same inference I drew. Even if not, you will still see why the conclusion appeared entirely reasonable to me, immediately on June 11, before the medical examiner's report was released.

2.2 Deduction, Induction, Abduction

The foregoing cases should give you some idea of what abduction is and how it works in practice. To acquire a more systematic understanding of abduction,

we first need to make some basic distinctions concerning inference. Abduction is normally thought of as one of three major types of inference, the other two being deduction and induction. The distinction between deduction, on 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 *must* be 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 Φs are Ψs.
> Object *a* is a Φ.
> Hence, *a* is a Ψ.

If all Φs are Ψs, and *a* is a Φ, then there is simply no way for *a* not to be a Ψ. Any argument of the above form is said to be (deductively) *valid*.[3]

Not all inferences are of this variety. Consider, for instance, the inference of "Alice is rich" from "Alice lives in Chelsea" and "Ninety-seven percent of the people who live 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 Alice be a member of the small minority of nonrich inhabitants of Chelsea. Hence, the said inference is *invalid*.

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 frequency information, whether provided numerically, as in the preceding example about Alice (which falls into the inductive category), or in a more qualitative fashion (e.g., "Most people living in Chelsea are rich").

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 nongray ones, and infer from this that all elephants are gray because that would provide the *best explanation* for why you have observed so many

3. Some authors include in the class of deductively valid arguments those that hold in virtue of their form *and* certain meaning postulates, such as the argument with premise "Marble *a* is red" and conclusion "Marble *a* is colored" (see, e.g., Carnap, 1952).

gray elephants and no nongray ones. You would thereby make an abductive inference.

The best way to distinguish between induction and abduction may be the following: while both are *ampliative*, meaning that the conclusion has more information content than is in the premises, 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.

Not all philosophers agree that there is a distinction to be made between inductive and abductive inference. As intimated previously, Harman believes that inductive inference is really a form of abduction. For instance, he would argue that we infer that Alice is rich in the above example because, if true, that would best explain why she lives in Chelsea, given the background knowledge that 97 percent of the Chelsea population are rich. There is research suggesting that people actually tend to prefer explanatory information over statistical information in making inferences (e.g., Bes et al., 2012). Hence, maybe people *would* reach the conclusion that Alice is rich on the basis of explanatory considerations rather than by merely drawing on the provided frequency information. To me, however, the appeal to explanation seems strained in this case.

This book concerns abduction, so if inductive inferences are really just a particular type of abductive inferences—which I do not believe—then so much the better: it would expand the book's market. It *would* be problematic if Richard Fumerton (1980, 2017) were right: that what philosophers tend to regard as abductive inferences are in fact merely inductive ones, albeit topped with a reference to explanation that is not doing any real work.

Fumerton may certainly be right that *some* cases advertised as paradigm cases of abduction actually only exploit frequency information, or at any rate that their appeals to explanation are gratuitous. For instance, elsewhere (Douven, 2017a) I used a little story from the opening pages of Hilary Putnam's *Reason, Truth and History* (Putnam, 1981) to illustrate the use of abductive reasoning. In this story, you are walking along the beach when you see what looks like a picture of Winston Churchill in the sand. You realize that what you see could be the trace of an ant crawling on the beach. But I suggested that a by far better explanation in this case is that someone intentionally drew a picture of Churchill in the sand—which is what you would supposedly be justified to believe. Here, one might claim that your inference would be justified merely by extrapolating from previous experiences. It is not so much

that you will have seen many portraits of known persons in the sand and that all those were known to have been drawn by humans. To the best of my recollection, the closest to a portrait in the sand that I have ever seen was a smiley face. However, your reference class need not be confined to portraits in the sand. Most likely, you will never have seen any portrait—on paper, on canvas, on a T-shirt, or on the sidewalk—that was not drawn by someone but instead was a random configuration of, for example, dust specks that just *happened* to look like a portrait of a known person.

Even if this is granted, however, it is not nearly enough to establish the general claim that abductive inferences are just misunderstood inductive inferences. In particular, I take the previous section to consist of cases that are much more naturally understood as cases of abductive reasoning than as cases of inductive or, for that matter, deductive reasoning. The appeal to explanation in those cases appears anything but gratuitous.

To start with the head-in-Jackson case, it is clear that the relevant conclusion could not have been reached on the basis of a deductive argument: that a head and a headless body found in each other's proximity belong together does not follow from anything we know about heads, bodies, decapitations, or anything else. If it did, what was the medical examiner needed for? In that case, the Jackson police should have a put a logician on the case.

As for the possibility that the conclusion was really reached on the basis of frequency information, note that there was no such information on the CNN website. There was nothing about murder rates in Jackson (or anywhere else) nor was there information about findings of heads and headless bodies and how often, in such findings, the head and headless body were from the same person. Rather, I suspect that such findings are so rare that there are simply no reliable statistics that one could consult. And even if reliable statistics *were* available in this case, they were not available to *me*, and so they did not play a role in the reasoning that led *me* to arrive at the conclusion that head and body were from the same person.

Nor could we easily have reasoned by analogy, which is sometimes treated under the heading of inductive inference (Carnap, 1945, 1980; Skyrms, 1993a; Hill & Paris, 2013; Paris & Vencovská, 2017).[4] It was not as if last year there

4. Lindenbaum-Hosiasson (1941) sees reasoning by analogy as being distinct from inductive reasoning, the latter being concerned with increasing our confidence in laws or general hypotheses, the former with increasing our confidence that a certain observable but as yet unobserved fact obtains. A number of later authors (e.g., Thagard, 1989, 2000; Schurz, 2008a)

had been the finding of a leg and a legless body near each other, which then turned out to belong together, and the year before the finding of a shoulder and a shoulderless body near to each other, and so on. Such situations had not occurred, at least not to my knowledge, so that again, even if they had occurred, I could not justifiably reason by analogy from those situations to the conclusion that the head and body that were found in Jackson, Mississippi, were from the same person.

But then what did make me, and presumably many others who read the same story on June 11, so certain that the head and body belonged together even before the medical examiner's conclusion was reported in the media? It is hard to see how the conclusion, on June 11, 2017, could *not* have been the outcome of an abductive inference. At a minimum, that the head and body belonged together was, at the time, clearly the best explanation of the evidence known then. To be sure, it was not the only explanation: there might have occurred two killings, say, followed by two decapitations, with the head of one victim being dumped close to where the body of the other victim was dumped, whether coincidentally or as planned by the perpetrators. But, at least in light of the information on the CNN website on June 11, no other candidate explanation appeared nearly as plausible as the conclusion I drew: that the head and the body were from one and the same person. It was not a conscious decision on my part to make an abductive inference in this case, but to the extent that I have introspective access to my own reasoning, I would say that this is exactly the inference I did make then.

Similarly in Newton's case, the "mode of intelligence" that stunned his colleagues is most plausibly thought of as involving abductive reasoning. It is exceedingly difficult to think of a reasonable set of premises—reasonable from Newton's perspective at the time—from which the conclusion that the Dutch had won follows deductively. Nor did the Dutch—or any other nation that possessed a sizable fleet in the second half of the seventeenth century—invade England frequently enough for Newton's reasoning to be naturally construed as inductive. Instead, it seems that what led Newton to his conclusion is that a Dutch victory was the *best explanation* for his evidence. There are various potential explanations of why the sound of the cannon fire grew louder and louder that do *not* involve a Dutch victory. For instance,

---

have presented analogical reasoning as being essential to abductive reasoning, or at least as a mode of reasoning involved in some abductions.

the British fleet might have defeated the Dutch, but then that victory might have been followed by a mutiny in which the British seamen turned against their own headquarters. However, this and other potential explanations are topped, in terms of explanatory goodness, by the hypothesis that the Dutch fleet had beaten the British.

Or again, in the case of Thomson's reasoning about what his experimental results showed, the conclusion that cathode rays consist of negatively charged particles does not follow logically from those results, nor could Thomson draw on any relevant frequency information. More generally, as Gerhard Schurz (2008a; 2019, section 11.2.5) and Ruth Weintraub (2017) argue, it is hard to imagine how we could ever arrive at postulating an unobservable entity merely on the basis of inductive reasoning. It is much more reasonable to suppose that Thomson could "see no escape from the conclusion" about cathode rays because that conclusion was the best—in this case presumably the only plausible—explanation of his results that he could think of. I take this and the other cases just discussed to offer strong support for the claim that abduction is a type of inference genuinely distinct from induction.

