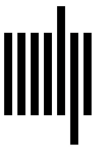
The Art of Abduction

By: Igor Douven

Citation:

The Art of Abduction
By: Igor Douven
DOI: 10.7551/mitpress/14179.001.0001
ISBN (electronic): 9780262369923
Publisher: The MIT Press
Published: 2022



The MIT Press

1

Introduction

> [T]he wisdom we mortals possess does not merely consist of remembering things past and apprehending the present, but on the basis of these two activities being able to predict the future, which is considered by serious men to be the highest form of human intelligence.
>
> —Giovanni Boccaccio, *The Decameron*

1.1 Why You Should Want to Know About Abduction

If you are consulting this book with malicious intent, looking for ways to expand your criminal pursuits, I encourage you to stop reading and to better your life. However, you may instead be a researcher—a cognitive scientist, a psychologist, a linguist, or a philosopher—who is interested in a type of inference that gives pride of place to explanatory considerations and that makes judgments on the quality of explanations and ultimately on explanatory superiority, which are crucial to what to believe and how to change our beliefs. You may have heard some good things about this type of inference but also, more likely, some bad things, and you may be wondering what to make of those things or how to reconcile them. In that case, I encourage you to read on.

If you are a psychologist, I encourage you to read on because there is accumulated evidence that abduction and, more generally, explanatory considerations play a key role in a great number of high-level cognitive processes, including categorization (e.g., Williams & Lombrozo, 2010, 2013; Edwards et al., 2019; Vasilyeva & Lombrozo, 2020), generalization (e.g., Lombrozo & Gwynne, 2014), learning (e.g., Chi et al., 1994; Baillargeon & DeJong, 2017;

Rittle-Johnson & Loehr, 2017; Walker et al., 2017; Ruggeri, Xu, & Lombrozo, 2019), understanding (Keil, 2006; Legare & Lombrozo, 2014; Walker & Lombrozo, 2017), interpreting behavior (Jern, Derrow-Pinion, & Piergiovanni, 2021), semantic processing (Krzyżanowska, 2015; Douven, 2016a; Douven et al., 2018, 2020; Mirabile & Douven, 2020; Douven, Elqayam, Mirabile, in press; Stewart et al., 2021; Rostworowski, Pietrulewicz, & Będkowski, in press), and reasoning and belief change (Pennington & Hastie, 1988, 1992, 1993; Douven & Schupbach, 2015a, 2015b; Johnston et al., 2017; ch. 3 examines in detail experimental evidence showing how pivotal explanation is in the realm of reasoning and belief change).

It has been suggested that abduction is also operative at a more basic level of cognitive functioning. Hermann von Helmholtz (1867, section 26) famously argued that perception is not the passive reception of sensory input that it might naïvely be thought to be but that it involves active, albeit typically unconscious, inference: A tower that we see in the distance projects only a tiny image on our retinas. If we perceive the tower as tall, that is the result of an inference that we make on the basis of the retinal image and an estimate (based on previous experience) of the distance between us and the tower. In this connection, Helmholtz (1867, p. 447ff) refers to John Stuart Mill's work on inductive inference. That may already be an implicit appeal to abduction, if Gilbert Harman (1965, 1968) is right that induction is a special kind of abduction (see also Thagard, 1978; Lycan, 1988; Weintraub, 2013). Leaving induction and its relation to abduction to the side for the moment, note passages in Helmholtz (1867) that suggest that he saw a clear connection between perception and explanatory reasoning. For instance, he attributes the occurrence of perceptual ambiguities—"when various . . . interpretations of an impression are possible"—to the fact that a person may vacillate between different explanations of the impression, or that different persons may settle for different explanations (Helmholtz, 1867, p. 440).

Helmholtz strongly influenced Irvin Rock's work on the relation between perception and inference, which is more explicit in relating perception directly to explanation. Rock (1983) spends several chapters discussing how explanatory reasoning helps to resolve cases of perceptual ambiguity. In chapter 6 of that book, he advances the hypothesis that what we perceive is guided by explanatory reasoning. More exactly, his claim—buttressed by references to the outcomes of a great number of experiments that Rock and others conducted

Figure 1.1: Illusion of a transparent square overlaying a tiled square (*left*), which according to Rock supports the hypothesis that explanation-guided inference is involved in perception; the illusion is disrupted when no easy explanation of the stimulus in terms of transparency is available (*right*). (*Source*: Rock, 1983, p. 140.)

over the years—is that the percept that we settle on is the one that offers the best explanation of the stimulus.

Consider the left panel of figure 1.1, which is reproduced from Rock (1983, p. 140). People tend to perceive this as a tiled square overlaid by a smaller square that is transparent, because—Rock argues—this offers the most elegant explanation of the stimulus. An alternative is to see the square as consisting of a number of unrelated figural fragments, but that explanation is not nearly as elegant. Note also how easily the illusion is disrupted by complicating the explanation made in terms of an overlaying transparent square, as in the right panel of figure 1.1. When we look at the left panel, we imagine that the transparent square is tinted and that the tint has a *consistent* effect on how we perceive the different colors of the tiles as seen through the overlying square. For the right panel, to maintain the explanation of looking through a transparent square for the right panel, we would have to assume *either* that the square is itself divided into four parts, each part tinted differently, and that the overlying square is rotated relative to the larger square exactly so that the different tints are aligned with the tiles, *or* that the underlying square does not consist of four tiles, two of which are black and two of which are light gray, but either is tiled in a more complicated way or has four tiles that are colored in more complicated ways, *or* make still more convoluted assumptions. That is obviously too much, which is why the transparency illusion does not, or not easily, arise for the right panel.

Figure 1.2: We tend to see this as a bas-relief footprint illuminated from above, but in reality it is an indented footprint illuminated from below. (Photo courtesy of Manuel Cazzaniga.)

Rock's book describes many more well-documented effects concerning perception, which he takes to indicate that in general we perceive what best explains the stimuli. These effects include stroboscopic motion, in which a stationary object *seems* to move, for instance, because different parts of it are illuminated shortly after each other (Rock & Ebenholtz, 1962; Sigman & Rock, 1974), anorthoscopic form perception, in which we perceive an object by viewing only small parts of it at a time (Rock & Halper, 1969; Rock & Gilchrist, 1975), and the kinetic depth effect, which (among other things) lets us interpret projections of three-dimensional objects onto a screen as shadows of three-dimensional objects rather than as transformations of two-dimensional objects (Rock & Smith, 1981). Rock (1983, ch. 6) canvasses various other potential explanations for these and other effects (e.g., empiricism, according to which interpretation is a mere matter of frequency of previous exposure, and the constancy view, which lets us prefer the interpretation that keeps the percept constant through changes of viewing angle or location), but he argues that his proposal offers a better explanation for those effects than do the competing accounts. Thus, he argues abductively for an abductive theory of perception.

