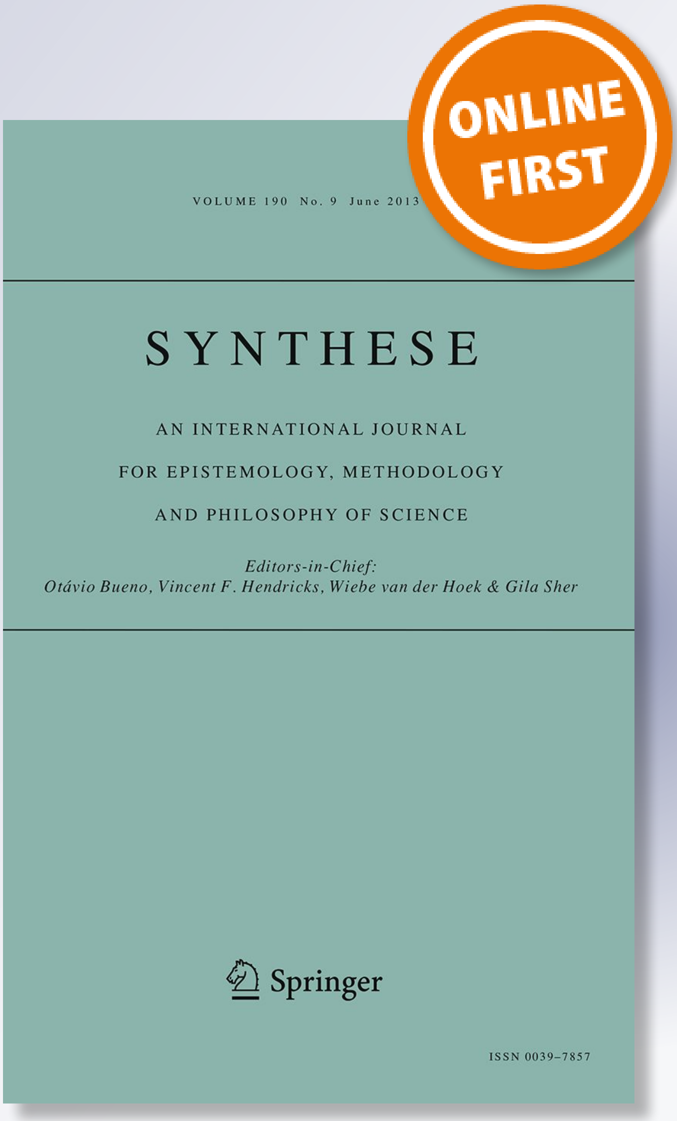
Inference, explanation, and asymmetry

Kareem Khalifa, Jared Millson & Mark Risjord

Synthese
An International Journal for Epistemology, Methodology and Philosophy of Science

ISSN 0039-7857

Synthese
DOI 10.1007/s11229-018-1791-y



Springer

Synthese
https://doi.org/10.1007/s11229-018-1791-y

S.I.: INFERENTIALISM

Inference, explanation, and asymmetry

Kareem Khalifa[1] · **Jared Millson**[2] · **Mark Risjord**[3]

Received: 30 August 2016 / Accepted: 20 April 2018

Abstract Explanation is asymmetric: if A explains B, then B does not explain A. Traditionally, the asymmetry of explanation was thought to favor causal accounts of explanation over their rivals, such as those that take explanations to be inferences. In this paper, we develop a new inferential approach to explanation that outperforms causal approaches in accounting for the asymmetry of explanation.

Keywords Explanation · Inference · Symmetry problem

1 Introduction

A surefire way to embarrass a theory of explanation is to show that it fails to respect the commonsense idea that explanation is an asymmetric relation. Give or take some rare exceptions, if A explains B, then B does not explain A. In the lore of philosophical accounts of explanation, the origin myth almost always includes reference to a flagpole and its shadow.

