Why abduction, not deduction, is indefeasible

Stephen Biggs and Jessica Wilson

(Forthcoming in *Inductive Metaphysics*, Andreas Hüttemann and Gerhard Schurz, eds.)

Abstract

It is often assumed that (rationally compelling, good) reasoning is defeasible if and only if it relies on an invalid argument, and that reasoning is indefeasible if and only if it relies only on valid arguments. We here argue that this common assumption is incorrect. We first argue that this equivalence should not be treated as definitional; we then argue that some reasoning is defeasible despite relying only on valid arguments, while other reasoning is indefeasible despite relying on invalid arguments. More specifically, we argue that deduction's being subject to rebutting defeaters renders it defeasible, despite its relying only on valid arguments, while abduction's being immune to rebutting defeaters renders it indefeasible, despite its relying on invalid arguments. We then offer an explanatory diagnosis of the errant assumption at issue, suggesting that it reflects both a too-quick generalization from the case of (enumerative) induction and a general tendency to conflate metaphysical and epistemic notions. Finally, we suggest that a kind of reasoning is an ultimate arbiter of disputes in a domain if and only if it is indefeasible in that domain, so that abduction is an ultimate arbiter of disputes in any domain in which it is operative.

Introduction

Epistemologists attribute defeasibility to reasoning (e.g., Pollock 1987), reasons (e.g., Sturgeon 2014), justification (e.g., Huemer 2001), and (less often) beliefs (e.g., Beltzer and Loewer 1988). Whether the property at issue is the same in each case is often unclear, even if some family resemblance is obvious. Here, we shall address only defeasible reasoning, leaving potential generalizations of our results to reasons, justification, and beliefs for another occasion.

What is defeasible reasoning? In introducing his account of this notion, Pollock offers the following characterization:

> [W]hen one judges the color of something on the basis of how it looks to him […] [s]uch reasoning is defeasible, in the sense that the premises taken by themselves may justify us in accepting the conclusion, but when additional information is added, that conclusion may no longer be justified. For example, something's looking red to me may justify me in believing

> that it is red, but if I subsequently learn that the object is illuminated by red lights and I know that that can make things look red when they are not, then I cease to be justified in believing that the object is red. (1987, p. 481).

What makes reasoning defeasible, Pollock suggests, is that additional information can defeat its conclusion, where a conclusion *c* is defeated by *x* if and only if *c* is justified independently of *x* but unjustified given *x*, where *x* is (at least often) evidence.[1] Crucially, the additional information does not defeat the conclusion simply by defeating an operative premise—learning about the illumination defeats the conclusion that the thing is red without defeating the lone premise that the thing looks red to him. For Pollock, this is what makes reasoning defeasible: not that its conclusion can be defeated, but rather that its conclusion can be defeated even when its premises remain undefeated. Conversely, for Pollock, what makes reasoning indefeasible is not that its conclusions cannot be defeated, but rather that its conclusions can be defeated only by defeating one or more of its premises. This characterization reflects that whether reasoning is defeasible or rather indefeasible depends on the nature of the reasoning itself, not on the status of the inputs on which it operates.

In this respect, defeasibility resembles (deductive, here and elsewhere) invalidity:[2] just as what makes an argument invalid is not that its conclusion can be false, but rather that its conclusion can be false even if its premises are true, what makes reasoning defeasible is not that its conclusion can be defeated, but rather that its conclusion can be defeated even if its premises remain undefeated. Likewise, indefeasibility resembles validity: just as a valid argument is valid not in that it invariably produces truth, but rather in that it invariably preserves truth, indefeasible reasoning is indefeasible not in that it invariably produces conclusions that cannot be defeated, but rather in that it invariably preserves lack-of-defeat. In both cases, the relationship between the premises and conclusion, rather than the status of the conclusion alone, is at issue.

While this similarity between defeasibility/indefeasibility and invalidity/validity is merely structural, it is typically assumed that the connection runs much deeper. Specifically, it is typically assumed that reasoning is defeasible if and only if it relies on an invalid argument. Koons (2022) makes this assumption when he characterizes defeasible reasoning as follows:

[1] This characterization of defeat follows an early proposal from Chisholm (1966/1989). We will talk of premises and conclusions' being defeated (undefeated, justified, unjustified, etc.) as shorthand for talk of beliefs in premises and beliefs in conclusions' being defeated (undefeated, justified, unjustified, etc.).

[2] Here and throughout, when we talk of validity and invalidity, we have in mind deductive validity and invalidity.

> Reasoning is defeasible when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provide support for the conclusion, even though it is possible for the premises to be true and the conclusion false.

In a similar spirit, Jaszczolt (2022) claims that "[some] laws of reasoning are 'defeasible' in the sense that if the antecedent of a default rule is satisfied, then its consequent is normally, but not always, satisfied", and Genin and Huber (2022) maintain that "it is clear […] that defeasible reasoning" is that which relies on an invalid argument.

If the assumption that reasoning is defeasible if and only if it relies on an invalid argument is correct, then all and only deductive reasoning is indefeasible (since all and only deductive reasoning relies on only valid arguments), and so both (enumerative) induction and abduction are defeasible (since both rely on invalid arguments). Indeed, the assumption connecting defeasibility to invalidity is sometimes expressed by equating defeasible reasoning with non-deductive reasoning. Pollock, for example, says that "What distinguishes deductive reasoning from reasoning more generally is that the reasoning is not defeasible" (1987). Strasser and Antonelli (2019) follow suit, contrasting deductive reasoning with defeasible reasoning:

> Our previous examples [of inductive, abductive, and probabilistic reasoning] are instances of ampliative reasoning. It is based on inferences for which the truth of the premises does not guarantee or necessitate the truth of the conclusion as in deductive reasoning. Instead, the premises support the conclusion defeasibly, e.g., the conclusion may hold in most/typical/etc. cases in which the premises hold.