It is also not hard to understand, abductively, why we tend to perceive figure 1.2 as a bas-relief (i.e., convex) footprint, although in reality the photo shows a normal concave footprint lit from below instead of, as is normally the case, from above. Judgments of explanation quality are typically relative to background assumptions, and precisely because objects are normally illuminated from above, the stimulus presented by figure 1.2 is best explained by

assuming that the footprint rises from the sand. But that judgment is easily overruled: if people are given clues that there might be a light source *below* the footprint, they no longer perceive the footprint as rising out of the sand but instead see it as indented (Morgenstern, Murray, & Harris, 2011). It is readily appreciated how, with this kind of clue, the assumption that we see a bas-relief footprint will probably no longer be the best explanation of the stimulus.

This explanation of the bas-relief illusion is particularly interesting because the illusion has often been alleged to provide evidence in favor of the so-called Bayesian brain hypothesis, according to which the human nervous system operates largely as if it obeyed Bayesian prescriptions. From this perspective, the illusion highlights that we always start from prior probabilities or simply priors: when the brain computes the percept from the stimulus, it assigns a high prior to the light-from-above hypothesis and a low one to the light-from-below hypothesis. Jakob Hohwy (2013) makes a scientifically and philosophically informed case for this Bayesian approach to perception (though see Colombo, Elkin, & Hartmann, 2021). Although he does stress the importance of abduction for his case, he sees it as an essentially Bayesian principle (p. 25). He does not support this by any arguments, however, and as we explain later in this chapter, the claim is not as easy to maintain as he may think.

Jerry Fodor (2001) also recognizes the key role that abduction plays in cognitive processing. As he puts it, "by all the signs, the cognitive mind is up to its ghostly ears in abduction" (p. 78). In his opinion, that makes the prospects for real progress in cognitive science dim, because, in his opinion, "we do not know how abduction works [and so] we do not know how the cognitive mind works . . . " (Fodor, p. 78).[1] I am inclined to agree that at a minimum we do not know *much* about abduction and also that this hampers progress in cognitive science—and not only in that field. That was one important motivation for

1. Fodor (2001) thinks that, in particular, the computational theory of mind, which he otherwise favors, cannot account for the reliability of abduction (which he takes as a given). As Rellihan (2009) notes, however, there are independent reasons for holding that an adequate theory of mind must be hybrid, acknowledging the importance of both computational and associative mechanisms, and abduction is most plausibly accounted for in terms of the latter. Moreover, in view of the recent successes of deep learning (LeCun, Bengio, & Hinton, 2015; Goodfellow, Bengio, & Courville, 2016), Fodor's critique of associative theories of mind has come to appear much less compelling than it did twenty years back.

writing this book. Later in this chapter, I present some ideas about what may have stood in the way of promulgating a better understanding of abduction.

If you are a linguist, I encourage you to read on because abduction has been said to be fundamental to determining what a speaker means by an utterance. Specifically, it has been argued that decoding utterances is a matter of figuring out the best explanation of why someone said what she said when she said it. Even more specifically, authors working in the field of pragmatics have suggested that hearers invoke the Gricean maxims of conversation (Grice, 1989) to help them work out the best explanation of a speaker's utterance whenever the semantic content of the utterance is insufficiently informative for the purposes of the conversation, or is *too* informative, or is off topic, or is implausible, or is otherwise odd or inappropriate. As Yan Huang (1994, p. 2) puts it, "What pragmatics does is to provide a set of . . . *explanatory* principles which constrains the interpretation or production of an utterance whose linguistic representation has already been antecedently cognized" (emphasis mine).[2]

To go into a little more detail, pragmatics starts from the observation that true utterances can still be misleading. The semantic content of a speaker's utterance can be φ, but if the best explanation of why she made the utterance on the given occasion is, or entails, that she (also) wanted to convey ψ, then her utterance can be misleading even if φ is true—because ψ may be false. For example, we will typically interpret an utterance of

(1) Some of the guests are bringing flowers

as implicating (to use the technical term for what is conveyed over and beyond semantic content) that not *all* guests are bringing flowers. After all, if that were the case, it would be difficult to explain why the speaker did not simply *say* that all guests are bringing flowers (assuming a context in which the speaker is in a position to see whether all, or only some, guests are bringing flowers). So, if all guests are bringing flowers, then an utterance of (1) is misleading, even if the speaker is not *saying* anything false.

Or suppose a graduate student tells his supervisor

(2) You have published some papers that I really like.

2. In the same vein, see Bach and Harnish (1979), Dascal (1979), Hobbs (2004), Gabbay and Woods (2005, ch. 4), Chien (2008), and Douven (2012a); see also various papers in Bunt and Black (2000).

The supervisor can see two possible explanations of why the student uttered this sentence. One is that the student wanted to convey that he read some (but not all) of her papers and liked all the ones he read; the other is that he read some or all of her papers and liked some of those he read and some not so much. Depending on context, one or the other explanation may appear better to the supervisor. For instance, the first explanation may top the second if the supervisor is a prolific writer and the student is known to be somewhat lazy. In that case, the supervisor may infer that the student did not read all of her papers. If, on the other hand, the supervisor has so far only published a handful of papers, and the student is known for his critical attitude and for not being intimidated in the least by any of the professors, that may favor the second explanation. Naturally, the two explanations may also be in close competition, in which case the inference of an implicature can be guarded at best. This explains why implicatures can vary in strength (Douven, 2012a; Douven & Krzyżanowska, 2019). Furthermore, the example highlights a phenomenon that is discussed in chapter 3, to wit, that the extent to which the best explanation stands out as being the best can vary and can have a significant impact on people's willingness to infer the truth of that explanation.

If you are a philosopher, you may not need to be convinced of the importance of abduction, given how common appeals to this form of reasoning are in your field. For instance, social epistemologists have argued that our trust in other people's testimony rests on abductive reasoning (e.g., Harman, 1965; Adler, 1994; Fricker, 1994, 2017; Lipton, 1998). According to Jonathan Adler (1994, pp. 274–275), "[t]he best explanation for why the informant asserts that *P* is normally that . . . he believes it for duly responsible reasons and . . . he intends that I shall believe it too." That is why we are normally justified in trusting an informant's testimony.

And a number of philosophers of science hold that abduction is a cornerstone of scientific methodology (e.g., Boyd, 1984, 1985, 1990; Harré, 1986, 1988; Psillos, 1996, 1999; Lipton, 2004; Schupbach, 2017). According to Timothy Williamson (2017, pp. 334–335), "[t]he abductive methodology is the best science provides," and Ernan McMullin (1992) even goes so far as to call abduction "the inference that makes science." Relatedly, it has been said that abduction is the predominant mode of reasoning in medical diagnosis: physicians tend to go for the hypothesis that best explains the patient's symptoms (Josephson & Josephson, 1994; Dragulinescu, 2016; Stilgenbauer & Baratgin,

2018).[3] We look at some examples of the use of abduction in science in chapter 2 and consider abduction specifically in the context of medical decision making in chapters 6 and 7.[4]

In philosophy, abduction is frequently invoked in objections to so-called underdetermination arguments. We consider the issue of underdetermination at some length in chapter 2, and chapter 8 is devoted in its entirety to arguably the most prominent underdetermination argument in philosophy or anywhere: the argument for Cartesian skepticism. For now, suffice it to say that underdetermination arguments proceed by arguing that the data, or whatever counts as such in a given debate, are insufficient to warrant the adoption of this or that hypothesis, where the hypothesis at stake often appears commonsensical to us (that there is an external world, that others are conscious beings like us, that thanks to science we know a good deal about the mechanisms and processes causing observable events, and so on). While perhaps most prominent in epistemology and the philosophy of science, such arguments have either been advanced or are implicit in other areas of philosophy as well. Concomitantly, we find appeals to abduction in all those areas.