The symmetry problem has served as an expedient way to disqualify a position that we call *Explanation-As-Inference* (hereafter: EAI). For instance, Bromberger

✉ Jared Millson
jmillson@agnesscott.edu

Kareem Khalifa
kkhalifa@middlebury.edu

Mark Risjord
mrisjor@emory.edu

[1] Middlebury College, 303 Twilight Hall, 14 Old Chapel Rd, Middlebury, VT 05753, USA

[2] Agnes Scott College, 303 Buttrick Hall, 141 E. College Ave, Decatur, GA 30030, USA

[3] Emory University, S414 Callaway Hall, 201 Dowman Drive, Atlanta, GA 30322, USA

Published online: 28 April 2018

(1965) used symmetry to critique Hempel's (1965) covering-law model of explanation. Similarly, Kitcher's (1989) unificationist theory purported to restore explanation's asymmetry, but it faced other, searching symmetry problems (see Barnes 1992). By contrast, the symmetry problem has been a great advertisement for causal approaches to explanation. According to these views, explanation's asymmetry follows effortlessly in the wake of causation's asymmetry. Given the long shadow that the symmetry problem casts, it is no wonder that causal approaches to explanation seem to enjoy a privileged status in contemporary philosophy of science (Strevens 2008; Woodward 2003).

Despite their prominence, causal theories of explanation face their own challenges with respect to the asymmetry of explanation. A growing body of literature shows that some scientific explanations are noncausal (see Reutlinger 2017). While the mere existence of these explanations challenges causal theories, noncausal explanations also raise a further, hitherto unnoticed, problem. Noncausal explanations exhibit asymmetries. This suggests that the ultimate source of explanatory asymmetry may not be causal, and it undermines an important dialectical motivation for adopting a causal theory. Hence, the holy grail would be an analysis that accommodates causal and noncausal explanations, and accounts for the asymmetries of both.

In this essay, we shall offer a new set of necessary conditions for explanation that provides a unifying framework for handling asymmetries. While we consider ourselves members of the EAI family, we make bold departures from our predecessors. Section 2 specifies the contours of the symmetry problem when both its causal and noncausal variants are taken on board. Section 3 then presents our new approach, which we call the defeasible inference model of explanation (DIME). Section 4 shows how DIME solves some of these symmetry problems without adverting to causal concepts. Finally, Sect. 5 considers a tougher challenge to DIME, in which causal concepts must be invoked in order to capture the relevant asymmetries. We argue that even though causation figures prominently in solving this last class of symmetry problems, it does so in a way that supports, rather than diminishes, the thesis that inference is the ultimate locus of explanation.

2 The symmetry problem

Historically, the symmetry problem was pitched as a challenge to Hempel's deductive-nomological (DN) model of explanation. The DN Model was one of three versions of Hempel's covering law theory, along with inductive-statistical (IS), and deductive-statistical (DS) explanations. In all three, the explanandum is inferred from premises. The premises, in turn, include at least one law of nature. A DN explanation of E is a sound deductive argument of the form $C_1, \ldots, C_k, L_1, \ldots, L_r \vdash E$, where $L_1, \ldots, L_r$ are universal laws of nature and $C_1, \ldots, C_k$ are initial conditions. Furthermore, all premises must be indispensable to the argument's validity.

While most discussions of the symmetry problem involve a flagpole and its shadow, Sylvain Bromberger's original example is somewhat more colorful:

> There is a point on Fifth Avenue, M feet away from the base of the Empire State Building, at which a ray of light coming from the tip of the building makes

> an angle of θ degrees with a line to the base of the building. From the laws of geometric optics, together with the "antecedent" condition that the distance is M feet, the angle θ degrees, it is possible to deduce that the Empire State Building has a height of H feet. Any high-school student could set up the deduction given actual numerical values. By doing so, he would not, however, have explained why the Empire State Building has a height of H feet…(Bromberger 1966, p. 92).

At the risk of belaboring what any high-school student could do, two inferences are involved in Bromberger's example:

$$\tan\theta = \frac{H}{M},\ \theta = 60°,\ H = 1,454\,\text{ft} \ \vdash\ M = 839.5\,\text{ft} \qquad \text{(Classic Tower)}$$

$$\tan\theta = \frac{H}{M},\ \theta = 60°,\ M = 839.5\,\text{ft} \ \vdash\ H = 1,454\,\text{ft} \qquad \text{(Classic Shadow)}$$

Since both inferences are deductions from a law and initial conditions, both count as explanations according to Hempel's criteria. But as Bromberger points out, only the first is plausibly an explanation. The problem easily generalizes. Many physical laws are expressed by equations treating one variable as a function of others. These laws permit the value of any one variable to be deduced from the values of the others. Like the building and its shadow, only some of these inferences are explanations.