Is reasoning defeasible if and only if it relies on an invalid argument? Is all and only deductive reasoning indefeasible? We argue that this common assumption is incorrect. We first argue that the assumption should not be taken to result from the definition of defeasible reasoning; specifically, we suggest that defeasible reasoning should be defined as that which is susceptible to defeat, not as that which relies on an invalid argument (§1). We then argue that deduction is defeasible despite using only valid arguments (because even deductive reasoning is subject to rebutting defeaters) (§2) and that abduction is indefeasible despite using invalid arguments (because abduction is holistic and so immune to rebutting defeaters) (§3). We go on to offer an explanatory diagnosis of why the assumption that reasoning is defeasible if and only if it relies on an invalid argument is so common, despite being false (§4). Finally, we suggest that abduction's being indefeasible renders it the ultimate arbiter of disputes in any domain in which it is operative (§5).

1. Operative definitions of defeasible and fallible

Defeasibility is a property of reasoning, not arguments. Invalidity is a property of arguments, not reasoning. Accordingly, defeasibility and invalidity are distinct properties. Nonetheless, people often talk of reasoning's being invalid, reflecting that it can be useful to talk of a property that stands to reasoning as invalidity stands to arguments. Since an argument is invalid if and only if its conclusion can be false even if all its premises are true, the corresponding property applied to reasoning (call it *X*) can be defined as follows:

> A form of reasoning ('procedure') R is *X* =def. a conclusion drawn through a meticulous application of R to some premises can be false even if all the premises are true.[3]

A conclusion drawn through a meticulous application of a procedure to some premises can be false even if all the premises are true if and only if that procedure relies on an invalid argument. Accordingly, the common assumption that reasoning is defeasible if and only if it relies on an invalid argument can be expressed as follows: reasoning is defeasible if and only if it is *X*.

Should 'defeasible' be defined as *X*? Suppose that defeasible =def. *X*. Then, whether a procedure is defeasible is a matter of the relationship between the truth-values of premises and conclusions: if a conclusion drawn through a meticulous application of a procedure to some premises can be false even if all the premises are true, then the procedure is defeasible; otherwise, it is indefeasible. The truth-values of premises and conclusions need not be, and ordinarily are not, facts about the epistemic status of reasoners; they are rather a matter of metaphysics, broadly construed (i.e., they are a matter of how things are, not a matter of how believers' justification is in particular). Accordingly, if defeasible =def. *X*, then we should assign defeasibility to metaphysics, not epistemology—in the sense of "assign" that Kripke has in mind when he says that "we should assign [necessity] to metaphysics", not epistemology (1971, 150). But defeasibility should be assigned to epistemology, not metaphysics; hence 'defeasible' should not be defined as *X*. This observation does not in itself falsify the common assumption that reasoning is defeasible if and only if it relies on an invalid

[3] An application of a form of reasoning is meticulous if it is as careful as that which an ordinary person can engage in. Such meticulousness is far less demanding than the idealization at issue in appeals to ideal reasoning—Chalmers, for example, maintains that ideal reasoning requires agents with "far greater than normal human capacities" (2012, 62). We restrict the applications at issue here to those that are meticulous, because a form of reasoning should not be rendered defeasible simply because its results can be defeated when it is done badly. This restriction is superfluous if a non-meticulous application of deduction is not an application of deduction, e.g., if affirming the consequent is not an application of modus ponens, but it becomes relevant if defeasibility is not defined through invalidity.

argument, but it does imply that any such equivalence is not definitional.

How, then, should ‘defeasible’ be defined? Recall the characterization from Pollock noted at the outset: reasoning is defeasible if and only if it is such that “the premises taken by themselves may justify us in accepting the conclusion, but when additional information is added, that conclusion may no longer be justified”. Following our earlier elucidation, the property Pollock characterizes (call it *Y*) can be defined as follows:

> A procedure R is *Y* =def. a conclusion drawn through a meticulous application of R to some premises can be defeated even if all the premises remain undefeated.

Should ‘defeasible’ be defined as *Y*? Suppose that defeasible =def. *Y*. Then, whether reasoning is defeasible is a matter of justification, specifically, of the relationship between one’s justification for beliefs about premises and conclusions—if a conclusion drawn through a meticulous application of a procedure to some premises can be defeated (i.e., rendered unjustified) even if those premises remain undefeated (i.e., remain justified), then the procedure is defeasible; otherwise, it is indefeasible. The justificatory statuses of premises and conclusions are facts about the epistemic status of reasoners (which is not to say reasoners’ intrinsic properties determine those facts). Accordingly, if defeasible =def. *Y*, then defeasibility belongs to epistemology—in the sense of “belongs” that Kripke has in mind when he says that a priority “belongs, not the metaphysics, but to epistemology” (1971, 150). More generally, defining ‘defeasible’ as *Y* captures that defeasibility is first and foremost a matter of susceptibility to defeat. We therefore shall define ‘defeasible’ such that defeasible =def. *Y*, treating this as a specification of Pollock’s earlier characterization.