For instance, in metaphysics David Lewis (1973, 1986) is famous for defending modal realism, according to which other possible worlds are as real as our own world. Although typically met with an "incredulous stare"—as Lewis (1973, p. 86) acknowledges—he advocates the position because, he thinks, it offers the best explanation for a vast array of phenomena, including how we use modal expressions like "necessary" and "possibly" and how we talk about causes and effects. Similarly, Williamson (2016) gives an abductive argument for the hypothesis that there are objective modalities (which does not entail modal realism but still goes against, for instance, standard empiricist thinking according to which all modalities have a merely verbal status; see, e.g., van Fraassen, 1977). Philosophers of language have invoked abduction in debates

3. Interestingly, Patel and Groen (1986, 1991) and Patel, Groen, and Arocha (1990) found that the *degree* to which physicians rely on abductive reasoning may vary with their level of experience. In particular, they report a number of experiments showing that less experienced physicians rely more heavily on abductive reasoning than their more experienced colleagues.

4. Although they are not strictly speaking scientists, judges are truth-seekers whom we expect to reason as carefully as scientists. Michelon (2019) shows that they frequently rely on abduction, not only when assessing the evidence that is brought before the court but also in determining which legal principles have been tacitly assumed in relevant jurisprudence. On the role of abduction in the law, see also Walton (2005).

about the metaphysical status of truth and its supposed bearers. It has been argued, for instance, that minimalist conceptions of truth (see, e.g., Horwich, 1990) are metaphysically too lightweight to carry the explanatory workload required of truth by our current best theories of meaning and interpretation (for discussion, see Devitt, 1991, pp. 31–32; Davidson, 1996; Horsten, 2011, ch. 5). It has further been argued that we are warranted to believe in the existence of propositions because of the various explanatory roles that they play (e.g., van Elswyk, 2019). Possibly the oldest use of abduction in metaphysics is found in the philosophy of religion, in which authors through the ages have tried to argue that the existence of god best explains this or that part of reality, like the beauty or order of the world, human morality, or that people have religious beliefs (e.g., Hildebrand & Metcalf, in press; for discussion, see Schupbach, 2005, and De Cruz & De Smedt, 2015).[5]

Outside of metaphysics, Williamson (2017) argues that the case for classical logic over nonclassical logics is abductive: classical logic gives the best *overall* account of our various deductive practices, even if perhaps this or that nonclassical logic offers a more elegant solution of the semantic paradoxes.[6] And in his *Doing Philosophy* (Williamson, 2018), he notes that platonists defend their position "because they find it crucial for the best explanation of what mathematicians are doing" (p. 44). Indeed, in the same book (ch. 6) he argues extensively that abduction is central to philosophy in general, even as central as it is to science.

Finally, at the risk of sounding overly ambitious, I encourage you to read on for the sake of scientific progress. There are different ways in which science can progress. There is the kind of progress that makes headlines: the possibility of quantum entanglement over hundreds of miles established, a new disease discovered, the oldest yet humanoid skull found, a new treatment for cystic fibrosis developed, the first close-up pictures of the moons of Jupiter, artificial neural networks able to accurately predict protein structures from sequences of amino acids, and so on. But there is much scientific progress of a more prosaic variety about which the general public will not necessarily

5. On the role of abduction in metaphysics, see also Hawley (2006), Sider (2009), Biggs and Wilson (2017), and Schurz (2021).

6. On the role of abduction in logic, see also Priest (2005, p. 217), Russell (2015, 2019), and Hjortland (2019).

hear but that may be as important and consequential as the more newsworthy findings.[7]

For example, in recent decades much progress has been made in the field of data visualization. Thanks to the development of heat maps in the 1990s, we can more easily recognize patterns in the expression of genes in different tissues than was previously the case (Weinstein, 2008). Similarly, the development of so-called actograms—a kind of diagrams for representing the activity of organisms throughout the day—has contributed much to our understanding of the circadian rhythm (Sheredos et al., 2013).

More to the present point is Ian Hacking's (1984, p. 70) observation that "[t]he quiet statisticians have changed our world—not by discovering new facts or technical developments but by changing the ways that we reason, experiment and form our opinions about it." Progress is to an important part due to the fact that we have collected more data, which in turn is to an important part due to our having improved and expanded our *techniques* for collecting data. But progress is no less due to the fact that we have learned more about what we can and cannot conclude from the data in our possession and that we have learned more about how to determine what the data tells us about our theories and ultimately the mechanisms and processes that generated the data. In other words, we have learned more about what can be learned from the data we have. We owe much of this to Hacking's "quiet statisticians."

The quote from Hacking might be understood to suggest that the development of statistics has been completed. The opposite is the case, however: the development of statistics, and more generally of research methods, is an ongoing process that is accompanied by discussion and disagreement but that from time to time produces results that virtually all in the scientific community regard as progress. This concerns, for example, the development of statistical techniques that allow us to make more accurate estimates or to determine more precisely how certain we can be of a prediction.

Here is an illustration that is also relevant to subsequent discussion, in which we look in detail at some empirical data about explanatory reasoning. Regression analysis is one of the most commonly used statistical techniques for determining the influence (if any) of one or several variables (the "independent variables" or "predictor variables") on another variable (the "dependent

7. In fact, if Stegenga (2018) is right, medical progress, which is arguably the kind of progress most consistently covered by the media, tends to be overpraised.

variable" or "outcome variable"). The result of such an analysis can help us predict the value that the dependent variable will take on the basis of measured values of the independent variables. Suppose, for instance, that you are interested in the relationship between font size and reading speed. To investigate this, you let people read texts that are set in the same font but in different sizes, and you measure the speed at which they read the text. No doubt there will be variation in your participants' reading times, and if you are lucky, a regression analysis will find a significant relationship between reading time and font size.