Subsequent proponents of EAI tried to rule out the asymmetries by further restricting the inferences that count as explanations. For instance, where Hempel required only that the explanandum be validly deduced from the explanans, Kitcher's unificationist model required that the explanandum be derived in a particular way.[1] Each explanatory derivation is an instance of a more abstract argument pattern that specifies the kinds of premises and inference-rules that may be used to derive the explanandum. A *systemization* of the corpus of accepted statements K is any set of general argument patterns that derive some members of K from others. Explanation consists of using instances of the "best" systemization, *E(K)*, as measured according to the following criteria:

1. *Acceptability* Each step of each instance of a general argument pattern in *E(K)* must be deductively valid, and acceptable relative to *K*.
2. *Scope* Unification increases in proportion to the size of the conclusion set of the number of acceptable instances of *E(K)*.
3. *Stringency* Unification increases in proportion to the strictness of the argument patterns in *E(K)*.
4. *Number of patterns* Unification decreases in proportion to the number of general argument patterns in E(K).

Kitcher's view accounts for the asymmetry in Bromberger's example. Kitcher proposes that our present explanatory store contains an "origin and development" (OD)

[1] Other unificationists who endorse EAI include Bangu (2016), Friedman (1974), Schurz and Lambert (1994), Schurz (1999). Space prohibits extensive discussion of their views on symmetry.

argument pattern, according to which spatial dimensions of any physical object are derived from its origin and subsequent physical changes. Kitcher then invites us to consider a "shadow" (S) argument pattern, whereby the spatial dimensions of physical objects are derived from the length of their shadows. Classic Shadow would be an instance of S. Simply adding S to our explanatory store runs afoul of the fourth criterion of unification, above; it will therefore only be explanatory if it fares better along one of the other dimensions of unification. However, it appears that the acceptable conclusions that S generates are a proper subset of those that OD generates. Nor does S appear any more stringent than OD. Hence, Kitcher's unificationism is not susceptible to shadowy symmetries because the non-explanatory inferences do not unify a scientific domain.

Nevertheless, Eric Barnes' (1992) version of the symmetry problem threatens Kitcher's version of EAI. Barnes imagines a closed system of "Newtonian particles." Given a complete description of this system at a time, the state of the system at any later time can be determined through Newton's laws. Barnes argues that the deduction of future states of the system from the present state satisfies Kitcher's criteria for explanation. Newtonian mechanics, after all, is a paradigm of scientific unification. But Newton's laws also permit the deduction of past states from present states. So, since Newtonian mechanics fits Kitcher's criteria, both the forward and backward calculations must count as explanatory. But clearly, retrodicting past states from the present does not count as an explanation.

The most prominent diagnosis of explanatory asymmetries holds that inferences such as Classic Tower are explanatory because they track causal relationships, while the non-explanatory ones like Classic Shadow do not. Since nothing in the form of inference marks a causal relationship, many causal theorists of explanation argue that the inferences are superfluous, or, at the very least, are subservient to the more basic explanatory task of tracking causes. On such a view, all explanations represent causal relationships, and explanation is asymmetrical because causes are. EAI appears to have hit a dead end.

However, we should be suspicious of any diagnosis that identifies the asymmetry of explanation with the asymmetry of causation. Begin with the observation that some explanations are noncausal. For instance:

> Consider the fact that at every moment that Earth exists, on the equator (or on any other great circle) there exist two points having the same temperature that are located antipodally (i.e., exactly opposite each other in that the line between them passes through the Earth's center). Why is that? (Lange 2016, 7).

The answer to this question does not invoke local meteorological causes of each location's temperature. This would explain why each antipodal point has the temperature it does, but not why there *must* be two antipodal points with the same temperature. For that, we must appeal to a mathematical fact, the intermediate value theorem. This theorem states that for any real-valued, continuous function f, and real number u between $f(a)$ and $f(b)$, there exists a value $c \in [a, b]$ such that $f(c) = u$. For example, the square function is real-valued and continuous. According to the intermediate value theorem, this means that for any real number u between a^2 and b^2, there exists a number $c \varepsilon [a, b]$ such that $c^2 = u$. Surprisingly, this fact about real-valued, continuous functions tells us something about equatorial temperatures.