If great minds like Newton, Le Verrier, and Thomson rely on abduction, that certainly speaks in favor of this mode of reasoning. But great minds can go wrong and can, like the rest of us, be subject to biases. Thus we should still ask whether abduction is any good. That question takes up much of the rest of the book, but the preceding stories already allow us to make a couple of pertinent observations.

The first is that abduction is certainly not fail-safe, as the cases of Vulcan and Gödel's demons show. The Gödel case is special, of course, precisely because it involves someone whose mind was at least in some ways not functioning properly (even if in other ways it was functioning brilliantly). Given how background knowledge informs our judgments of explanation quality, we cannot expect abduction to be any good when applied by someone with deranged ideas about reality, as Gödel's certainly were. If one's picture of reality is already distorted, then one may well come to regard the existence of demons to best explain why there is so much misery in the world.[5]

Le Verrier's view of the world was *not* distorted, as far as is known. Not only was it reasonable for him to judge that the anomalous rate of precession of the perihelion of Mercury's orbit was best explained by the existence of

5. See also chapter 1, footnote 16.

an as yet undiscovered planet, but one could argue that the later judgment was even more reasonable than the earlier one, which had led to the discovery of Nepture. After all, in the case of Mercury, the judgment was additionally backed by the known success of the analogous earlier proposal. If a proposed hypothesis explaining a given phenomenon is similar to one known to explain a similar phenomenon, that should increase the degree of acceptability of the former (Lindenbaum-Hosiasson, 1941; Carnap, 1980; Thagard, 1989, 2000). Nevertheless, Le Verrier's later judgment turned out to be false.

That abduction sometimes leads to false conclusions is what you would expect, given that it is an ampliative type of inference. In itself, that is not a reason to dismiss it. It would of course be wonderful if we came by purely with deductive reasoning, such that our reasoning would never introduce falsehoods, provided that we took some care in checking the validity of the steps in our arguments. But as Gerhard Schurz and Ralph Hertwig (2019, p. 19) rightly remark, while deduction has "maximum ecological validity"—meaning that it is entirely general and can be relied on in every possible situation—it also has "very low applicability," in that "the prevalence of deductive inferences with *nontrivial* conclusions is low." In other words, we are practically bound to use ampliative types of reasoning, but then we must reckon with the fact that, at times, our conclusions will be false even if the premises from which we start are not.

Even though abductive reasoning sometimes produces falsehoods, it may still be *reliable*: it may be that on *most* occasions when we reason abductively we will reach a true conclusion, provided that we start from true premises. A quick count shows that of the six cases discussed in the previous section at least five exhibit abductive reasoning on the basis of true premises (the Gödel case may be the exception, although we do not know exactly what his premises were), and in four of those five the reasoning also led to a true conclusion. But the sample that we are considering is not only small, it is also highly nonrandom and hence not a good basis for an assessment. For all that has been shown, my being correct in the case of the head in Jackson was a lucky hit for a mode of reasoning that should still be advised against.

Nonetheless, the foregoing suggests that we might be able to show abduction reliable simply by counting the successful applications. We merely have to widen our sample and make sure that it is truly representative, and then the proportion of cases in which abduction worked should give us an indication of its reliability (or otherwise). That would amount to an inductive

justification of abduction. There have also been attempts to justify abduction abductively. I am sympathetic to all such attempts, at least to the extent that I believe the justification for abduction must proceed empirically on the basis of its success rate. Trying to justify abduction a priori, by deducing it from first principles, is a hopeless endeavor, I believe (see also section 6.2), just as we cannot hope to justify induction by deducing it from first principles (Schurz, 2019). In fact, even a deductive justification of deduction may be out of reach (Haack, 1976). So, I maintain that abduction should be justified empirically. However, doing so is not nearly as straightforward as one might initially think, at least not if the aim is to justify it *broadly*, as opposed to justifying it for a limited class of cases.

Note also that a justification of abduction must consist of more than a demonstration of its reliability. Abduction might be reliable but possibly not as reliable as some other ways of forming or changing our beliefs. In that case, it could be foolhardy to rely on abduction instead of some more reliable rule. It *could* be but need not be, because those other rules might have other defects that make them unsuited. However, the same may hold for abduction: perhaps it generally leads to true conclusions when applied on the basis of true premises, but actually relying on it could still get one into serious trouble (e.g., because it exposes one to Dutch bookies).

But the justificatory status of abduction is a topic for later chapters. We are still in the process of clarifying what abduction is. In the preceding discussion, we have positioned abduction relative to the other main types of inference. Next, we have a look at how abduction has been characterized by various philosophers. Philosophers have actually used the term "abduction" in two somewhat different senses. In the sense in which it is used most frequently in modern literature, it refers to the place of explanatory reasoning in *justifying* hypotheses; abduction in this sense is also known as "Inference to the Best Explanation." By contrast, in the older sense, "abduction" refers to the place of explanatory reasoning in *generating* hypotheses. This book concerns abduction strictly in the modern sense.[6] Nevertheless, we start with a historical digression by briefly going into how the term "abduction" was introduced in the philosophical literature.

6. However, I agree with Niiniluoto (1999a, p. S442) that it is not always possible to distinguish between the two roles of abduction. For a book-length treatment of how the two roles relate to one another, see Niiniluoto (2018).

2.2.1 Historical Digression: Peirce on Abduction

The term "abduction" was coined by Charles Sanders Peirce in his work on the logic of science. He introduced it to denote a type of nondeductive inference that was different from the already familiar inductive type. It is, however, a common complaint that no coherent picture emerges from Peirce's writings on abduction.[7] Yet it is clear that, as Peirce understood the term, "abduction" did not quite mean what it is currently taken to mean (Campos, 2011; McAuliffe, 2015; Yu & Zenker, 2018). One main difference between his conception and the modern one is that whereas according to the modern conception, abduction belongs to what the logical empiricists called the "context of justification"—the stage of scientific inquiry in which we are concerned with the assessment of theories—for Peirce abduction had its proper place in the context of *discovery*, the stage of inquiry in which we in the first place try to generate theories that may later be assessed. As he writes, "[a]bduction is the process of forming explanatory hypotheses. It is the only logical operation which introduces any new idea" (1934, section 172); elsewhere he writes that abduction encompasses "all the operations by which theories and conceptions are engendered" (1934, section 590). Deduction and induction, then, come into play at the later stage of theory assessment: deduction helps to derive testable consequences from the explanatory hypotheses that abduction has helped us to conceive, and induction finally helps us to reach a verdict on the hypotheses, where the nature of the verdict is dependent on the number of testable consequences that have been verified.[8]

7. Perhaps this is not surprising, given that he worked on abduction throughout his career, which spanned a period of more than fifty years. For a concise yet thorough account of the development of Peirce's thoughts on abduction, see Fann (1970); in addition, Niiniluoto (2018) contains valuable historical background information.

8. Schurz (2008a) defends a view of abduction that is very much in the Peircean spirit. In this view, "the crucial function of a pattern of abduction . . . consists in its function as a search strategy which leads us, for a given kind of scenario, in a reasonable time to a most promising explanatory conjecture which is then subject to further test" (Schurz, 2008a, p. 205). Note, however, that Schurz does acknowledge a role for abduction in the justification of hypotheses, though that role is, in his view, a relatively minor one. Schurz's paper is also of interest because of the rich typology of patterns of abduction that it puts forth. Most notably, he distinguishes between *selective* and *creative* abduction. The first is invoked when we face a large space of possibilities from which to select a hypothesis, and the second serves to introduce new concepts or entities. For instance, the abductive reasoning that led Newton to conclude

As Harry Frankfurt (1958) notes, however, the foregoing view is not as easy to make sense of as might at first appear. Abduction is supposed to be part of the logic of science, but what exactly is *logical* about inventing explanatory hypotheses? According to Peirce (CP 5.189), abduction belongs to logic because it can be given a schematic characterization, to wit, the following:

> The surprising fact, *C*, is observed.
> But if *A* were true, *C* would be a matter of course.
> Hence, there is reason to suspect that *A* is true.