Regression analysis has proven to be a more than highly valuable statistical technique. Not so long ago, however, a number of statisticians noted that in applications such as our fictional example, the method is not optimal. In that example, we assume that you will find variation in reading times that you can then relate to variation in font size. But it is to be noticed that part of the variation that you find may have to do with the fact that you have used different test subjects: not everyone is sensitive to variations in font size in the same way. This variation among test subjects can influence the outcome of your analysis in a way and to an extent that you do not want. In this experiment, you are not interested in individual differences among people; you want to establish the relationship between font size and reading speed *in general*.[8]

In response to this problem, so-called mixed-effects models have been developed (sometimes called "hierarchical linear models"). These models are intended to accommodate that there can be different sources of variance in the data, some of which may not be of interest to us. In this example, some variance will come from the fact that you manipulated the font size—the variance that interests you—but some will also come from the fact that the participants will vary in important respects. Some will be faster readers in general than others, some will have better eyesight than others, some will be more motivated to participate than others, some will be better rested at the

8. Stated more technically, the problem is that in standard regression analysis, the data are assumed to be independent and identically distributed. In particular, the independence part of this assumption—the probability of a data point taking a specific value is independent of which values are taken by the other data points—is typically violated when each participant contributed more than one data point: data points coming from the same participant will often be correlated. Violations of assumptions of statistical procedures are not always harmful, but violating the independence assumption in the case of a regression analysis *is* harmful (Judd, Westfall, & Kenny, 2012).

time of the experiment than others, and so on. Sometimes we are interested in precisely such differences. But in our fictional example we assume that you are *not* interested in those. Mixed-effects models help us to separate interesting from uninteresting variance in the data and thereby help us avoid discrediting the variable or variables of interest for failing to explain variance that is due merely to elements in our experimental design that we consider to be random (like the participants we happened to enroll, or the particular items we used as materials). Think of the way in which this is achieved as first creating a regression model for each participant separately and then constructing from those individual models a kind of aggregate model. It has been shown that this type of analysis tends to yield more accurate estimates of the particular parameters of interest than traditional regression models do (Baayen, 2008; Zuur et al., 2009; Singmann & Kellen, 2020).

My hope is that a more thorough study of abduction than we have seen so far will lead to a similar kind of scientific progress—that, more exactly, and with certain qualifications made subsequently, we will be able to get more mileage out of our evidence by paying special attention to explanatory connections between that evidence and whatever hypotheses that we are considering.[9] According to the current mainstream, however, there is no way to assign a role to such connections that would not make us deeply irrational; or at least there is no *interesting*, *substantive* way to do it. So, if we are to make the kind of progress that, I believe, a closer attention to abduction would allow us to make, we must first remove a number of persistent misunderstandings about it that prevent the mainstream from deeming abduction even worthy of serious attention.

In short, I encourage you to read on if you are interested in a mode of reasoning that appears central to numerous cognitive processes, that people tend to find entirely natural after you have told them about it, that may be so routine and automatic that often we do not even notice when we engage in

9. Here, I am in the company of Haig (2005a, 2005b), who expects much from the "abductive theory of method" that he proposed for the behavioral sciences. However, although he is not concerned to articulate a specific version of abduction, from the brief statement of abduction that he gives (e.g., Haig, 2005a, p. 381) it appears that he endorses the version labeled ABD1 in section 2.2.2, which is shown to be inadequate. He also does not address any of the concerns that have been raised about the tenability of abduction, and takes the reliability of abduction for granted (Haig, 2005a).

it, but that is also explicitly appealed to in various domains of research, and whose study finally raises the prospect of making some real scientific progress.

1.2 Two Common Misconceptions

If abduction is so ubiquitous, plays all these roles in the cognitive lives of people, and has the other features mentioned previously, why do we know so little about it and perhaps even, if Fodor is right, simply have *no* idea of how it works? When pressed, we may have difficulty saying what exactly abduction is. Moreover, many who feel attracted to the idea of abduction do not seem to know what to make of the standard objections to abduction. Finally, even if psychologists may be interested in various *descriptive* issues surrounding abduction, the increasing popularity of Bayesian statistics raises the question of why we should still care about the *normative* status of abduction. Indeed, in light of the successes about which Bayesian statistics can boast (Kruschke, 2011; Gelman et al., 2013; McElreath, 2020), any claim to the effect that the study of abduction could contribute to the progress of science would appear overblown, possibly even ludicrous.

To be fair, as is clear from the citations given at the outset of the previous section, psychological research into how people rely on abduction has been picking up steam in recent years. However, I firmly believe that abduction would have received still more attention were it not for some persistent misunderstandings about its normative status. I believe this even if Shira Elqayam and Jonathan Evans (2011) are right that psychologists should concern themselves with descriptive issues only and should leave normative issues to others.[10] Most psychologists working on reasoning appear to know the philosophical literature fairly well. If all they read there about abduction are deprecating remarks, they may not receive much incentive to prioritize empirical work on abduction. Rather, they may come to think of abductive reasoning as a further bias that may deserve no *special* attention, given that we already know quite a bit about biases. (See section 3.4 for more on this.) In philosophy, misunderstandings about abduction have hampered research into it to an even greater extent. It is an overarching aim of this book to rectify those misunderstandings. I begin by describing what I see as the two main ones.

10. This claim has been contested; see the responses to these authors in the same issue of *Behavioral and Brain Sciences* in which their target paper appeared.

1.2.1 Nothing but a Slogan

So far, we have given only the sketchiest characterization of abduction, as a mode of inference that makes explanatory considerations relevant to what we are licensed to believe. That may be excusable for an introductory chapter, but Bayesians and others have complained that it is difficult to find *anywhere* a statement of abduction that is more specific than such slogan-like characterizations. That alone can appear a reason to prefer Bayes's rule over abduction: rather than a slogan, Bayes's rule is a mathematically precise formula. It would thus seem that those of us who are still attracted to the idea of abduction have to provide a formulation of it that is nearly as precise as the formulation of Bayes's rule.

First, it is simply not true that there are no precise versions of abduction. In various chapters, we look at versions that are *as* precise as Bayes's rule and are also completely cast in the language of probability theory. Admittedly, while these versions serve important dialectical roles in the book, I am not vouching for any of them as adequate update rules. That is at least partly because I am not presently willing to commit to any specific theory of explanation,[11] nor am I willing even to commit to the claim that explanation quality can be measured in probabilistic or more general formal terms, as some authors have claimed or at least assumed (e.g., McGrew, 2003; Glass, 2007, 2012a, 2018, 2021a, 2021b; Schupbach, 2011a; Schupbach & Sprenger, 2011; Crupi & Tentori, 2012; Brössel, 2015; Cohen, 2016; Sprenger & Hartmann, 2019, ch. 7).[12]

11. Nor need I do so; see Prasetya (in press). Prasetya distinguishes the approach to explanationism taken in this book ("ampliative explanationism," as he calls) from the heuristic and objective approaches to explanationism (as defended by, e.g., Lipton, 2004 and Dellsén, 2018, and, respectively, Weisberg, 2009, Poston, 2014, and Climenhaga, 2017) and argues that those other approaches do commit one to particular theories of explanation.