To explain why a pair of antipodal points on the equator must have the same temperature, consider the function, $D(x)$, which is simply the difference between the temperature at x and its antipode, x'. Specifically, letting $T(x)$ represent the temperature at x, we can say $D(x) = T(x) - T(x')$. Since the temperature on the equator is a real-valued, continuous function, so is the difference between the temperature of any equatorial point and its antipode. Now, consider any point x on the equator that is warmer than its antipode x'. Since $T(x) - T(x')$ will be a positive number, $D(x) > 0$. Since $T(x') - T(x)$ will be a negative number, $D(x) > 0 > D(x')$. Of course, had we begun with a point cooler than its antipode, then the signs would have changed: $D(x) < 0 < D(x')$. So we can say quite generally that for any point, x and its antipode x', 0 is between $D(x)$ and $D(x')$. The intermediate value theorem then entails that there must be some point c where $D(c) = 0$. But this would be a point that has the same temperature as its antipode.[2]

Quite clearly, this explanation is asymmetric. The explanans of the foregoing explanation consists of the intermediate value theorem and the fact that temperature—and hence $D(x)$—is a continuous real-valued function. The explanandum is the fact that for any of the Earth's great circles, there are always two points with the same temperature. Obviously, this topological feature of the Earth's temperatures does not explain either the intermediate value theorem or the mathematical characteristics of $D(x)$. Moreover, the explanans is not a cause of the explanandum. Hence, contrary to the received wisdom, a failure of inferences to track causes is not the source of the explanatory asymmetry.

Indeed, the asymmetry is readily captured by inferential considerations alone. Only the explanation involves a valid inference, while transposing the explanandum with either the intermediate value theorem or the statement about $D(x)$ would produce a fallacy. Thus, not only is the asymmetry not causal, it is decidedly inferential.

To summarize, it appears that there are a variety of explanatory asymmetries and they are susceptible to different kinds of solution. Some symmetry problems, such as Bromberger's, appear to be solvable by either inferential or causal means. Others, such as Barnes' example of the Newtonian particles, appear to favor causal approaches over EAI. Still others, such as the asymmetry involving equatorial temperature, appear to favor EAI over any causal approach. Thus, contrary to the prevailing dogma, it is far from clear that causal accounts have gotten to the root of explanatory asymmetries. We take this opening as an opportunity to reenvision EAI. The net result will be one in which all three kinds of asymmetries can be seen to spring from a common inferential fountainhead.

3 The defeasible inference model of explanation (DIME)

Suppose, like us, that you are sympathetic to EAI, and want to solve these symmetry problems. Where did others go wrong? A provocative hint for solving symmetry

[2] For defenses for why this and other mathematical derivations of empirical facts are explanatory, see, e.g. Colyvan (1998) and Lange (2016).

problems is found in Kitcher's solution to Bromberger's problem. Kitcher shows how an inference's *comparative* failings disqualify it as an explanation. In particular, it seems as if the proper explanation of the tower's height—say an architect's design—will succeed where the shadow "explanation" fails. Similarly, many *causal* approaches to explanation hold that CLASSIC SHADOW is not explanatory because, when we hold the architect's design fixed, the tower's height would still be 1454 feet, even if the shadow's length were not 839.5 feet. As we have seen, gaps remain in both EAI and causal approaches. This suggests that they may be deploying the wrong basis of comparison.

We provide a new way of explicating the insight that a proper explanation succeeds where its competitors fail. Specifically, DIME holds that A explains B only if:

1. A and B are (approximately) true,
2. B is a nontrivial consequence of A,
3. the appropriate nontrivial inferences are bracketed, and
4. the inference from A to B succeeds where all others fail.

We will say that inferences satisfying the last three conditions are "sturdy." Importantly, in this paper we only provide necessary conditions for "A explains B." Hence, our analysis is only partial. This will not matter in what follows, since we will have provided enough of an analysis to solve the symmetry problem within an EAI framework. In future work, we intend to complete this analysis. To provide a better sense of DIME, Sections 3.1 through 3.4 discuss each of the four conditions in turn.

3.1 Explanation and truth

The least remarkable of our requirements for explanation is that the explanans and explanandum must be *true*. This conforms to common usage, where a false proposition is not the actual explanation. Presumably, several alternative accounts of "quality control" on the explanans and explanandum—e.g. involving different theories of truth, or appealing to significantly different semantic or epistemic properties than truth—can be wedded to Conditions 2 through 4 (i.e. sturdiness), and still furnish similar solutions to the symmetry problem, so we will mostly take this requirement for granted in what follows.