Although ‘defeasible’ should not be defined as *X*, talk of *X* is useful, if only when discussing the common assumption that reasoning is defeasible if and only if it relies on an invalid argument. We find ‘fallible’ to be a fitting label for *X* because talk of fallibility indicates the possibility of falsehood (hence, e.g., to say that a belief is fallible is to say that it could be false).

We therefore adopt the following definitions:

> A procedure R is defeasible =def. a conclusion drawn through a meticulous application of R to some premises can be defeated even if all the premises remain undefeated.
>
> A procedure R is indefeasible =def. no conclusion drawn through a meticulous application of R to some premises can be defeated without defeating one or more premises.

> A procedure R is fallible =def. a conclusion drawn through a meticulous application of R to some premises can be false even if all the premises are true.
>
> A procedure R is infallible =def. no conclusion drawn through a meticulous application of R to some premises can be false unless at least one premise is false.

We think that these definitions capture what many epistemologists have in mind when they say ‘defeasible’, ‘indefeasible’, ‘fallible’, and ‘infallible’—which is not to say that these terms are used univocally in the literature. Those who disagree can treat these definitions as stipulative, with the stipulations having been made to facilitate discussion of whether all reasoning is, as per the common assumption, such that a conclusion drawn through a meticulous application of it to some premises can be defeated even if all the premises remain undefeated if and only if it is such that a conclusion drawn through a meticulous application of it to some premises can be false even if all the premises are true.

As a preliminary application of these definitions: (enumerative) induction is both defeasible and fallible. Suppose that Ida infers at time t that all swans are white from the premise that every swan observed up through time t has been white. Suppose that her application of induction is meticulous—she goes to great pains to ensure that the set of observed swans is a sufficiently large and random sample. Finally, suppose that her belief in the conclusion that all swans are white is later defeated by the discovery of a previously unobserved black swan. Since belief in the conclusion of her meticulous application of induction is defeated without defeating belief in its premise, induction is defeasible. Since the conclusion is false even though its premise is true, induction is fallible.

For induction, defeasibility and fallibility coincide. Do they always coincide, as per the common assumption? That they are definitionally inequivalent does not answer this question—though it blocks one route to the common assumption, and is moreover suggestive. Establishing the inequivalence of defeasibility and fallibility requires identifying an infallible procedure that is defeasible or a fallible procedure that is indefeasible. We identify such procedures below, arguing in particular that deduction is both infallible and defeasible and that abduction is both fallible and indefeasible.

2. The infallibility and defeasibility of deduction

We first note that deduction is defeasible despite being infallible.

That deduction is infallible follows from relevant definitions: a procedure is deductive =def. it relies on only valid inferences; a procedure is infallible = def. no conclusion drawn through a meticulous application of it to some premises can be false unless at least one premise is false, which is to say no application relies on an invalid argument; and a procedure relies on only valid inferences =def. no meticulous application of it relies on an invalid argument.

That deduction is defeasible follows from its results' being subject to rebutting defeaters—i.e., defeaters that render a conclusion unjustified without rendering the premises or inference that produced the conclusion unjustified. Consider an example in which a conclusion of reasoning that relies on only valid arguments is defeated without defeating an operative premise or inference:

> Dina infers that a wall is probably red from the justifiably believed premises that it looks red to her and that it is probably red if it looks red to her. A moment later, she infers that the wall is probably not red from the justifiably believed premises that she painted it yellow earlier in the day and that it probably is not red if she painted it yellow earlier in the day.

Since Dina should not believe both that the wall probably is red and that it probably is not red, at least one of these conclusions is defeated by the other; hence, given her epistemic situation, she should reject one or both conclusions. Nevertheless, no premise is defeated—at least, none needs to be. No claim in the later reasoning defeats (or need defeat) a premise in the earlier reasoning: affirming that she painted the wall yellow earlier in the day, that the wall probably is not red if she painted it yellow earlier in the day, and that the wall is probably not red does not (or need not) defeat either the earlier premise that it looks red to her or the earlier premise that it probably is red if it looks red to her. Likewise, no claim in the later reasoning defeats (or need defeat) a premise in the earlier reasoning.

To be sure, accepting that the wall is probably not red defeats belief in the conjunction of those premises, but to defeat belief in a conjunction is not to defeat belief in any conjunct. Lotteries provide a relevant example. If a lottery drawing with a million tickets is to have at most one winner, one might justifiably believe of each ticket that it is probably a loser. If moreover the lottery rarely has any winner, one might justifiably believe that this drawing will have no winner, and correspondingly, believe the relevant conjunction, i.e., that ticket-1, ticket-2, etc. are probably all losers. Learning later that the lottery was won would defeat belief in the conjunction but not belief in any conjunct, i.e., it would remain that one justifiably believes that ticket-1 is probably a loser, that ticket-2 is probably a loser, and so on, even though one could no longer justifiably believe that all tickets are losers.

In the above scenario, the conclusion of Dina's application of deduction is defeated without defeating any of its premises. Deduction is thereby shown to be defeasible, in light of the operative definition, since one case suffices to establish the defeasibility of a form of reasoning. As above, deduction is (by definition) infallible. Hence deduction is both infallible and defeasible, and as such it follows that the infallibility of a form of reasoning does not imply its indefeasibility, contra the common assumption.[4]

3. The fallibility and indefeasibility of abduction

Since deduction is defeasible, one might suspect that every procedure is defeasible, such that a fortiori no procedure is both indefeasible and fallible. This suspicion is incorrect, however, as we'll now argue.