12. That may in fact be unlikely if, as some have argued (e.g., de Regt & Dieks, 2005; de Regt, 2017; but cf. Strevens, 2013), explanation is conceptually tied to understanding or kindred notions (e.g., articulated awareness; see Woody, 2004). The reason is that understanding may be too complex a notion to be captured in strictly probabilistic or formal terms. On the other hand, in a major approach to causality (Pearl, 1988, 2000; Sloman, 2005; Ali et al., 2010; Ali, Chater, & Oaksford, 2011; Fernbach & Erb, 2013; Oaksford & Chater, 2013, 2014, 2017, 2020a; Sprenger, 2018; Hartmann, in press), this notion is analyzed by means of graphical models, which are at the same time statistical models, encoding relations of probabilistic dependence and independence. Insofar as explanation is *causal* explanation, work on these so-called causal Bayes nets may also help to capture the notion of explanation and to quantify explanation quality. To be sure, not every explanation is a causal explanation (Wouters, 1995, 2007; Bueno

Nevertheless, in the following we often *pretend*—again for dialectical purposes—that explanation can be captured in probabilistic terms. But even if explanation quality is adequately captured by one or more formal measures, is it not still a problem that, as I said, we encounter a *variety* of precise versions of abduction? How in that case are we supposed to pick the right one? The answer is that, no, it is not necessarily a problem if there is not a best mathematically precise version of abduction, just as it is not necessarily a problem if there is *no* such version. The suggestion that I make is that the advocates of abduction should rather embrace the previous critique instead of trying to counter it. The suggestion, to be more exact, is that abduction is best thought of as a slogan such as that explanatory power is of confirmation-theoretic significance or similar circumscriptions previously discussed, and in each situation this slogan has to be fleshed out in the best possible way for that situation. Finding the best elaboration for a given situation may be a matter of experience or talent. Some people may be better at this than others, maybe because they have had more practice, or have better insights, or have a combination of the two. That may make them better scientists or generally more successful in life. Finding the elaboration of the slogan that best fits a given situation may even sometimes be a matter of luck. And we should reckon with the possibility that sometimes the elaboration can be made formally precise, and sometimes not.[13]

To clarify the suggestion, I start by offering a couple of analogies. At various junctures, we support claims by using computer simulations. The simulations were all written in Julia, a language for scientific computing (Bezanson et al., 2017). One of the main features of this language (which is not unique, though many languages do *not* have it) is called "multiple dispatch." This

& Colyvan, 2011; Mancosu, 2015; Lange, 2016; Kasirzadeh, in press; Reutlinger, Colyvan, & Krzyżanowska, in press; Saatsi, in press), but if Kitcher (1981) and others are right, then at least some types of noncausal explanation can be understood in terms of coherence (internal coherence, coherence with the data, or coherence with background knowledge). Inasmuch as the currently best accounts of coherence are all of a probabilistic variety (e.g., Shogenji, 1999; Olsson, 2002, 2005; Bovens & Hartmann, 2003; Fitelson, 2003; Meijs, 2005; Douven & Meijs, 2007; Meijs & Douven, 2007; Siebel & Wolff, 2008; Schupbach, 2011b; Schippers, 2014; Koscholke, 2016; Koscholke & Jekel, 2017; Schippers & Schurz, 2017; Koscholke & Schippers, 2019; Koscholke, Schippers, & Stegmann, 2019), such types of explanation can still be analyzed in strictly probabilistic terms.

13. This is also why I am not hopeful that we can have a general logic of abduction, parallel to deductive logic. For some interesting attempts, see Aliseda (2006) and Meheus and Batens (2006).

feature offers the capability of defining different so-called methods for a given function; the methods pertain to different argument types or combinations of such types. For instance, in Julia although the + operator always performs the same operation at some abstract level of description, at the most basic level what happens when that operator is called on a number of arguments is different for different types of arguments or combinations of types of arguments. One can see this clearly by comparing the assembly code (a very basic representation of the Julia code, close to the "metal") for two operations, for example, +(5, 6) and +(5.0, 6.0). In the first case, + operates on integers, and in the second, it operates on *floating point numbers*, or *floats*, which are your computer's way of representing as best it can the real numbers.[14] As a result, the assembly code is quite different in the two cases, even though the higher-level code—the code that we write—contains nothing to suggest this.[15] This is quite similar to what I claim about abduction, namely, that at a general level it can be described as a mode of reasoning that assigns special weight to explanatory factors but that this broad description needs further filling in, where the filling-in may vary per application context.

Here is another analogy: psychologists studying problem solving have identified a number of general strategies that people use to solve problems. Among them is a strategy often called "divide and conquer," which comes down to dividing a problem into more manageable subproblems, solving the subproblems, and obtaining a solution to the initial problem by collating the solutions to the subproblems. This strategy will appear familiar to everyone because we all frequently rely on it. It is a strategy that parents recommend to their children and teachers to their students. For instance, mathematics teachers tell their students that although there are several ways to calculate the area of a pentagon, the simplest one is to cut the pentagon into three triangles,

14. How well it can represent them depends on the computer's architecture, but no computer can represent them perfectly because computers are finite things (see Goldberg, 1991, or Overton, 2001, for details). To see the difference between how your computer stores, for instance, the number 1 as an integer and how it stores the same number as a float, you can use Julia's `bitstring` function, which yields the bit representation of a number, and compare the output of `bitstring(1)` with that of `bitstring(1.0)`. See appendix E on how to use Julia.

15. To inspect assembly code in Julia, one can use the `@code_native` macro; for our example, compare the output of `@code_native +(5, 6)` with that of `@code_native +(5.0, 6.0)` To see all the methods associated with a function, one can call the `methods` function on it. For instance, executing `methods(+)` in Julia will show the 184 methods for the + function that are currently implemented in the language.

calculate the areas of those triangles, and then add the outcomes. Similarly, but at a more advanced level, they instruct students to integrate a function over a complicated region by trying to divide the region into smaller regions such that one obtains nice bounds on one's integrals and then summing up the integrals over those smaller regions. In a very different example, piano teachers recommend that their students study difficult passages by grouping the notes in certain ways, practicing the groups separately, and after each group has been mastered, playing the passage in its entirety.

But note that the divide-and-conquer strategy, too, needs filling in. For what are the subproblems? And how should one collate the solutions to the subproblems to obtain the solution to the overarching problem? This is difficult, if not impossible, to say in general. Discovering the best way to divide a larger problem and the best way to put the partial solutions together is sometimes far from straightforward, and some people may be better at these things than others. In piano practice, it is known that one way of grouping the notes in a passage can be much more efficient than others in reducing the difficulty of that passage, and some piano teachers are famous for having a special knack for finding clever groupings for tricky passages; the best of them will even tell you that different groupings may work for different students (no two hands are the same, as the saying goes, though the more relevant observation is probably that no two brains are the same). Successful mathematical modeling is also often a matter of identifying the most expedient way to split a problem into smaller ones.