But Frankfurt rightly remarks that this is not an inference leading to any new idea. After all, the new idea—the explanatory hypothesis *A*—must have occurred to one *before* one infers that there is reason to suspect that *A* is true, for *A* already figures in the second premise.

Frankfurt then goes on to argue that a number of passages in Peirce's work suggest an understanding of abduction, not so much as a process of *inventing* hypotheses but rather as one of *adopting* hypotheses, where the adoption of the hypothesis is not as being true or verified or confirmed but as being a worthy candidate for further investigation. On this understanding, abduction could still be thought of as part of the context of discovery. It would work as a kind of selection function or filter, determining which of the hypotheses that have been conceived in the stage of discovery are to pass to the next stage and be subjected to empirical testing. The selection criterion is that there must be a reason to suspect that the hypothesis is true, and we will have such a reason if the hypothesis makes the relevant observed facts a matter of course. This would indeed make better sense of Peirce's claim that abduction is a logical operation.

Nevertheless, Frankfurt ultimately rejects this proposal as well. Given, he states, that there may be infinitely many hypotheses that account for a given fact or set of facts (which Peirce acknowledged), it can hardly be a sufficient condition for the adoption of a hypothesis (in the preceding sense) that its truth would make that fact or set of facts a matter of course. At a minimum, abduction would not seem to be of much use as a selection function. One may doubt whether this is a valid objection, however. Previewing some of the discussion forthcoming in this chapter in connection with underdetermination arguments, I note that it is by no means clear that "accounting for

that the Dutch had won was of the former kind, whereas the abductive reasoning that led Thomson to postulate the existence of electrons was of the latter.

a given fact" is to be identified with "making that fact a matter of course." For all Frankfurt says, for a hypothesis to account for a fact, it is enough if it entails that fact. It is dubitable that entailment is sufficient for explanation, or at least for what we would regard as a satisfactory explanation. Given that it seems reasonable to read the phrase "making a given fact a matter of course" as "giving a satisfactory explanation of that fact," one could argue in response to Frankfurt's objection that even if there are an infinity of hypotheses that account for a given fact, there may still be only a handful that could be said to explain it satisfactorily. It is for Peirce scholars to decide whether this proposed interpretation is plausible in the light of Peirce's further writings.

Even if "making a given fact a matter of course" can be read as "giving a satisfactory explanation of that fact," it is remarkable that there is in Peirce's writings on abduction no reference to the notion of *best* explanation. Some satisfactory explanations might still be better than others, and there might even be a unique best one. This idea is crucial in all recent thinking about abduction. Therein lies another main difference between Peirce's conception of abduction and the modern one, which is the only one with which we are concerned from here on.

2.2.2 How Contemporary Philosophers Characterize Abduction

The core idea of abduction is often said to be that explanatory considerations have confirmation-theoretic import, or that explanatory success is a (possibly fallible) mark of truth, or something similar. As mentioned previously, critics have pointed out that formulations like these are slogans at best and that they can be cashed out in a great variety of ways. "Which is it?", the critics want to know. As also noted, I think it is fine to conceive of abduction as a slogan that can be filled in per context of use. Nevertheless, a number of philosophers have proposed general statements of abduction that are meant to go beyond the known slogans. We look at these not just for the sake of completeness but also because some of them might be, if not universal explications, at least candidate fill-ins for certain contexts. Moreover, it is important to point out that a number of well-known objections against abduction that some might take to militate against the idea of abduction in general really pertain only to specific proposed precisifications.

That is for instance true of the following version, which is common in textbooks:

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

This is often immediately followed by the comment that although hopefully headed in the right direction, it may be slightly too strong, and that the inference that abduction warrants is only to the *probable* truth of the best explanation, or to the *approximate* truth of the best explanation, or to the *probable approximate* truth.

The main problem with ABD1 is not solved by any of these suggestions, however. Because abduction is ampliative, as explained earlier, it will not be a sound rule of inference in the strict logical sense, however exactly it is explicated. It can still be *reliable* in that it mostly leads to true conclusions whenever all premises are true. An obvious necessary condition for ABD1 to be reliable in this sense is that *most of the time* when E is true and H best explains E, 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, because ABD1 takes as its premise only that some hypothesis is the best explanation of the evidence *as compared with 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 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 that we have come up with, because otherwise ABD1 may well lead us to believe "the best of a bad lot" (van Fraassen, 1989, p. 143).

How reasonable is it to suppose that this extra requirement is usually fulfilled? Not at all, presumably. To believe otherwise, we would have to 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 upon, inter alia, the absolutely best explanation of this data. After all, hardly ever will we have considered, or will it even be possible to consider, *all* potential explanations. As Bas van Fraassen (1989, p. 144) compellingly shows, it is a priori rather implausible to hold that we are thus privileged.

In response to this "argument of the bad lot," one might argue that the challenge to show that the best explanation is always or most often among the hypotheses considered can be met without having to assume some form of

privilege. Given the hypotheses that we have managed to come up with, we can always generate a set of hypotheses that jointly exhaust logical space. Suppose that $H_1, \ldots, H_n$ are the candidate explanations that we have so far been able to conceive. Then simply define $H_{n+1} := \neg H_1 \wedge \cdots \wedge \neg H_n$ and add this new hypothesis as a further candidate explanation to the ones that we already have. Obviously, the set $\{H_1, \ldots, H_{n+1}\}$ is exhaustive, in that one of its elements must be true. Following this simple procedure would seem enough to make sure that we never miss out on the absolutely best explanation.[9]

Alas, there is a catch. Even though there may be many hypotheses H_j that imply H_{n+1} and, had they been formulated, would have been evaluated as better explanations for the data than the best explanation among the candidate explanations that 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 that we have as competing explanations the Special Theory of Relativity and Lorentz's version of the ether theory. Then, following the preceding proposal, we may add to our candidate explanations the further hypothesis that neither of these two theories is true. But surely this new hypothesis will be ranked quite low qua explanation—if it will be ranked at all, which seems doubtful because 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 that we consider.[10]

Van Fraassen (1989) took the argument of the bad lot to show that if abduction can be made to work at all, it must be in the form of a probabilistic rule akin to Bayes's rule, but different from it in that it explicitly factors in explanatory considerations. And such rules, he argued, can be shown to be incoherent by Lewis's dynamic Dutch book argument. Before we turn to probabilistic versions of abduction and the issue of their coherence, it is to be pointed out that the argument of the bad lot is not nearly as general as van Fraassen appears to suppose.

9. See Lipton (1993) for a proposal along these lines. See Schupbach (2014) for a different response, and see Dellsén (2017) for discussion. McCain and Poston (2019) dispel a seeming objection to abduction related to the argument of the bad lot.

10. The argument of the bad lot, the response to it summarized in the preceding paragraph, and the problem with that response, pointed out here, were all foreshadowed in a rudimentary form in van Fraassen (1980, pp. 21–22).

A promising response to the argument begins with the observation that it 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 a better explanation of the evidence than the other hypotheses available (see Kuipers, 2000, p. 171). This asymmetry 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 that it requires an absolute premise. For instance, following Alan Musgrave (1988) or 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 $H_1, \ldots, H_n$ of E, infer the truth of *that* H_i that explains E best, *provided* H_i is satisfactory / good enough qua explanation.

Even apart from the asymmetry problem, ABD2 immediately strikes us as more plausible than ABD1. Who, after all, would want to infer the truth of a poor explanation of the evidence, even if the competing explanations were poorer still? But what counts as satisfactory or good enough? Lipton and Musgrave make no attempt to make these notions precise. An interesting proposal has been made by Finnur Dellsén (2021). In a nutshell, his proposal is that for an explanation to be good enough it must have survived sufficiently many attempts to replace it with a better alternative. One imagines that what counts as sufficiently many attempts and even what counts as a better alternative may vary somewhat from one context to another, which would be consistent with our idea that abduction is a slogan that needs to be filled in per context of use.