We have parenthetically added that the explananans and explanandum may be *approximately* true. For many scientific explanations, the premises are known, strictly speaking, to be false (Cartwright 1983). For instance, it would be miraculous if the Empire State Building stood at a *perfect* 90° angle to 5th Avenue, *exactly* at 1454 feet, etc. Nonetheless, the building's height explains the shadow's length for roughly the reasons implied by CLASSIC TOWER. In Sect. 3.2, we sketch how approximation and defeasibility pair naturally with each other.

3.2 What is a nontrivial consequence?

For our purposes, nontrivial consequences have three key features. Each of them maps on to properties of explanation. First, nontrivial inferences are *irreflexive*. This accords with the idea that explanation is also an irreflexive relationship, e.g., that the shadow being 839.5 feet long does not explain why it is 839.5 feet long.

Second, nontrivial inferences are *premise consistent*. Since a contradiction explains nothing, the classically valid inference pattern *Ex Falso Quodlibet* (where a contradiction entails any proposition) cannot be an explanation. As its name suggests, premise consistency requires that the premises of a nontrivial consequence be consistent.

Third, nontrivial inferences are *defeater-sensitive*, i.e. there are conditions under which the inference ceases to be good. For instance, suppose we explain a person's symptoms by appeal to her disease. In such a case, one may infer the symptoms from the presence of the disease. However, if we found out that the person was taking an effective treatment—one that would eliminate the symptoms—then the inference from the disease to the symptoms is no longer acceptable. We would look elsewhere for the explanation.

One way to capture defeater-sensitivity is to make the consequence relationship defeasible. Defeasible consequence relations are disrupted when certain additional propositions—called *defeaters*—are considered. We will represent defeasible inferences this way:[3]

$\Sigma \mid$ The patient is infected with the*Varicella zoster*virus (CHICKENPOX)

$\mathrel{|\!\!\!\sim}_{\Theta}$ The patient's skin is covered in red spots

Here, "Σ" denotes a set of background conditions, which we discuss at greater length in Sect. 3.3.[4] Additionally, "$\mathrel{|\!\!\!\sim}_{\Theta}$" denotes a defeasible consequence relationship. "Θ" denotes a set of defeaters, e.g. "The patient has received effective treatment." Roughly, a good defeasible inference of this sort will turn bad if members of Θ are true.[5]

To capture the defeater-sensitivity of explanation, we will be treating nontrivial consequence relationships as defeasible in the manner just described. This is an important departure from earlier proponents of EAI. The consequence relation of classical logic is not defeasible in the sense articulated here. If the sequent $A \vdash B$ is classically valid, no new information will disrupt the inference from A to B. However, one could also capture the defeater-sensitivity of explanation using classical logic by including true information that various defeaters do not obtain in one's premises. Since no technical

[3] Throughout this paper we will use capital Roman letters for sentences of a formal language and capital Greek letters for sets of these sentences.

[4] The graphical separation of Σ from the premises of a defeasible inference via "$\mid$" is intended to remind the reader that the conclusion of the inference follows from the information contained in the premises and *not* from the contents of the background set. The latter figures solely in the determination of the inference's defeat-status.

[5] More precisely, an inference is defeated whenever its premises or background set contain a sentence that is logically equivalent to a member of the defeater set. This characterization of defeat allows for defeaters that are false to be considered in the evaluation of inferences—what we call *expedient* defeaters below. Although we refrain from providing it here, all of our informal references to *defeat* in the present text conform to the formal definition given in Millson et al. (2018). Two features of this formal definition are worth pointing out. First, a disjunction defeats an inference if both the disjuncts (or their logical equivalents) belong to the defeater set, and, second, a conjunction defeats the inference if at least one of the conjuncts (or their logical equivalents) belongs to the defeater set.

details about one's choice of logical system figure prominently below, readers more inclined toward explanatory deductivism are free to slide information contained within Θ into the premises.