Again, one of the claims argued for in this book is that it is no different with abduction: different situations may call for different ways of letting explanatory considerations guide one's belief changes. Picking the right way for the given situation is a skill that may come more easily to some people than to others but that most—hopefully, all—of us possess to some degree and may be able to hone, with enough dedicated practice.[16] Various authors (e.g., Day & Kincaid, 1994; Lipton, 2004) have argued that abduction is context sensitive in that its use heavily relies on background knowledge, most notably,

16. This is not to deny that in the hands of *some* people abduction is a risky tool. As Mirabile and Horne (2019) show, people prone to "conspiracist ideation" (Swami et al., 2011), who have a general tendency to adopt conspiracy theories, often come to hold such theories on explanatory grounds. There are at least parts of the practice of abductive reasoning that these people have not fully mastered or even not mastered at all.

for making judgments of explanatory goodness.[17] The context sensitivity that flows from the current proposal is more extreme, in that it makes the exact shape that abductive reasoning should take dependent on context as well (while also acknowledging the kind of context sensitivity discussed by previous authors).

The idea advanced here can also be related to Rudolf Carnap's work on concept explication. Concept explication, as Carnap conceived it, is "[t]he task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept" (Carnap, 1947a, pp. 7–8). He saw this task as central to much of what philosophers do, and he was clear about when—in his view—it has been carried out successfully. Next to being more exact, the replacing concept should also bear some similarity to the replaced concept, and it should meet some adequacy conditions, like being simple and fruitful (Carnap, 1950, p. 5). Think of the slogan-like formulation of abduction as the not quite exact concept. Then, as subsequently shown, that can be given more exact explications that are arguably adequate, albeit that an explication that is adequate in one situation need not be so in another situation. In that second situation, however, a different explication of the slogan may be adequate.[18]

My proposal is meant to be in accord with a more general claim that has been forcefully advocated in cognitive science and psychology, to wit, that philosophers have been wrong in their "universalist" approach to the question of the rationality of belief change. Gerd Gigerenzer and various of his collaborators, and more recently also Elqayam, have argued that the idea—prominent among philosophers—that rationality can be captured by a number of universally valid principles has been an illusion all along. According to these researchers, we can understand the notion of rationality only if we take into account (1) the various ways—biological and cognitive—in which humans are limited; (2) the goals that people have; and (3) the environment or environments in which humans operate and interact in pursuit of their goals. Biological, cognitive, and environmental constraints, as well as goals, can all vary greatly from one individual to another and from one moment to another. This led Elqayam, Gigerenzer, and others to conclude that

17. On page 5, we discussed an example of the context sensitivity of abduction even as it functions in lower-level cognitive processes.

18. Thanks to Youness Ayaita for helpful discussion here.

whether someone reasons or acts rationally depends on whether the reasoning or behavior helps the person achieve whatever goal or goals she pursues in whichever environment she is in, given the cognitive resources available to her in that environment. Abduction, being a broadly circumscribed rule that can, however, be tailored to one's needs, depending on the circumstances, fits much more naturally with this so-called ecological conception of rationality than the supposedly one-size-fits-all rule that Bayesians pretend to offer. This is one of the key points argued in chapters 6 and 7.

The foregoing is not to say that the current popularity of Bayesianism is completely undeserved, nor is that something argued in this book. As explicitly pointed out, adopting Bayesian norms may be the right thing to do in certain contexts. But Bayesian *imperialism*, according to which those norms should govern our behavior, cognitive as well as practical, in each and every context, is misguided. This imperialism has gone at the expense of proposals that assign a guiding role to explanatory considerations. It is shown here how taking into account such considerations can offer important benefits, which should be reason to give the said proposals a fairer hearing than philosophers and also many psychologists have so far been willing to do.

To be clearer still about the position that I advance in this book, I refer to William Lycan's (2002, p. 417) useful typology of versions of "explanationism" (as he and some others refer to the broad idea that explanation has confirmation-theoretic significance). In this typology, "Weak Explanationism" is the view that we can rationally reach a conclusion via abductive reasoning; however, it is left open whether abduction is a fundamental form of reasoning or parasitic on, say, Bayesian reasoning. "Sturdy Explanationism" adds to Weak Explanationism the claim that abduction is fundamental indeed. Whereas both Weak and Sturdy Explanationism are compatible with there being other forms of rational nondeductive reasoning, "Ferocious Explanationism" is *not*: according to this position, abduction is the only such form.[19] Finally, the strongest version of explanationism—which Lycan calls (rather infelicitously, in my opinion) "Holocaust Explanationism"—is the view that abduction is the only rational form of reasoning *tout court*, meaning that, for instance, even our reliance on deductive reasoning is to be justified on explanatory grounds. In terms of this typology, the position to be defended in the following is closest to Sturdy Explanationism, the main qualification

19. Harman qualifies as a ferocious explanationist in this typology; see page 2.

being that, in my view but not in Lycan's (as far as I can tell), it would be wrong to think of abduction as a single specific rule of inference; "abduction," as explained, is to be conceived as a blanket term denoting a broad idea to be filled in differently in different contexts.

1.2.2 Right but Boring, or Interesting but Wrong

Over the last two decades, Bayesian confirmation theory has firmly established itself as the dominant view on confirmation; currently, one cannot effectively discuss a confirmation-theoretic issue without making clear whether and, if so, why one's position on that issue deviates from standard Bayesian thinking. Abduction assigns a confirmation-theoretic role to explanation. In the probabilistic versions previously hinted at, and subsequently examined in great detail, explanatory considerations help to make some hypotheses more credible and others less so. By contrast, Bayesian confirmation theory makes no reference at all to the concept of explanation. Does this imply that abduction is at loggerheads with the prevailing doctrine in confirmation theory? Answer *yes*, and many people are likely to tell you that although your answer suggests that abduction might be *interesting* for offering some genuine alternative to Bayes's rule, it also shows that you are on the wrong track. Those people will claim that any systematic way for changing your beliefs that deviates from Bayes's rule will make you irrational. If, on the other hand, you answer *no*, then the mainstream will think that you may be right but not that your idea presents a significant challenge. If you are not offering a real *alternative* to Bayes's rule (or some generalization thereof), then how is your proposal not boring?

A number of authors have argued that abduction is compatible with Bayesianism indeed, but that this does not make abduction boring. To the contrary, it should be considered a much-needed supplement. At this time, the fullest defense of this view has been given by Peter Lipton (2004, ch. 7); as he puts it, Bayesians should also be explanationists.[20]

This requires some clarification. What could it mean for a Bayesian to be an explanationist? In order to apply Bayes's rule and determine the probability for hypothesis H after learning evidence E, the Bayesian agent has to determine the probability of H conditional on E. For that, she needs to

20. For similar defenses, see Okasha (2000), McGrew (2003), Poston (2014, ch. 7), and Niiniluoto (2018, ch. 6).

assign unconditional probabilities—*prior probabilities*, or *priors*—to H and E as well as a probability to E given H, called "the likelihood of H on E." How is the Bayesian to determine these values? Probability theory gives us *more* probabilities once we have *some*; it does not give us probabilities from scratch. When H implies E or the negation of E, or when H is a statistical hypothesis that bestows a certain chance on E, then the likelihood follows "analytically."[21] However, this is not always the case, and even if it were, there would remain the question of how to determine the priors. This is where, according to Lipton, abduction comes in. In his proposal, Bayesians ought to determine their prior probabilities and, if applicable, likelihoods on the basis of explanatory considerations.