We might want to add a further qualification to ABD2. For as Alexander Bird (2010) notes, even if there is a best explanation for the evidence at hand and that best explanation is also good enough, it is still an open question how much better an explanation it is than its closest competitor(s). The answer to this question, he thinks, may make a difference in what we can infer from the evidence. Specifically, he suggests that given a best and good-enough explanation, we are entitled to infer its truth only if it is significantly better than its closest competitor. If the second-best explanation of the evidence is

perfectly good as well and is just barely topped by the best one, then according to Bird, "our faith in that slightly better one must be slim" (Bird, 2010, p. 346).[11]

A second way to formulate a symmetric or congruous version of abduction is 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., 1984, 1992, 2000; see also Niiniluoto, 2018, p. 56):

ABD3 Given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, if H_i for some i with $1 \leqslant i \leqslant n$ explains E better than H_j for all j: $1 \leqslant j \leqslant n$ such that $j \neq i$, infer that H_i is closer to the truth than any of those other hypotheses.

Clearly, ABD3 requires an account of closeness to the truth, but many such accounts are on offer today (see ch. 5), some of which would certainly be suitable for our purposes. Still, while it circumvents the best-of-a-bad-lot objection, ABD3 could be criticized for being too weak. After all, it gives no indication whatsoever of how far from the truth the best explaining hypothesis is.

There is a more general concern that one might have about replacing ABD1 by either ABD2 or ABD3. Abduction has been touted as a powerful weapon by scientific realists. It is supposed to assist us in choosing between (or among) empirically equivalent theories, thereby breaking the deadlock of underdetermination, as discussed in the next section, but it has also been wielded more directly by realists, who proclaimed that their position yields the best explanation for the instrumental successes of modern science and that this gives good reason for holding realism to be true (see section 6.2 for more on this).

Van Fraassen has been careful to state the realism–antirealism debate as being, ultimately, about the goal or goals of science. According to him, realists deem the goal of science to consist of uncovering the hidden processes and mechanisms behind the phenomena and thereby to arrive at theories that are true not just as far as they pertain to the observable part of the world but also as far as they pertain to the unobservable part. For antirealists, or at least the sort of antirealists to whom van Fraassen belongs, the goal is only to arrive at empirically adequate theories, which give a true description of the observable

11. In the same vein, see Josephson and Josephson (1994, ch. 1).

part of the world but not necessarily of the unobservable part. Note that this official statement of the opposition between realism and antirealism is entirely silent on the question of whether science is *able* to arrive at theories that are also true of the unobservable. However, antirealism would not be of much interest if it conceded that science does typically give us theories going beyond mere empirical adequacy but that nevertheless it is not *essential* to science to do so; the former is what we really care about. And indeed, in much of his work, van Fraassen has sought to undercut realist arguments to the effect that science is in the process of uncovering what lies beneath the phenomena. His attacks on abduction belong to this part of his work.

But if we reject ABD1 and adopt ABD2 or ABD3 in its stead, can abduction still serve realist purposes? With regard to ABD2, van Fraassen could ask why it should matter to the truth of a theory that it explains the evidence to *our* satisfaction. We might be way too undemanding. How to tell? And ABD3 leaves open the possibility that even our best-explaining theories are *very* far removed from the truth, which is incompatible with realism, at least in spirit. In response, it could be said that if ABD3 does not support full-blown realism, it still allows us to hold on to a position interestingly stronger than van Fraassen's brand of antirealism. For even if aiming at truth is too ambitious a goal, if ABD3 is a reliable rule of inference, we would be definitely underplaying our hand if we aimed no higher than at empirical adequacy. For then the goal that is minimally within the reach of science is to bring us ever closer to the truth, also on a theoretical level. (I am assuming here that in general scientists are able to think up hypotheses explanatorily superior to the ones already available.)

That is a weaker position than scientific realism as commonly conceived, which maintains that we have good reason to believe that contemporary scientific theories are approximately true and not only that they are closer to the truth than their predecessors were.[12] However, those realists not satisfied with the just-mentioned convergent realism should not despair if no stronger explication of the idea of abduction could be coherently maintained. As a moment's reflection shows, ABD3 licenses an inference to the unqualified truth of the *absolutely best* or *perfect* explanation. That is to say, if one is sure that however many other potential explanations for the data may have

12. This is rough. For a more careful statement of scientific realism, see Hoefer and Martí (2020).

gone unconsidered, none can equal the best of those that have been considered, then it makes no difference whether one applies ABD1 or ABD3. Consequently, if one is able to show that realism, as commonly conceived, is a perfect explanation for the observed predictive accuracy of science (i.e., that no unborn hypothesis can even equal realism as an explanation of that fact), then it suffices to show that ABD3 is a reliable rule of inference in order to defend the strong version of realism. I am skeptical that this can be accomplished (Douven, 1995). However, I would be content with a convergent realism as suggested here, as are others who consider themselves to be in the realist camp (e.g., Niiniluoto, 1999b; Kuipers, 2000). What is more, in chapter 8, I argue that we can probably do better than that, specifically that we are in a position or can move to a position from which we can ascertain the existence of many unobservables.

A more general criticism, which pertains to all these above explications of abduction as well as to their variants, is that they are stated in *categorical* terms. All these versions issue a yes-or-no verdict: they either do or do not license an inference, which for most authors means that one either is or is not warranted in accepting or believing a hypothesis, or believing that it is closer to the truth than its rivals, or believing that it is probably true, and so on.[13] That may suggest that abduction belongs to, in Richard Foley's (1992) terminology, the epistemology of (categorical) belief rather than the epistemology of degrees of belief. For many Bayesians, this alone would disqualify abduction because as they see it, the notion of categorical belief has no place in scientific philosophy; categorical belief talk is loose talk. And those Bayesians who *are* willing to condone the notion of categorical belief and so might be open to considering seriously a version of abduction cast in categorical terms do so only on the condition that the connection between categorical beliefs and probabilities or degrees of belief can be made formally precise—and the prospects of satisfying

13. For instance, according to Musgrave (1988, p. 239, italics omitted), "[i]t is reasonable to accept a satisfactory explanation of any fact, which is the best available explanation of that fact, as true." And Psillos holds that abduction "authorises the acceptance of a hypothesis *H*, on the basis that it is the best explanation of the evidence" (2004, p. 83, italics omitted), and that the guiding idea behind abduction is that "a hypothesis is accepted on the basis of a judgement that it best explains the available evidence" (2007, p. 442). Similarly, according to Greco (2015, p. 510), in abductive reasoning "one concludes that something is the case on the grounds that this best explains something else one believes to be the case."

that condition are bleak at best (Douven & Williamson, 2006; Douven & Rott, 2018; but cf. Chandler, 2013; Wenmackers, 2013; Leitgeb, 2017).[14]

I have two comments on this. First, I am not aware of an argument to the effect that advocates of abduction are committed to the epistemology of belief and so should reject out of hand any explication of the core idea of abduction in probabilistic terms. While such explications are sparse, they are not altogether absent from the literature. It was previously mentioned that van Fraassen, in the context of criticizing abduction, claimed that it had a chance of being tenable only if it could take the form of a probabilistic update rule, similar to Bayes's rule. He then went on to suggest the following version of abduction, which is of precisely that kind:[15]

EXPL Let $\mathcal{H} = \{H_i\}_{1 \leqslant i \leqslant n}$ be a set of mutually exclusive and jointly exhaustive hypotheses, and let Pr and Pr′ designate a person's degrees-of-belief function before and after learning evidence E, where it is assumed that $\Pr(E) > 0$. Then that person updates her degrees of belief abductively precisely if it holds for all j that

$$\Pr'(H_j) = \frac{\Pr(H_j)\Pr(E \mid H_j) + f(H_j, E, \mathcal{H})}{\sum_{k=1}^{n}\left(\Pr(H_k)\Pr(E \mid H_k) + f(H_k, E, \mathcal{H})\right)},$$

with

$$f(H_j, E, \mathcal{H}) = \begin{cases} c & \text{if } H_j \text{ best explains } E, \\ 0 & \text{otherwise,} \end{cases}$$

for some c: $0 < c < 1$.