To see what it would take for a procedure to be indefeasible, return to the case of Dina. Dina's reasoning is defeasible because her meticulously drawn conclusion is subject to a rebutting defeater. This observation suggests that if a procedure is to be indefeasible, then any meticulous application of it must be immune to rebutting defeaters. The converse also holds: if any meticulous application of a procedure is immune to rebutting defeaters, then no conclusion drawn through a meticulous application of that procedure can be defeated without defeating an operative premise, which is to say the procedure is indefeasible.

Is there a procedure any meticulous application of which is immune to rebutting defeaters? Yes. As we'll now argue, contrary to common assumption, abduction is such a procedure.

Abduction proceeds by assessing the extent to which each of a range of candidate hypotheses (claims or theories) satisfies certain abductive principles ('theoretical virtues'), such as principles of ontological parsimony, ideological simplicity, compatibility with other beliefs, empirical adequacy, fruitfulness, and so on. To use abduction when deciding among competing hypotheses is to infer to the truth of (more weakly: likely truth of; yet more weakly: justified belief in) the hypothesis that best

[4] Hawthorne (2007) describes a case that potentially provides an additional route to the conclusion that deduction is defeasible: "Even though I have carefully worked through a mathematical proof that *p*, I will not know *p* if I get empirical evidence that I am mad, or that human or mechanized experts have agreed that not-*p*, or that there is a priori gas in the area, or that I have made lots of mistakes using a very similar proof technique in the past, or that lots of smart people are inclined to laugh when they hear my proof. Were such experiences part of my history, then certain episodes in which I make a mistake would then (arguably) count as relevantly similar, destroying knowledge in the case at hand" (2007, 209-210). While Hawthorne invokes his case in the course of evaluating the content and usefulness of the a priori/a posteriori distinction, his assessment of the case, if correct, would imply that deduction is defeasible. That said, whether higher-order evidence can defeat first-order beliefs (as in Hawthorne's case) remains controversial (see Horowitz 2022 for relevant discussion); hence in our discussion we focus on cases in which first-order evidence defeats first-order beliefs.

explains some target explanandum or evidence, where the underlying abductive principles and their weightings determine how hypotheses are to be ranked. As Harman (1965) describes the procedure:

> In making this inference one infers, from the fact that a certain hypothesis would explain the evidence, to the truth of that hypothesis. In general, there will be several hypotheses which might explain the evidence, so one must be able to reject all such alternative hypotheses before one is warranted in making the inference. Thus one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than any other hypothesis, to the conclusion that the given hypothesis is true. […] Such a judgment will be based on considerations such as which hypothesis is simpler, which is more plausible, which explains more, which is less ad hoc, and so forth. (89)

One schematic characterization of an abductive argument (see Douven 2021 for variations on the theme) is as follows:

(1) All and only candidate hypotheses H1, H2, … Hn can explain the explanandum, F.[5]
(2) Hypothesis H_i would best explain F.
Therefore,
(3) Hypothesis Hi is true (or likely to be true, or in any case justifiably believed).

The first premise encodes the range or set of candidate explanations of the explanandum, and the second premise encodes the outcome of attention to the various abductive principles, their comparative satisfaction, and their comparative weightings, which outcomes in turn enter into the 'all things considered' judgement whereby one hypothesis is deemed to best explain (or, one might prefer: accommodate) the explanandum.

There remain, of course, questions about various aspects of abductive deliberation—for example, about how to identify (or whittle down, if need be) the range of candidate hypotheses, about which abductive principles there are and how they should be weighted, and about what to do when incompatible hypotheses are equally 'best' by lights of the operative principles and weightings. For present purposes we aim to remain neutral about how specifically these questions can or should be answered.[6] Here it is worth noting that for present purposes it doesn't matter whether there is

[5] Here and throughout, we remain neutral on the ontological category—true claim(s), fact(s), state(s) of affairs, property instance(s), etc.—of the explanandum F.

[6] For consideration of the principles at play in abduction see, e.g., Thagard 1978, Lipton 1991, 2004, Beebe 2009 (esp. 609–611), and Mackonis 2013. For attempts to formalize certain abductive principles see, e.g., McGrew 2003 and Schupbach and Sprenger 2011.

consensus, now or ever, about (in particular) abductive principles and weightings. Perhaps in the fullness of methodological time there will come to be such consensus; but if not, then we may assume that among the premises of any given application of abduction will be one registering the operative principles and weightings, with this ‘abductive premise’ encoding knowledge of the existence of, and in the usual case commitment to rejecting in the case at hand, alternative principles and/or weightings.

We also set aside general skepticism about abduction (contra, e.g., van Fraassen 1980, 2004). We rather assume that abduction is a form of reasoning that is, when properly implemented, rationally compelling or good. We see this as dialectically apropos, given the popularity of and seeming need for abduction in everyday life, in the sciences, and in philosophy, as per Lipton’s claim that “Inference to the Best Explanation is a popular account” of inference (1991, 1), Douven’s claim that “Most philosophers agree that abduction (in the sense of Inference to the Best Explanation) is a type of inference that is frequently employed, in some form or other, both in everyday and in scientific reasoning” (2021), Ladyman’s claim that “naturalists must agree that inference to the best explanation is indispensable in science” (2007, 184), and Sider’s claim that competing metaphysical positions “are treated as tentative hypotheses about the world, and are assessed by a loose battery of criteria for theory choice” (2009, 385).