Exactly how are explanatory considerations to guide one's choice of priors? The answer to this question is not as simple as one might initially think. Suppose that you are considering what priors to assign to a collection of rival hypotheses and you wish to follow Lipton's suggestion. How are you to do this? A seemingly obvious—although somewhat vague—answer may go like this: Whatever exact priors you are going to assign, you should assign a higher prior to the hypothesis that best explains the available data than to any of its rivals (provided there is a single best explanation).[22] Note, though, that your neighbor, who is a Bayesian but thinks confirmation has nothing to do with explanation, may well assign a prior to the best explanation that is even higher than the one you assign to that hypothesis. In fact, his priors for best explanations may even be consistently higher than yours, not because in his view explanation is somehow related to confirmation—it is not, he thinks—but just because. . . . In this context, "just because" is a perfectly legitimate reason, because any reason for fixing one's priors counts as legitimate by Bayesian standards. According to mainstream Bayesian epistemology, priors (and sometimes likelihoods) are up for grabs, meaning that one assignment of priors is as good as another, provided that both are coherent (that is, they obey the axioms of probability theory). Lipton's recommendation to the Bayesian to be an explanationist is meant to be entirely general. But what should your

21. This claim assumes some version of Lewis's (1980) Principal Principle (see p. 78), and it is controversial whether or not this principle is analytic; hence the scare quotes.

22. To forestall misunderstanding, your prior for a hypothesis is not the probability you assign to that hypothesis before *any* data are in. The terms "prior" and "posterior" are to be understood relative to whichever data are newly obtained and subsequently to be accommodated. In Lindley's (1972, p. 2) oft-cited phrase, "Today's posterior is tomorrow's prior."

neighbor do differently if he wants to follow the recommendation? Should he give the same prior to any best explanation that you, his explanationist neighbor, give to it, that is, *lower* his priors for best explanations? Or should he rather give even *higher* priors to best explanations than he already does?

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

A more interesting answer to the question of how explanation is to guide one's choice of priors has been given by Jonathan Weisberg (2009). We said that mainstream Bayesians regard one assignment of prior probabilities to be as good as any other. So-called objective Bayesians do not do so, however. These Bayesians think priors must obey principles beyond the probability axioms in order to be admissible. Objective Bayesians are divided among themselves over exactly which further principles are to be obeyed, but at least for a while they agreed that the Principle of Indifference (Keynes, 1921) is among them. Roughly stated, this principle counsels that, absent a reason to the contrary, we give equal priors to competing hypotheses. However, in its original form the Principle of Indifference may lead to inconsistent assignments of probabilities and so can hardly be advertised as a principle of rationality (Gillies, 2000, ch. 3). The problem is that there are typically different ways to partition logical space (the space of all logically possible worlds) that all appear equally plausible given the problem at hand and that not all of them lead to the same prior probability assignment, even assuming the Principle of Indifference (see section 3.2.1 for details). Weisberg's proposal amounts to the claim that explanatory considerations may favor some of those partitionings over others. Perhaps we will not always end up with a unique partition to which the Principle of Indifference is to be applied, but it would already be progress if we ended up with only a handful of partitions. In that situation we could arrive in a motivated way at our prior probabilities by proceeding in two steps, namely, by first applying the Principle of Indifference to the partitions separately, thereby possibly obtaining different assignments

of priors, and by then taking a weighted average of the priors thus obtained, where the weights, too, are to depend on explanatory considerations. The result would again be a probability function—the uniquely correct prior probability function, according to Weisberg.

The proposal is intriguing as far as it goes but, as Weisberg admits, in its current form, it does not go very far. For one thing, it has nothing to say about how explanatory considerations are to determine the weights required for the second step of the proposal. For another—and this seems the bigger problem—it may be idle to hope that taking explanatory considerations into account will in general leave us with a manageable set of partitions, or that even if it does, this will not be due merely to the fact that we are overlooking a great many prima facie plausible partitionings of logical space to begin with.

Another interesting suggestion about the connection between abduction and Bayesian reasoning is that the explanatory considerations may serve as a heuristic to determine, even if only roughly, priors and likelihoods in cases in which we would otherwise be clueless and could do no better than guessing (Niiniluoto, 1999a; Okasha, 2000; McGrew, 2003; Lipton, 2004, ch. 7; Cabrera, 2017; Dellsén, 2018). This suggestion is sensitive to the well-recognized fact that we are not always—indeed are quite often *not*—able to assign a prior to every hypothesis of interest or to say how probable a given piece of evidence is conditional on a given hypothesis. Consideration of that hypothesis's explanatory power might then assist us in gauging the prior to assign to it or the likelihood to assign to it on the given evidence.

Bayesians, especially the more modest ones, might want to retort that the Bayesian procedure is to be followed only if (1) priors and likelihoods can be determined with some precision and objectivity, or (2) likelihoods can be determined with some precision and priors can be expected to "wash out" as more and more evidence accumulates, or (3) priors and likelihoods can both be expected to wash out. In the remaining cases, Bayesians might say, we should simply refrain from applying Bayesian reasoning. A fortiori, then, there is no need for an abduction-enhanced Bayesianism in these cases. Further, some incontrovertible mathematical results indicate that, in the cases that fall under (1), (2), or (3), our probabilities will converge to the truth anyway (Gaifman & Snir, 1982). Consequently, in those cases there is no need for the kind of abductive heuristics that the previously mentioned authors suggest.[23]

23. For similar concerns, see Weisberg (2009, section 3.2).

Finally, Stathis Psillos (2000) proposes yet another way in which abduction might supplement Bayesian confirmation theory, a way very much in the spirit of Peirce's conception of abduction, which is briefly discussed in section 2.2.1. The idea is that abduction may assist us in selecting plausible candidates for testing, where the actual testing will follow Bayesian lines. However, Psillos (2004) concedes that this proposal assigns a role to abduction that will strike committed explanationists as being too limited.

If the explanationists' attempt to cozy up with Bayesians makes their proposal boring, attempts to sell abduction as *deviating* from Bayes's rule are nonstarters, or so a number of influential authors hold. Why is that? We are typically given one of two arguments: the dynamic Dutch book argument and the argument from inaccuracy minimization. These arguments are too intricate and too important—they have been the main obstacles to a wider acceptance of abduction—to be dealt with in a single section. We devote a whole chapter to them. To briefly address them here: the dynamic Dutch book argument is alleged to show that using any rule for belief change other than Bayes's rule may lead one to assess as fair a number of bets that together ensure a financial loss, come what may, and thus in the end that using any rule for belief change other than Bayes's rule makes one irrational; the argument from inaccuracy minimization purports to show that if one uses some non-Bayesian rule for belief change, then one's mental representation of the world is bound to be less sharp than it would be were one to use Bayes's rule instead.