14. What to make then of the widely shared intuition that categorical and graded belief do hang together closely? As Douven and Elqayam (2021) argue, it is possible to account for this intuition even if there is no *formal* connection between the two. Drawing on recent work on dual processing (e.g., Evans, 2007a, 2008, 2010a, 2010b), they show how categorical and graded belief can be *psychologically* connected in a way that explains the said intuition but that almost certainly cannot be captured in logico-mathematical terms. For related takes on the psychological dimension of the categorical-versus-graded-belief contrast, see Weisberg (2020) and Nagel (2021).

15. In the present context, it would make sense to call this rule "ABD4," but I have been calling it "EXPL" in various papers for about twenty years, and I have become somewhat attached to this label, which I is why I am using it here. Stephan Hartmann suggested, jokingly, that I call the principle "van Fraassen's principle," in analogy of van Fraassen's (1989) calling the Principal Principle "Miller's principle" to "honor" Miller's (1966) attempt (unsuccessful, as others soon discovered; e.g., Howson & Oddie, 1979) to refute the Principal Principle.

This looks more complicated than it is. It is helpful to think of updating via EXPL as consisting of three simple steps: in the first step, we update on the evidence via Bayes's rule; in the second, we add a bonus c to the post-Bayesian-update degree of belief assigned to the best-explaining hypothesis; and finally, we renormalize so that the degrees of belief assigned to the members of $\mathcal{H}$ sum to one in the end.

As stated here, EXPL needs some qualifications. What if two or more hypotheses are tied for explanatory bestness? The obvious solution in that case is to split the bonus evenly between or among those hypotheses. Furthermore, it could be said that, in most instances, we will judge explanation quality not just in light of the most recently acquired piece of evidence, but in light of the totality of our evidence; indeed, that might be what we *should* do (Carnap, 1947b; Good, 1967). If so, the bonus should go to the hypothesis (hypotheses) that best explains (explain) our new evidence E together with whatever evidence we already possessed. And finally, it is noted that EXPL really provides a continuum of abductive rules, one for each value of c. Van Fraassen uses the specific instance of EXPL with $c = .1$ to argue that abduction is incoherent—a claim with which we take issue in chapter 4—though the exact value of c is inessential to his argument; that is to say, any other value would have done just as well for that argument (except the value 0, which, if it were permitted, would effectively reduce EXPL to Bayes's rule). We shall encounter instances of EXPL on numerous later occasions, and comments on van Fraassen's proposal are postponed to those occasions. For now, the important point is that abduction is not tied to the notion of categorical belief, but can take a probabilistic form that allows giving some weight to judgments of explanation quality in the process of adjusting one's degrees of belief in the light of new evidence.[16]

The second comment is that although most (but not all) attention in the remainder of the book goes to probabilistic versions of abduction, this should not be taken to indicate that I believe that there is something wrong with categorical versions or that such versions have only secondary importance. To the contrary (and to repeat), a central message of the book is that it is fine to regard abduction as a slogan that, upon use, is to be explicated in a way that

16. Note, incidentally, that using an instance of EXPL would not be inconsistent with bringing explanatory considerations to bear on the assignment of priors and / or likelihoods, in the way suggested by Lipton and others. For reasons given previously, it may be hard to make this suggestion precise, but that is another matter.

depends on the context. Given that we sometimes think and talk about what we believe (categorically) and sometimes think and talk about how confident we are in this or that, it seems entirely reasonable to suppose that we will sometimes want to have at hand a categorical explication of abduction and sometimes a probabilistic one (if we are open to reasoning abductively at all). Nonetheless we will focus mainly on probabilistic versions because the main criticisms of abduction have come from proponents of Bayesianism, and I see no other way to counter those criticisms effectively than by relying on versions of abduction that are somewhat similar to Bayes's rule.

I end this section with two more general remarks. First, it merits emphasis that there is no need to choose between abduction in the Peircean sense and abduction in the more modern sense. One can consistently commit to a role for explanatory considerations in the process of generating theories (Peircean abduction) and a role for such considerations in the process of trying to justify a theory. Indeed, this is what Brian Haig (2005a) does in his abductive theory of method that was mentioned previously (see ch. 1, footnote 9).[17] In his view (Haig, 2005b), for instance, exploratory factor analysis is an abductive technique (abductive in the Peircean sense) that facilitates the generation of theories by suggesting a number of so-called latent variables that together might be sufficient to explain the phenomena at hand (the "manifest" variables). But he also sees a key role for explanatory reasoning later in the scientific process, when a number of potential explanations have been generated and we are to choose in which of those to invest our confidence. In particular, he sees a key role here for the concept of explanatory coherence, as proposed by Paul Thagard (1989, 2000), which allows one to rank theories according to how well they do with respect to criteria of explanatory goodness.

This brings me to the second remark, which concerns those criteria in relation to the idea of explanatory bestness. In section 1.2.1, I have stated that I aim to remain noncommittal regarding explanation. For the purposes of this book, it is immaterial whether explanations are best thought of as arguments of a specific type (in the manner of Hempel, 1965), as unifying different and seemingly unrelated pieces of evidence (Friedman, 1974; Kitcher, 1981), or as revealing the cause or causes of the explanandum (Salmon, 1984).

17. Niiniluoto (2018) also sees a role for abduction both in the Peircean and in the modern sense, as does Schurz (2008a), as mentioned in footnote 8 in this chapter. However, in Schurz's view the role that abduction plays in the context of justification is only minor; its primary role is in the context of discovery.

Indeed, according to some authors, a plurality of concepts of explanation are needed to understand how that notion functions in science and in everyday life, and it is a mistake to hold that there is exactly one correct account of explanation (Woody, 2015; Lange, 2016; Colombo, 2017; Khalifa, Doble, & Millson, 2020; Schupbach and Sprenger, in their 2011 paper, are also open to this possibility). Here it suffices to note that all these positions are compatible with the widely held view that explanation quality is a matter of scoring high on the so-called theoretical virtues, which include simplicity, scope, fruitfulness, internal coherence, coherence with background knowledge, and mathematical elegance.[18]

It has often been remarked that these virtues may conflict with each other (e.g., McMullin, 1996; Niiniluoto, 1999a; Schurz, 2008a) and then will have to be balanced against one another. Exactly what relative importance they are to be given may differ per context of application, and even how they are to be interpreted may vary (Longino, 1995). For instance, there is a plethora of measures of coherence (see ch. 1, footnote 12 for references), and here, too, it has been argued that we should be open to the possibility that there is not one best such measure but that different measures may be called upon on different occasions (Schippers, 2014). Note that this view on explanation quality and its relation to abduction dovetails nicely with the general message that abduction is best seen as a slogan that needs to be precisified on a per-case

18. See, for instance, McMullin (1996), Papineau (1997, p. 9), and Niiniluoto (1999a). Different authors may weigh the individual virtues differently. For instance, Einstein (1950, p. 62) thought that especially *simplicity* should guide theory choice, because "we are justified in feeling that Nature is the realisation of what is mathematically simplest." Kornblith (1993, p. 33) highlights the importance of *scope*—the ability of a theory to account for phenomena that would seem to have nothing to do with each other—when he explains why Brownian motion and Gay-Lussac's law of combination of gases together provide such strong evidence for the existence of atoms: "Although either phenomenon, by itself, might perhaps be explained by some as yet unthought hypothesis, the fact that two entirely disparate phenomena are naturally explained by one and the same hypothesis lends dramatic support to the atomic theory." And Zahar (1973) argues that Einstein's Special Theory of Relativity superseded the Lorentz–Fitzgerald–Poincaré version of the ether theory because it was more *fruitful* in that it suggested various new theories to be tested. Also recall Whewell's (1847) idea that a *consilience of inductions*—which combines elements of simplicity and scope—can provide a theory with a "stamp of truth." For empirical work on the theoretical virtues, see Lombrozo (2007); Khemlani, Sussman, and Oppenheimer (2011); Colombo, Bucher, and Sprenger (2017); Johnson, Valenti, and Keil (2019); and Shimojo, Miwa, and Terai (2020).

basis: the precisification may pertain as well to the criteria by which we judge explanations.