Is abduction defeasible, as is commonly assumed? A procedure is defeasible if and only if some conclusion of a meticulous application of it is subject to rebutting defeaters. But, as we now argue, no conclusion of a meticulous application of abduction is subject to rebutting defeaters. To explain why this is so, we generalize from an example in which one abducts to a conclusion and then later encounters new evidence that defeats that conclusion.

> Anna considers what would best explain the fact (F) that no one has ever seen a leprechaun. She uses abduction to conclude that leprechauns do not (concretely, in what follows) exist. The following abductive principle (‘Parsimony’) informs her abduction: for any hypotheses H and H∗, if H∗ posits more kinds than H posits, and H and H∗ are otherwise equal so far as other abductive principles are concerned, then H better explains the explanandum than H∗. Anna uses this principle to reason abductively as follows: Where F is the explanandum that no one has ever seen a leprechaun, H and H* are the only candidate explanations for F, and H* is equivalent to H concatenated with the claim that there are leprechauns,
>
> > 1 H* = H plus the claim that there are leprechauns. (given)
> > 2 H* posits more kinds than H does. (1)

3 If (2), then H explains F better than H* does, unless explanatory considerations that support H* equal or outweigh the support H receives from its positing fewer kinds than H*. (Parsimony)

4 H explains F better than H* does, unless explanatory considerations that support H* equal or outweigh the support H receives from its positing fewer kinds than H*. (2, 3)

5 Explanatory considerations that support H* do not equal or outweigh the support H receives from its positing fewer kinds than H*.

6 H explains F better than H* does. (4, 5)

7 All and only H and H* can explain F. (given)

8 H is true. (6, 7; by abduction)

9 Leprechauns do not exist. (8)

> Later, Anna has perceptual experiences as of leprechauns chastising her for denying their existence. Anna then (correctly, let's suppose) retracts her conclusion that leprechauns do not exist. Accordingly, the conclusion of her (initial) meticulous application of abduction is later defeated.

If H had still explained F better than H* even given this new evidence, E, then E wouldn't have defeated H.[7] Hence, E defeats the conclusion that leprechauns don't exist only because H* explains F as well as or (in this case) better than H, given E. Anna's reasoning includes the premise—namely, 5—that explanatory considerations that support H* do not equal or outweigh the support H receives from its positing fewer kinds than H*. Accordingly, E defeats H only by defeating a premise in Anna's reasoning. This defeater, therefore, is not a rebutting defeater; rather E is an undercutting defeater, i.e., a defeater that renders a conclusion unjustified by rendering either an operative inference or premise (as in this case) unjustified. That E defeats H, then, doesn't suggest that abduction is defeasible.

This example generalizes to any possible evidence. Any evidence that defeats Anna's initial conclusion must be such that given that evidence explanatory considerations supporting ~H equal or outweigh explanatory considerations supporting H; otherwise, that evidence would not defeat H. Hence, any evidence that defeats Anna's initial conclusion must be such that it defeats a premise in the reasoning that produced that conclusion—namely, 5. This holds not only for perceptual evidence

[7] If evidence is factive, then this example presupposes that Anna's new perpetual experience is veridical. Here and throughout, we remain neutral about whether evidence is factive (cf., e.g., Williamson 2000, Mitova 2014) or not (cf., e.g., Carnap 1928, Quine 1968). Thanks to an anonymous reviewer for raising this issue.

but also for evidence that Anna incorrectly weighted some abductive principle, since such evidence would defeat Anna's conclusion only if it suggested that H is explanatorily inferior to some alternative. Anna's abductive reasoning, therefore, is not subject to rebutting defeaters.

This result generalizes to all abductive reasoning. Any meticulous application of abduction includes a premise (as per 5) encoding that considerations supporting competing hypotheses don't equal or outweigh those used to support the chosen hypothesis—call such a premise a 'catch-all premise'. To see why this premise is necessary, notice that any application that excludes a catch-all premise would implausibly lead abductors to infer the truth of the hypothesis that is explanatorily best in some respects without affirming that it is explanatorily best overall—more on this below. Accordingly, in any case in which a conclusion of a meticulous application of abduction is defeated, that conclusion is defeated only because an operative premise (the catch-all premise) is defeated. Abduction, therefore, is not subject to rebutting defeaters.

Why must any meticulous application of abduction include a catch-all premise? The theoretical virtues identified in abductive principles (e.g., qualitative ontological parsimoniousness, ideological simplicity, internal consistency) constitute explanatory goodness collectively, not in isolation. To see why, notice that appealing to qualitative ontological parsimony alone would implore abductors to affirm (in every case, and regardless of competing considerations) the most ontologically parsimonious hypothesis, according to which there is only one or perhaps even no existing object! Even if such a hypothesis is true, no one intends to establish monism or ontological nihilism simply by endorsing a principle of parsimony—a point reflected both in Ockham's dictum that one shouldn't multiply entities *beyond necessity* and in the ceteris paribus clause that appears in Parsimony (which recurs in any plausible version of an abductive principle—see Biggs and Wilson 2017 and 2021 for discussion). Because the characteristics identified in abductive principles constitute explanatory goodness only collectively, abductive principles must function collectively, with any application of abduction affirming that the chosen hypothesis does best across that collective, not merely on the principles noted in that application. The catch-all premise simply reflects this holistic nature of abduction.