As we shall see, neither argument is ultimately compelling, and each fails for more than one reason. Perhaps the simplest but also most important reason for buying into neither is that, even if they were sound (which they are not), absolutely nothing follows from the mere claim that vulnerability to dynamic Dutch books and / or failure to minimize inaccuracy is *really, really bad*. That could be a reason to abandon abduction if it had at the same time been shown that there is nothing *really, really good* to be had by reasoning abductively, something that is out of reach for people sticking to Bayes's rule, or any other rule. Furthermore, none of the critics of abduction has even bothered to consider that possibility. In fact, as shown subsequently, it is more than a possibility; there is positive reason to hold that abduction offers benefits that we would forego by consistently adhering to Bayes's rule. Put more generally, if you believe we should judge epistemic norms (such as rules for belief change) by their consequences, then make sure that you are not too narrowly focused on only certain *kinds* of consequences, possibly

leaving others that might be just as relevant, or even more relevant, out of consideration. In my view, many Bayesians criticizing abduction have erred in this regard.

My hope is that, if you do read on, I will ultimately be able to convince you that abduction is interestingly different from Bayes's rule and that there is nothing intrinsically untoward about it. As mentioned, abductive reasoning may involve a bit of art, of imagination, and of creativity; and although—as with all matters artistic—some of us may have a greater inherent talent for abductive reasoning than others, I firmly believe that we can all improve our reasoning skills by getting a better understanding of how abduction works (see, in the same vein, Mirabile, 2020).

1.3 Overview

The following are the main goals I hope to achieve in this book:

(1) To clarify what abduction is and to explain why we should care about it. I argue that abduction is the general idea that explanation has a role to play in confirmation, where however this idea needs to be specialized per context of use; I further argue that we should study abductive reasoning both for normative and for descriptive reasons.

(2) To convince the reader that, in assessing the rationality of abductive reasoning, our standards should attend to known facts about human psychology: the study of abduction should be part of an effort to develop a comprehensive theory of rationality for *real* people, not for a future make of computationally unconstrained robots.

(3) To show that the bad things you may have heard about abduction are vastly overblown and, at a minimum, are misinterpreted: the standard arguments against abduction make assumptions the advocates of abduction are by no means forced to commit to, and even if those arguments are granted, they only show that there can be costs attached to reasoning abductively, not that this mode of reason is cost-ineffective.

(4) To show that there are actually quite a few good things to report about abduction, in particular that, in the right hands, abduction is a powerful tool: computer simulations are used to show that, in various types of contexts, we may be better off using abduction than the currently more popular Bayes's rule; it will also be shown that abduction can do

the sort of heavy lifting that philosophers have always hoped it would do for them.

Next is an overview of how each chapter is meant to help realize these goals.

Chapter 2 states the idea of abduction in greater detail by examining abduction from a number of different angles. Abduction is distinguished from other modes of inference; we briefly trace its historical roots; some applications are described; we consider some common definitions of abduction in the philosophical literature; and we discuss the problem of underdetermination, which is the problem that most often leads philosophers to invoke abduction.

Chapter 3 examines reasoning research concerning how judgments of explanation quality impact people's beliefs and especially changes of belief. Special attention is given to research showing that people deviate from Bayesian principles in ways one would expect to occur if they were influenced by explanatory considerations. The aim is to raise psychologists' interest in explanatory reasoning. An equally important goal of the chapter is to make clear that if one wants to maintain that any violation of Bayesian principles betokens irrationality, then people are massively irrational (which some philosophers may not care about, but some certainly will). Moreover, the chapter provides some inspiration for formulating probabilistic versions of abduction that are compared with Bayes's rule along dimensions of performance that arguably matter to us.

Chapter 4 canvasses the two main objections that have been brought against abduction, both already mentioned: the dynamic Dutch book argument and the inaccuracy-minimization argument. The chapter first argues that it is a mistake to think that abduction, or any other rule for belief change, can be judged in isolation from other rationality principles, including decision-theoretic principles. Instead, it is to be viewed as part of a package of rules, and there are packages that include abduction and that allow one to use abduction without having to live in fear of Dutch bookies. The chapter then argues that critics who level the charge that abduction fails to minimize inaccuracy overlook that there are different senses of inaccuracy that are relevant in the context of the present debate, and that while abduction may fail to minimize inaccuracy in one sense, there is a different, arguably more important, sense of inaccuracy that is not touched at all by these critics' arguments.

Chapter 5 argues that one need not even grant that abduction fails to minimize inaccuracy in *any* sense! The same chapter also takes issue with the

claim that, in contrast to the dynamic Dutch book argument, the inaccuracy-minimization argument is a distinctively epistemic (as opposed to pragmatic) argument.

That abduction does not actually fall prey to the dynamic Dutch book and inaccuracy-minimization arguments does not automatically mean that it has anything to recommend it. Chapter 6 gives abduction a positive grounding. It uses computer simulations to show that versions of abduction can vastly outperform Bayes's rule in terms of obtaining a best balance between the speed with which our probabilities converge to the truth and the overall accuracy of those probabilities. That is precisely because abduction is more flexible and can be fine tuned to meet the specific needs of the situation in which we happen to be. The chapter also demonstrates how evolution may have predisposed us to reason abductively.

Chapter 7 adds a social perspective to the foregoing by studying artificial agents who update, either by Bayes's rule or by some version of abduction, on evidence they receive directly from the world but whose belief states are also impacted by those of others in their community. In such a social setting, updating via abduction appears to be particularly advantageous. Here, too, we find grounds for believing that evolution may have favored groups of epistemically interacting agents who reason abductively.

Chapter 8 highlights another benefit of abduction by putting it to work in probably the best-known underdetermination argument, namely, the argument for skepticism regarding the external world. Various authors have argued that we are warranted in believing in the existence of an external world because that belief best explains our sensory evidence. I explain why the argument for that claim is not as straightforward as it has appeared to many, in particular by pointing out that previous authors arguing abductively against the skeptic have done insufficient justice to skeptical concerns about admitting as evidence anything that goes beyond the purely phenomenal and also about rules of reasoning—most notably, abduction—that might enable one to move from premises about the phenomenal alone to a conclusion about the external world. The chapter argues that we can acknowledge those concerns and still be able to resolve the skepticism debate.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in EB Garamond by the author in LuaLaTeX.

Library of Congress Cataloging-in-Publication Data

Names: Douven, Igor, author.
Title: The art of abduction / Igor Douven.
Description: [Cambridge, Massachusetts] : Massachusetts Institute of Technology, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021031324 | ISBN 9780262046701 (paperback)
Subjects: LCSH: Abduction (Logic) | Reasoning. | Practical reason. | Bayesian statistical decision theory.
Classification: LCC BC199.A26 D68 2022 | DDC 160—dc23
LC record available at https://lccn.loc.gov/2021031324