2.3 Abduction and Underdetermination

Sometimes all it takes to settle a dispute is to obtain more data. Suppose that we disagree about whether Bob is at his desk. We walk over to his office, and there he is, sitting at his desk. Done! Obtaining more data is not always so easy. Newton and his friends had to wait a little before the data came in that confirmed that the Dutch had won. When Newton concluded that the Dutch had won, the data alone were not enough to settle the issue. Sometimes obtaining data that would settle an issue may be out of reach, even forever. If our dispute concerns the question of whether Bob is a conscious being, someone with an inner life, or is only *behaving* as if he had an inner life, then there appears to be no equivalent to walking over to his office or waiting for an eyewitness report; no amount of data may be enough to show who of us is right, at least not conclusively, and not on its own. Similarly if we are wondering whether we are actually *seeing* Bob at his desk or instead are *hallucinating* him to be there; or more generally still, if we cannot agree on what the data are that might help us settle our dispute.

All these situations exhibit a phenomenon known as "underdetermination," meaning that in these situations, our data is compatible with various hypotheses of interest being true. That predicament will often only be temporary, as in the case of Newton or even more so in the case in which all we need to do to settle our dispute is walk over to Bob's office. But philosophers have come up with cases in which, according to them, the predicament is permanent: no amount of data could ever warrant a choice for one or the other hypothesis.

So far in this chapter we have looked at a number of uncontroversial applications of abduction, and we have also reviewed various proposals to explicate abduction. This section discusses the most paradigmatic use of abduction in philosophy, to wit, as a tool to break the deadlock in alleged cases of underdetermination. Advocates of abduction believe that, sometimes, where the data cannot settle an issue, the data together with an appeal to explanatory considerations *can*. That is so whether the underdetermination is thought to be only temporary or permanent, although the underdetermination cases that have been most widely discussed in the philosophical literature are invariably

of the permanent variety. Accordingly, the uses of abduction that have been most widely discussed also concern precisely such cases.

It will therefore be useful to describe underdetermination cases generically and to point at reasons for why we should care about them. Arguably, applications of abduction in cases of permanent underdetermination are the more interesting ones. It is thus worth also discussing why we should believe that there are such permanent cases of underdetermination. Finally, we want to be more explicit about how abduction may facilitate to resolve them.

2.3.1 What Does It Mean to Say That One Thing Is Underdetermined by Another?

Underdetermination is a central issue in various areas of analytic philosophy. Underdetermination claims are often adduced to argue that our epistemic position vis-à-vis a given part of reality is less impressive than we would have hoped or thought it was, and in any event there is usually a lot at stake in arguments concerning some underdetermination claim(s). Some well-known philosophical debates can be regarded as turning, at bottom, on whether or not a given underdetermination claim must be accepted and concomitantly on whether or not we must resign ourselves to some modest (typically, very modest) epistemic position concerning the part of reality at issue.

For instance, in the philosophy of science one frequently encounters claims to the effect that a particular theory is (or is not) underdetermined by the evidence, or even that all scientific theories (or at least all those belonging to a certain interesting class of theories) are underdetermined by the evidence and even by all evidence that we might ideally possess. What is typically, and roughly, meant by such claims is that having *all* the available evidence will still not allow us to determine the truth-value of the theory, respectively of any theory (or any theory belonging to some designated class). To make this both more precise and more general, we can let underdetermination be a relationship between distinct classes of propositions and hold for different combinations of "know" and "justifiedly believe."[19] We might for instance say that one class of propositions, C_1, ⟨know, know⟩-underdetermines another class of propositions, C_2, if and only if knowing every member of C_1 is not

19. To make this *entirely* general, one might even consider any combination of epistemic attitudes, though I doubt that others than the just-mentioned ones will yield philosophically interesting underdetermination claims.

enough to know any member of C_2. Similarly, C_1 ⟨know, justifiedly believe⟩-underdetermines C_2 if and only if knowing every member of C_1 is not enough even to be justified in believing any member of C_2.

Although not usually stated in this way, most underdetermination claims encountered in the philosophy of science seem to be about ⟨know, justifiedly believe⟩-underdetermination of some given class of rival scientific theories by a class of propositions expressing (relevant) evidence—and not only those in the philosophy of science. One of the most central underdetermination claims in epistemology can be rendered as follows: The class of propositions expressing all your sense data throughout your entire life, as well as those you might have had at a certain moment in your life, ⟨know, justifiedly believe⟩-underdetermines the class of propositions {You are a brain in a vat, You are an embodied brain}. And a well-known underdetermination claim from the philosophy of mind is that the class of truths about a person's behavior ⟨know, justifiedly believe⟩-underdetermines, among many others, the class of hypotheses {The person has an inner life, The person does not have an inner life}. Unless stated otherwise, by "underdetermination" is meant ⟨know, justifiedly believe⟩-underdetermination.

2.3.2 *Why Is Underdetermination Philosophically Interesting?*

Nothing philosophically really interesting follows from an underdetermination claim in itself. Suppose that C_1 underdetermines C_2, so that knowing every member of C_1 will not justify us in believing any member of C_2. That does not mean that justifiedly believing in members of C_2 is impossible. Perhaps we simply do not need to know any member of C_1 in order to come to have justified beliefs in, or even knowledge of, the members of C_2, for instance, because we have direct epistemic access to the latter or epistemic access via some third class of propositions, C_3.

In interesting underdetermination claims, there is always some allegedly important epistemic distinction between the two classes of propositions referred to in the claim. To one class we are supposed to have a fairly direct cognitive access or at least a direct access given certain idealizing assumptions which in the context of the discussion are mostly deemed innocuous (or, at any rate, permissible). To the other class we at least *prima facie* seem to have cognitive access, if at all, only via the former, that is, it seems that the

propositions in the latter class can be known or at least justifiedly believed, if at all, only because we can know the propositions in the former.

Arguments involving underdetermination claims come in two main varieties, which one might call the "modus tollens type" and the "modus ponens type." Arguments of the former type are meant to establish either the existence of a class of data to which we have some kind of cognitive access—although it is not obvious that we have that access—or the existence of one or more rules of inference that we justifiably rely on, although it is not obvious that we rely on that / those rule(s), or at least it is not obvious that we are justified in doing so. The modus ponens variety is meant to establish some form of skepticism.

In arguments of the modus tollens type, it is standardly presented as a given that we know / justifiedly believe all or at least some members of a class of propositions, C_1, and that there is at least some initial plausibility to the thought that our knowledge of / justified belief in these propositions depends entirely on our knowledge of (some of) the members of another class of propositions, C_2. But, it is then claimed, C_2 underdetermines C_1. The point then typically argued for is *either* that our knowledge of / justified belief in the elements of C_1 (insofar as we have it) must, appearances to the contrary notwithstanding, depend on more than just our knowledge of (some of) the members of C_2 (if it depends on that at all) *or* that there must be other rules than those most obviously available to us (like the rules of first-order logic) by dint of which we can come to have knowledge of / justified beliefs in (some of) the elements of C_1 on the basis of our knowledge of the members of C_2. This is almost invariably accompanied by some proposal as to what the something more, or the other rule(s), could be.

We find an underdetermination claim of this type in Lewis (1986, p. 107), when he discusses the requirements for a functional theory of mental content and argues that in order to be able to assign content to functional states, we must rely on principles of fit, roughly to the effect that the assignment of contents to a person should tend to make her behavior come out as serving her desires according to her beliefs. But, writes Lewis (1986, p. 107),

> principles of fit can be expected to underdetermine the assignment of content very badly. Given a fitting assignment, we can scramble it into an equally fitting but perverse alternative assignment. Therefore a theory of content needs a second part: as well as principles of fit, we need "principles of humanity," which create a presumption in favour of some sorts of content and against others.

This is a kind of attempted transcendental deduction of the existence of principles that we use in interpreting each other, where—note—it is taken as a given that interpretations (or assignments of contents) are in fact not generally underdetermined.