To put the point briefly: any meticulous application of abduction uses at least one abductive principle by way of establishing the explanatory superiority of one hypothesis (or a set of hypotheses) over any competitor; moving from the premise that a given abductive principle, P, supports a hypothesis, H, to the conclusion that H best explains the explanandum requires affirming that explanatory considerations that support competing hypotheses don't equal or outweigh the support that H receives from P; so, any meticulous application of abduction includes a catch-all premise.

To be sure, abductors don't always make a catch-all premise explicit, as when philosophers and scientists claim that a given hypothesis is superior to competitors on grounds that it is simpler, without further claiming that the chosen hypothesis is moreover superior all things considered. But any such abductor assumes a catch-all premise, since no advocate of abduction thinks that it confers justification on the hypothesis that performs best on a subset of abductive principles even though that hypothesis is explanatorily inferior to some alterative when all abductive principles are accounted for.

Notice that the catch-all premise isn't restricted to the evidence one has when performing the abduction. If the catch-all premise were so restricted, then E would be a rebutting defeater rather than an undercutting defeater for Anna's initial conclusion that leprechauns don't exist—since 5 would be replaced with 5*: Given the evidence at hand, explanatory considerations that support H* do not equal or outweigh the support H receives from its positing fewer kinds than H*. Contrary to this restriction, no one thinks that one should infer the truth of the hypothesis that best explains an unrepresentative subset of evidence, any more than anyone thinks that one should believe the simplest theory, regardless of competing considerations; rather, advocates of abduction think that one should infer the truth of the hypothesis that best explains all evidence.

It doesn't follow that justifiably believing the catch-all premise in a given application of abduction requires exploring all evidence that could bear on the hypotheses at issue. Such justification only requires justifiably judging that further investigation of competing hypotheses is unlikely to reverse the results of one's present inquiry. We take it that abductors often justifiably make such judgements, even though they rarely, if ever, exhaustively explore all evidence or apply every abductive principle. At any rate, since any meticulous application of abduction must include a catch-all premise, our assumption that general skepticism about abduction is false implies that either (as we think) justifiably believing the catch-all premise doesn't require exploring all evidence or exploring all evidence is often (in practice) possible for ordinary abductors.

Finally, notice that one can't make any procedure indefeasible by simply adding an analogue of the catch-all premise to it. To see why, consider what (enumerative) induction's analogue of a catch-all premise would be and what would happen were induction to include that analogue. One possible analogue is this: no explanatory consideration that one did not address in that application of induction would overturn the result of that application had it been included. Requiring any meticulous application of induction to include this premise would transform it into abduction—such that induction would justify whichever conclusion best explains relevant evidence. Another possible analogue holds that observing entities that one did not observe in an application of induction wouldn't overturn the conclusion of that application. Requiring any meticulous application of induction to include this premise would transform induction into deduction: from the premises that all observed xs have been Fs and that no unobserved xs are not Fs, one can deduce that all xs are Fs. Relatedly, abduction would serve a purpose even for a reasoner who has observed the history of the universe, allowing the abductor to discern which explanations she should believe given those observations, while induction would serve no purpose, since she would already know what holds outside any given sample prior to applying induction. More generally, whereas (as above) any application of abduction excluding the catch-all premise is broadly irrational, applications of induction excluding any analogue are rational (skepticism about induction aside).

In sum: reflecting that abduction is holistic and the corresponding nature of abduction principles, any meticulous application of abduction includes a catch-all premise; therefore, no such application is subject to rebutting defeaters; therefore, contrary to common assumption, any meticulous application of abduction is indefeasible.

Nonetheless, abduction is (as per usual) fallible (not truth-preserving), since a best explanation can be false. Abduction is (deductively) invalid, no matter how well it is performed: even in the ideal case where the abductor explores all relevant hypotheses, accounts for all evidence, deploys all abductive principles, and weights all principles correctly, thereby identifying the hypothesis that is the unique best explanation of the explanandum, that hypothesis can be false. This fallibility is sometimes blamed on the contingency of abductive principles and the corresponding purported a posteriority of abduction. Contrary to that assessment, we take the fallibility of abduction to be orthogonal to the status of abductive principles as necessary or contingent and the status of abduction as a priori or a posteriori. Indeed, despite affirming the fallibility of abduction, we think that abduction is an a priori mode of inference (see Biggs and Wilson 2017, 2020; see Biggs and Wilson 2021 for arguments that Kant and Carnap agree with our assessment; see also Bonjour 1998, Swinburne 2001, Peacocke 2003).

We conclude, then, that abduction is reasonably taken to be both indefeasible and fallible.

Might there be a procedure different from abduction which is both indefeasible and fallible? Our discussion so far suggests that a procedure is indefeasible if and only if it immune to rebutting defeaters. We find it doubtful that any two procedures that can deliver competing results can both be immune to rebutting defeaters; for since the procedures can deliver competing results, the results of (at least) one should be capable of defeating results of the other. Accordingly, we find it doubtful that procedures that can deliver competing results can both be indefeasible.

That said, it might be that multiple procedures that invariably deliver the same results (and hence are such that the results of one are incapable of defeating those of another) are all immune to rebutting defeaters. This might be the case with abduction and other procedures that are plausibly taken to be holistic, especially the method of reflective equilibrium (which plausibly requires using all relevant considerations to find the hypothesis that best fits all one's evidence) and (appropriate) Bayesian inference (which plausibly requires using all relevant considerations (encoded in priors and likelihoods) to establish the probability of a theory given some evidence). In articulating the method of reflective equilibrium, Daniels (2020) talks not of mere coherence but rather of "acceptable coherence", which "requires that our beliefs not only be consistent with each other [. . .] but that some of these beliefs provide support or provide a best explanation for others"; and various authors suggest that Bayesian inference and abduction are deeply connected, perhaps because the former incorporates abductive principles (Weisberg 2009), perhaps because abduction is simply Bayesian inference described differently (Henderson 2014).[8] At any rate, we doubt that a procedure that is distinct from abduction and delivers competing results could be indefeasible, since those results could (in principle) serve as rebutting defeaters for a meticulous application of abduction, which is (as above) immune to such defeaters.

4. An explanatory diagnosis

We have argued that, contrary to common assumption, neither defeasibility nor fallibility implies the other. More specifically, we have argued that defeasibility and fallibility are both definitionally inequivalent (since the former should be defined by appeal something epistemic—namely, justification-preservation—while the latter should be defined by appeal to something metaphysical—namely, truth-preservation) and extensionally inequivalent (since deduction is infallible and defeasible and abduction is fallible and indefeasible). Why, then, have defeasibility and fallibility been commonly taken to imply each other? We offer two speculative diagnoses.

First, the conflation might be due, in part, to a too-quick generalization from the case of (enumerative) induction. In our earlier example, the conclusion that all swans are white, inferred from the premise that all observed swans were white, was defeated by an encounter with a black swan. In this and many standard cases of induction, defeat occurs via the conclusion's being shown to be false, notwithstanding the truth of the premises. But then discovering the state of affairs—the premises true and the conclusion false—that makes induction invalid also shows that it is defeasible. And so, one might naturally come to suppose more generally that—at least in the absence of reasons to think

[8] See Biggs and Wilson (in progress) for further discussion.

otherwise—defeasibility and fallibility, even if not definitionally equivalent, are in any case equivalent, such that (in particular) deduction, being infallible, must thereby be indefeasible, and abduction, being fallible, must thereby be defeasible.

But the generalization, while natural enough, is too hasty. Recall the case of Dina, who initially deduces that the wall probably is red and later deduces that it probably is not red. In this case, one conclusion defeats the other—and perhaps each defeats the other—but neither shows that the other is false; for even if the first conclusion is in fact true, as we might suppose the conclusion of Dina's initial reasoning is, considerations that are orthogonal to the reasoning that led to that conclusion might serve to reveal that the conclusion is unjustified, notwithstanding the truth of the premises. Hence it is that deduction is defeasible, despite being infallible; more generally, hence it is that any procedure unable to appropriately accommodate every consideration bearing on the justification of the conclusions of that procedure will be defeasible, whether it is infallible (as deduction) or not (as induction). Conversely, if a procedure is capable of appropriately accommodating all such considerations, it will thereby be indefeasible, even if it is fallible (as abduction)—since it will account for anything that could defeat its conclusions.

Ultimately, then, the neat alignment between fallibility and defeasibility present in the case of induction doesn't generalize; but it takes recognition of an underappreciated (albeit common) route to defeat to see just how this can be.

Second, the conflation of defeasibility and fallibility might be due, in part, to a conflation of epistemic and metaphysical notions. Recall Kripke's observation about necessity and a priority:

> It's certainly a philosophical thesis, and not a matter of obvious definitional equivalence, either that everything a priori is necessary or that everything necessary is a priori. Both concepts may be vague . . . But at any rate they are dealing with two different domains, two different areas, the epistemological and the metaphysical. (Kripke 1980, 36)

Similarly, defeasibility and fallibility are "dealing with two different domains, two different areas, the epistemological and the metaphysical", and it is therefore "a philosophical thesis, and not a matter of obvious definitional equivalence" that all defeasible reasoning is fallible nor that all fallible reasoning is defeasible. Nonetheless, we suspect that many epistemologists who have assumed this equivalence have thought it to be definitional—indeed, the tendency to introduce defeasible reasoning as that which relies on an invalid argument or that which is not deductive suggests as much.

5. Abduction as the ultimate arbiter of disputes

We close by noting that abduction's being indefeasible plausibly gives it a special epistemic status: namely, abduction's being indefeasible plausibly makes abduction an ultimate arbiter of disputes in any domain in which it is operative. As a first pass, a procedure is an ultimate arbiter of disputes in a domain if and only if in that domain one should follow that procedure at the end of the day, i.e., regardless of what any competing considerations suggest. More carefully, a procedure P is the ultimate arbiter of disputes in a domain D if and only if for any claim q in D, one who concludes that q through a meticulous application of P, and is justified in believing the premises of that application, is thereby justified in believing q, and moreover is unjustified in believing ~q or withholding belief in q, regardless of what any distinct procedure suggests. As above, a procedure is indefeasible if and only if no conclusion of a meticulous application of it can be defeated without defeating an operative premise, which is to say that one who is justified in believing its premises is justified in believing its conclusions, regardless of the dictates of any alternative procedure. Accordingly, a procedure is an ultimate arbiter of disputes in a domain if and only if it is indefeasible in that domain. Abduction, then, is plausibly taken to be an ultimate arbiter of disputes in any domain in which it is operative, precisely because it is indefeasible. Whether abduction is moreover *the* ultimate arbiter of disputes in any domain in which it is operative depends on whether any procedure other than abduction invariably converges with abduction—a project we leave for another time.