The other type of argument involving underdetermination claims is more common in philosophy. The basic structure of arguments of this type is the following: We can know / justifiedly believe the members of a class of propositions, C_1, if at all, only by knowing the propositions in another class, C_2; but C_2 underdetermines C_1; hence we cannot know any member of C_1. Well-known examples of this type are the Cartesian argument for external-world skepticism and various arguments for more restrictive forms of skepticism, such as skepticism about other minds. To this type also belongs one of the main arguments—if not *the* main argument—for scientific antirealism, the position in the philosophy of science that counsels agnosticism as the proper epistemic attitude vis-à-vis scientific theories, because, it is claimed, scientific theories are underdetermined by the evidence.

2.3.3 Why Believe Underdetermination Claims?

While underdetermination claims are found in many areas of philosophy, the most detailed discussion of underdetermination has taken place in the philosophy of science, specifically in the debate between scientific realists and scientific antirealists. Central to this debate is whether science is in the business of uncovering the hidden structure of reality, laying bare the unobservable goings-on responsible for the observable goings-on, as scientific realists hold, or whether it can give us, at best, theories that are empirically adequate, meaning that they correctly represent the observable part of the world but not necessarily the unobservable part, as scientific antirealists and in particular so-called constructive empiricists (van Fraassen, 1980) hold.[20] A main motivation for scientific antirealists is that they believe theory choice in science to be underdetermined.[21] Precisely because underdetermination has received so much detailed attention in the philosophy of science literature,

20. At least this is the interesting take on what is at stake in this debate. The boring take on it is that the debate is strictly about the goal of science and not about what science may be able to achieve (see p. 49).

21. Underdetermination is not the only motivation for scientific antirealism. For a systematic discussion of the various reasons scientific antirealists have advanced for their position, see Psillos (1999).

we focus on underdetermination in science for a while. As subsequently discussed, much of that discussion carries over to underdetermination arguments in other areas.

The standard antirealist argument for the thesis that scientific theories are underdetermined by the evidence involves two premises. The first is the following:

EE For each scientific theory there are empirically equivalent rivals,

where an empirically equivalent rival to a theory is a contrary theory (a theory inconsistent with it) that at least in light of the data alone—any possible data—will necessarily be accorded the same confirmation-theoretic status. Naturally one could consider weaker versions of EE ("for most theories . . . ," "for many theories . . . ," "for some theories . . . ," "for all theories with such-and-such features . . . ," and so on), which in combination with the subsequent premise KE, would all seem to yield somewhat different versions of scientific antirealism, but for simplicity we stick to EE here.

The important thing to note is that if EE is correct, then no matter how many empirical tests a theory has already passed, this success cannot on its own be taken as an indication that the theory is true, for each of its empirically equivalent rivals will or would pass the same tests just as successfully. Thus, unless the data refute a theory, no amount of them suffices to determine its truth-value.

Although EE does not by itself yield any antirealist conclusions, it does do so together with the following premise sometimes called "Knowledge Empiricism":

KE If the data alone do not suffice to determine a theory's truth-value, then nothing does.

Indeed, from EE and KE it follows directly that the truth-value of any scientific theory must forever remain beyond our ken. Crucially, in the context of our present discussion, notice that KE says, in effect, that explanatory power is at most of pragmatic value and has no epistemic significance.

Arguments for EE either extrapolate from supposed historical cases of empirical equivalence or try to prove formally the existence of empirically equivalent rivals to any scientific theory. As to the former, we are often pointed to the empirical equivalence of the Special Theory of Relativity and the ether

theory in the Lorentz–Fitzgerald–Poincaré version (briefly mentioned previously) and respectively that of standard quantum mechanics and Bohmian mechanics. As to the latter, John Earman (1993) has proposed various plausible formalizations of the notion of empirical equivalence and used them to prove some propositions that can all be regarded as establishing interesting versions of EE (see Douven & Horsten, 1998, for discussion). Similar results have been obtained by other authors.

Arguments for KE typically try to raise doubts about the truth-conduciveness of any *prima facie* reasonable candidate criterion for theory choice beyond conformity with the data. For instance, antirealists have argued that there is no a priori reason to believe that reality is simple rather than complex.[22] A further point they have raised is that even if it be granted that the world is simple in *some* sense of that word, it still need not be simple in the parochial sense that its nature or structure is easy to grasp for creatures with *our* cognitive capacities (van Fraassen, 1980, p. 90).

2.3.4 What Can One Say in Response to Underdetermination Claims?

Not many philosophers are happy to accept scientific antirealism. It is not surprising, then, that several responses to the preceding antirealist argument for underdetermination are to be found in the literature. It merits remark that most of the responses now to be discussed, which all concern the antirealist argument for underdetermination, have close parallels in debates about underdetermination in other parts of philosophy.

One type of response against the antirealist argument denies EE, or at least maintains that currently we have no reason to believe that there exist (interesting) empirically equivalent rivals to any scientific theory. We might regard as an early token of this the logical positivists' argument that apparently empirically indistinguishable rivals are really just notational variants of one another. But their response was based on a verificationist view of meaning that nowadays is almost universally regarded a failure.

It seems a better strategy to tackle directly, or at least to try to raise doubts about, the arguments that have been given in support of EE. For instance, it may be pointed out that historical evidence for the thesis is very sparse. Advocates of the thesis point time and again to the two historical cases mentioned in the previous section. Patently, however, having two actual cases of

22. Some realists agree. See, for instance, Niiniluoto (1999b, pp. 183–184).

empirical equivalence in science seems hardly enough to support the claim that all scientific theories have empirically equivalent rivals nor even that a substantial number of theories have such rivals. Yet that is about all the historical evidence that we have ever been given! This objection might be countered by saying that the sparseness of actual examples of empirically equivalent rivals is explained by the fact that in scientific practice it is typically hard enough to come up with even one theory that fits the data and is also consistent with accepted background theories, let alone with a number of such theories. But one can also, and perhaps with more right, draw an altogether different moral from this fact, as Gerard 't Hooft (1994, p. 27) for instance does about the fundamental laws of physics when he states that "[t]he requirement that [the fundamental laws of physics] must agree with the very restrictive postulates of both quantum mechanics and general relativity has up to now proved so difficult to realize in any physical model that one is tempted to suspect that not more than one model will exist which agrees with all this."

In an influential paper, Larry Laudan and Jarrett Leplin (1991) have argued that in fact no number of actual examples of allegedly empirically equivalent theories can support EE, because, they contend, such theories may really be only *temporarily* indistinguishable by the data.[23] They do not mean to suggest that cases of theories that happen to be indistinguishable in light of the data that we currently have manifest only the kind of underdetermination we found in the case of Newton's conclusion about the Dutch victory and so are unable to support EE. Rather their point is, first, that our conception of data may change over time and, in particular, that the line between the observable and the unobservable may shift due to new technological advances; and second, as is widely acknowledged, that theories have observational consequences only when conjoined with so-called auxiliaries, and over time we may come to hold different views about the hypotheses that we deem eligible to figure as auxiliaries in the derivation of observational consequences from theories, so that over time theories may come to have different observational consequences.[24]

Still, the antirealist would seem in her right to insist on a conception of data, or at least of the observable, that is *not* susceptible to change over time.

23. See also Leplin (1997), and see Douven (2000) for discussion.

24. Fodor and Lepore (1992) make basically the same point when they argue that "our knowledge of confirmation relations is a posteriori," so that there is no way to decide a priori which empirical beliefs are relevant to the confirmation of a given hypothesis.

Van Fraassen (1980) has argued that there is an epistemically significant distinction between claims whose truth-value can be ascertained by observation with the naked eye and ones whose truth-value cannot thus be ascertained. And, almost by definition, no technological advances are going to affect that distinction. As for the variability of auxiliaries, Richard Boyd (1984) may be right that advocates of EE can successfully respond to this point by reformulating the thesis in terms of "total sciences," which include both theories and auxiliaries.