Acknowledgements

Thanks to an anonymous reviewer for this volume for helpful comments. Thanks also to David Alexander, Ranpal Dosanjh, and participants in the 2023 'Inductive Metaphysics: Insights, Challenges and Prospects' workshop at the University of Dusseldorf (especially Cruz Davis, Siegfried Jaag, Markus Shrenk, and Gerhard Shurz), for useful discussion of portions of this material.

References

Beltzer, Marvin and Barry Loewer. 1988. A Conditional Logic for Defeasible Beliefs. *Decision Support Systems* 4: 129-142.

Biggs, Stephen, and Jessica Wilson, 2017. The A Priority of Abduction. *Philosophical Studies* 174:735–758.

Biggs, Stephen, and Jessica Wilson, 2020. Abductive two-dimensionalism: a new route to the a priori identification of necessary truths. *Synthese* 197:59–93

Biggs, Stephen, and Jessica Wilson, 2021. Abduction versus conceiving in modal epistemology. *Synthese* 198: S2045–S2076.

Biggs, Stephen, and Jessica Wilson. In progress. Abduction as Ultimate Arbiter of Modal (and Other)

Disputes.

Beebe, James R. 2009. The Abductivist Reply to Skepticism. *Philosophy and Phenomenological Research* 79:605–636.

Bonjour, Lawrence, 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Carnap, Rudolph. 1928/1967. *The Logical Structure of the World: Psuedoproblems in Philosophy*. Translated by Rolf George. University of California Press.

Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.

Chisholm, Roderick. 1966/1989. *Theory of Knowledge*. Englewood Cliffs, New Jersey: Prentice Hall

Daniels, Norman, 2020. Reflective Equilibrium. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Summer 2020 edition.

Douven, Igor, 2021. Abduction. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Summer 2021 edition.

Fraassen, Bas C. Van, 2004. *The Empirical Stance*. New York: Yale University Press.

Genin, Konstantin and Franz Huber, 2022. Formal Representations of Belief. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Fall 2022 edition.

Harman, Gilbert H., 1965. The Inference to the Best Explanation. *Philosophical Review* 74:88–95.

Hawthorne, John, 2007. A Priority and Externalism. In I*nternalism and Externalism in Semantics and Epistemology*, edited by Sanford C. Goldberg, 201–218. Oxford University Press.

Henderson, Leah, 2014. Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science* 65: 687-715.

Horowitz, Sophie. 2022. Higher-Order Evidence. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Summer 2022 edition.

Huemer, Michael. 2001. The Problem of Defeasible Justification. *Erkenntnis* 54: 375-97.

Jaszczolt, Katarzyna M., 2022. Defaults in Semantics and Pragmatics. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Summer 2022 edition.

Koons, Robert, 2022. Defeasible Reasoning. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Summer 2022 edition.

Kripke, Saul, 1971. Identity and Necessity. In *Identity and Individuation*, edited by M. K. Munitz, 135–64. New York: New York University Press.

Kripke, Saul. 1980. *Naming and Necessity*. Princeton: Princeton Univeristy Press.

Ladyman, James, 2007. Does Physics Answer Metaphysical Questions? *Royal Institute of Philosophy Supplement* 61: 179–201.

Lipton, Peter. 1991. The Best Explanation. *Cogito* 5: 9–14.

Lipton, Peter, 2004. *Inference to the Best Explanation*. Routledge/Taylor and Francis Group.

Mackonis, Adolfas, 2013. Inference to the Best Explanation, Coherence and Other Explanatory Virtues. *Synthese* 190: 975–995.

McGrew, Timothy. 2003. Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science* 54:553–567.

Mitova, Veli. 2014. Truthy Psychologism about Evidence. *Philosophical Studies* 172: 1105-1126.Peacocke, Christopher, 2003. *The Realm of Reason*. Oxford University Press.

Pollock, John. 1987. Defeasible Reasoning. *Cognitive Science* 11: 481-518.

Quine, W.V.O. 1968. Epistemology Naturalized. In *Ontological Relativity and Other Essays*, by W.V.O. Quine, 69–90. New York: Columbia University Press.

Schupbach, Jonah N. and Jan Sprenger. 2011. The Logic of Explanatory Power. *Philosophy of Science* 78:105–127.

Sider, Theodore, 2009. Ontological Realism. *In Metametaphysics: New Essays on the Foundations of Ontology*, edited by David John Chalmers, David Manley, and Ryan Wasserman, 384–423. Oxford: Oxford University Press.

Strasser, Christian and G. Aldo Antonelli, 2019. Non-monotonic Logic. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Summer 2019 edition.

Sturgeon, Scott. 2014. Pollock on Defeasible Reasons. *Philosophical Studies* 169: 105-118.

Swinburne, Richard, 2001. *Epistemic Justification*. Oxford University Press.

Thagard, Paul R., 1978. The Best Explanation: Criteria for Theory Choice. *Journal of Philosophy* 75:76–92.

van Fraassen, Bas, 1980. *The Scientific Image*. Oxford: Oxford University Press.

Weisberg, Jonathan, 2009. Locating IBE in the Bayesian Framework. *Synthese* 167:125-143.

Williamson, Timothy, 2000. *Knowledge and its Limits*. New York: Oxford University Press.