In their paper Laudan and Leplin further complain that while many philosophers of science seem to believe that there exists some algorithm for generating empirically equivalent rivals to any given theory, such an algorithm is nowhere to be found in the literature. This seems right. At the same time one wonders why such an algorithm should be called for. It seems that a proof of EE—whether or not that shows how *effectively* to construct empirically equivalent rivals—would offer the advocates of EE all that they could wish for. And as we saw previously, such proofs have for instance been given by Earman.[25]

It must be admitted, however, that at least the existing proofs of EE appear to rely on assumptions that are open to dispute. For instance, the proofs in Earman (1993) crucially depend on the assumption that scientific theories can be formulated in first-order languages, which may well be false. It may in effect be very hard to prove EE in a way that could suit the antirealist's needs. The main stumbling block here is that there seems to be no purely logical characterization of the notion of empirical equivalence. Of course, it is not uncommon to find empirically equivalent rivals defined as contraries that have the same logical consequences in the observational part of some designated vocabulary—which *is* a logical characterization. But while this characterization of empirical equivalence may be perfectly all right if hypothetico-deductivism is assumed, that confirmation theory is not part of the current orthodoxy. Indeed, it would be a mistake to think that the notion of empirical equivalence can be defined without (at least implicit) reference to a confirmation theory.

To buttress this point, it is good to briefly consider the underdetermination problem from a Bayesian perspective. We subsequently go deeper into

25. It should be noted that Earman's paper appeared after the publication of Laudan and Leplin's paper; it was in effect at least partly meant as a response to that paper.

the details of Bayesian confirmation theory; here an informal understanding suffices to appreciate that, for Bayesians, EE would be unable to establish any (interesting) underdetermination claim if empirical equivalence were defined in the way just suggested, because that two or more theories have the same observational consequences does not imply that they bestow the same likelihoods on all evidence statements that they do not entail (but are consistent with). And the latter are, from a Bayesian point of view, just as relevant to determining a theory's confirmation-theoretic status as are its observational consequences. In fact, given a purely subjective version of Bayesian confirmation theory, which imposes no constraints on rational degrees of belief beyond the axioms of probability theory, no underdetermination claim would seem to follow from EE even if empirically equivalent theories were defined to be theories that bestow the same likelihoods on all evidence statements. That confirmation theory would for instance allow one to assign a prior probability of 0 to all empirically equivalent rivals of a given theory, so that unless the data refute it, the theory may eventually come to have probability 1 (which is typically thought to suffice for justified credibility). On the other hand, for versions of Bayesian confirmation theory that are only slightly stronger, if the existence of empirically equivalent rivals in the just-defined sense is assumed, then interesting underdetermination claims *can* be derived. Suppose for instance we add to the subjective theory the apparently still quite weak principle that, given any set of empirically equivalent theories, there is no *unique* element of that set that receives highest prior probability. Then it is a direct consequence of Bayes's theorem that at no point in time will any of the theories have a unique highest posterior probability, no matter what evidence one may come to possess.

So, whether two theories will have the same confirmation-theoretic status given any amount of evidence depends at least in part on the confirmation theory that is being assumed, and thus proofs of EE may be of limited interest at best: even if it can be shown that all theories have empirically equivalent rivals given our current best confirmation theory, there is no guarantee that they will have these rivals given some still-to-be-developed confirmation theory that we may come to prefer one day.

However, it may be questioned whether EE is really all that important for antirealist purposes. For instance, one may wonder why an argument for underdetermination must be based on the claim that there *actually* exist empirically equivalent rivals to any scientific theory. Is it not enough to

observe that any scientific theory *might* have such rivals? But first, whether the mere possibility that these rivals exist really undercuts any confirmation that we might otherwise have for a given scientific theory will (again) depend on one's confirmation theory. Furthermore, most philosophers of science would regard a position based on the assumption of the mere possibility of empirically equivalent rivals as of academic interest at best and not as a live option. That may explain why no scientific antirealist has tried to argue for underdetermination along these lines. More generally, there are to the best of my knowledge simply *no* underdetermination claims, in philosophy or elsewhere, based on the mere possibility that a hypothesis of interest has an empirically equivalent rival.

There may be a better way to argue that even if EE cannot be maintained, an interesting argument for underdetermination can go through. Consider that it may be rather simple to devise—on paper, that is—an experiment that would enable us to distinguish between two theories whereas the experiment is practically impossible to carry out. This possibility is anything but academic. For instance, David Atkinson (2003, p. 216) argues that while string theory is testable in principle, in order to really test it "one would have to produce energies that are ten to the power sixteen . . . times higher than those that [the biggest particle accelerator] will produce in 2005." Indeed, he concludes that "[i]t seems safe to say that we will never be able to produce energies anywhere near this value, and that string theory can never be confronted with the crucial test of experiment" (p. 216). One does not have to agree with Atkinson's pessimism regarding the testability of string theory in order to appreciate how similar practical considerations may apply quite generally in science. And if every scientific theory should have rivals that are indistinguishable from it by any evidence we might be practically able to obtain, even if the theories *are* distinguishable by evidence that we could obtain in principle, that would seem to serve an underdetermination argument for scientific antirealism no less than does EE. Because every empirically equivalent rival to a theory is a rival that is indistinguishable from it given the evidence that could practically be obtained, but not vice versa, the claim that every theory has rivals indistinguishable from it by that sort of evidence is patently weaker than EE.

The other main type of response to the argument from underdetermination, a response that I believe to be more promising than the attack on EE, is of course to deny KE. KE is a better target anyhow if we are interested in underdetermination arguments in general and not just those concerning

scientific theories. After all, whereas it may be exceedingly difficult to come up with an empirically equivalent rival to a given *scientific* hypothesis, the bar is much lower when it comes to thinking up empirically equivalent rivals to the external-world hypothesis or the hypothesis that other people are conscious beings. References to hallucinations, dreams, envatted brains, zombies, and the like, have been generally found sufficient to raise doubts concerning our epistemic standing vis-à-vis the aforementioned hypotheses. Here, too, it has been questioned whether the rival hypotheses are really empirically equivalent to their nonskeptical counterparts (see, e.g., Putnam, 1994). Again, however, we can grant that these are cases of genuine empirical equivalence, and still no skeptical conclusion follows without KE. Thus, if we find grounds for gainsaying KE, that should help us across the board, both in the scientific and in the nonscientific domain.

Denials of KE nowadays mostly take the form of an attempt to defend abduction, which is also the approach taken in this book. If we accord confirmation-theoretic import to explanatory force, then if of two theories that both conform to the data one better explains those data than the other, that speaks epistemically in favor of the former. So, can we defend abduction? This question occupies us in much of the rest of the book. But before starting the defense of abduction, I go in some detail into what is known about how people actually rely on explanatory considerations in their reasoning. Some of those results serve as inspiration, further on, for formulating alternatives to the previously stated rule EXPL, and those rules will be instrumental in my defense of abduction.

To end this chapter, I should note that there is a third type of response to underdetermination—heralded by some of the American pragmatist philosophers, though no longer advocated by anyone—that could be viewed as an independent response but also as constituting an easy defense of abduction. According to this response, KE is false because theoretical virtues such as simplicity, scope, and more generally explanatory power, are truth-conducive *by definition*: possessing those virtues is simply constitutive of being true, together of course with being in accordance with the data. This response immediately raises a question: what if two empirically equivalent theories that are in accordance with the data do equally well in light of the theoretical virtues, too? Surely there is no guarantee that this will *never* happen. Here the pragmatist answer is that even though such theories may appear to be rivals, they are not really. Rather they are different, equally legitimate, conceptualiza-

tions of reality, and may both be true ("true in their conceptual schemes," as it is often put). In this vein, W. V. O. Quine (1992, p. 100) has suggested that one may "oscillate" between the different conceptualizations "for the sake of added perspective from which to triangulate on problems." Very similar passages are to be found in the writings of Putnam after his conversion to pragmatic realism (see, for example, his 1981; see also Decock & Douven, 2012). One reason why this type of response is no longer popular today is that most believe it to share all the problems that beset pragmatist accounts of truth and language generally (see, e.g., Schmitt, 1995). Like these accounts, the pragmatic response to underdetermination seems to issue in what many think is a self-refuting relativism. Recent work on concepts, in particular on the *rationality* of concepts, may come to rescue here. It has been argued that some conceptual schemes are more rational than others (Douven, 2020d; Douven & Gärdenfors, 2020), where however—in line with a conception of rationality to be endorsed subsequently (chs. 6.2 and 7)—there is no universally rational conceptual scheme, but rather which such scheme to use depends on context. But this is not a line of argumentation pursued in the present work.