One might object to this irenic picture we have suggested. In particular, explanations involving mathematically formulated physical laws walk and talk like classically valid inferences. Explanations in mathematized sciences, one might argue, are indefeasible through and through. This objection overlooks the way that modeling practices—such as idealization, abstraction, *ceteris paribus* clauses, and approximation—"hide" the defeasibility of explanation. Consider once again the approximations involved in the explanation of the shadow. If the building is too far from being perpendicular, the inference will not go through. The *defeater set*, Θ, captures just this kind of modeling consideration. Moreover, this set includes many of the model's other limitations. For instance, the law of geometric optics involves an idealization (light behaves as a ray) that breaks down under certain conditions (e.g. in quantum systems.) For deductivists to capture these aspects of modeling, they would have to add very similar information to CLASSIC TOWER's premises, and tweak its form so as to preserve its validity. Hence, treating explanations involving mathematically formulated laws as defeasible inferences is no worse than treating them as deductions, and might well be advantageous.

For these reasons, we will represent the inferences in Bromberger's example as:

$$\Sigma \mid \tan\theta = \frac{H}{M},\ \theta = 60^{\circ},\ H = 1,454\,\text{ft} \mathrel{\substack{\vert\!\sim \\ \Theta}} M = 839.5\,\text{ft} \qquad \text{(TOWER)}$$

$$\Sigma \mid \tan\theta = \frac{H}{M},\ \theta = 60^{\circ},\ M = 839.5\,\text{ft} \mathrel{\substack{\vert\!\sim \\ \Theta'}} H = 1,454\,\text{ft} \qquad \text{(SHADOW)}$$

Hereafter, we assume that for *us* to solve Bromberger's symmetry problem, we must show that TOWER is explanatory, but SHADOW is not.

3.3 Which nontrivial inferences should be bracketed?

DIME follows many other theories of explanation in recognizing the importance of background conditions for explanations. Our background set, Σ, plays two crucial roles: delimiting the set of explanations that compete for sturdiness and setting up a "level playing field" on which they compete. We discuss the first of these roles here. Section 4.1 discusses how background conditions level the explanatory playing field.

Cursory reflection reveals that any explanandum is the conclusion of an ungodly number of nontrivial inferences. Fortunately, many of these inferences can be "bracketed." For a given explanandum B, C is bracketed just in case the nontrivial inference from C to B is defeated by some proposition in the background set Σ.[6] For instance, consider an inference of the shadow's length from a set of premises describing a taller tower standing behind the Empire State Building. (Call this the TALLER TOWER

[6] In the formalism above, C can be bracketed with respect to B if and only if $\Sigma \mid C \mathrel{\substack{\vert\!\sim \\ \Theta}} B$ and there exists a sentence $D \in \Theta$ and $D \in \Sigma$.

inference.) Since no such tower exists, Taller Tower is defeated, and hence can be bracketed as a serious candidate for an explanation. It actually is bracketed if the fact that no taller tower exists is taken for granted.

Concepts kindred to bracketing have a long (if sometimes subterranean) history in theorizing about explanation. For instance, *ceteris paribus* clauses are often best interpreted as demands to bracket a class of potential explanatory factors. Admittedly, some accounts of *ceteris paribus* laws will not fit the specifically inferential mold that we have cast, but nevertheless play a similar role in "ruling out" ertswhile explanations. The host of positions about these laws reflects the diversity of positions in the explanation literature (see Reutlinger et al. 2015). This diversity, in turn, reflects the fact that theorists of explanation with otherwise opposing theoretical orientations feel obliged to account for how bracketing (in a broad sense) figures in explanation.

Bracketing has two important dimensions: the structure of defeat and the truth-value of the defeater. The first of these concerns the exact way that a defeater undermines an inference. In the case of Taller Tower, the inference is bracketed because the premises are false. In this case, we assume that the defeater in the background set is simply the negation of one or more of the premises. Call this *premise defeat*. However, sometimes the premises of a bracketed inference are true, but, owing to a defeater, the conclusion does not follow. For example, return to Chickenpox. Suppose that the patient has an allergy that frequently results in red spots but that only the *Varicella* virus is causing the red spots in this particular case. Then the defeater might be that the patient was not exposed to any allergens. In this case, both the true premises of the bracketed inference (that the patient's allergies frequently result in red spots) and its defeater (that the patient was not exposed to allergens) are in the background set of Chickenpox. Call this *inference defeat*.[7]

Turn next to the truth-value of the defeater. In paradigmatic cases of both premise and inference defeat, the defeaters are true or *veridical*. However, in certain cases, false defeaters are admitted into Σ for pragmatic purposes. Call these *expedient* defeaters. For instance, certain causes may bear on an effect, but should nevertheless be bracketed because one's interests lie in other causes. In this case, the background set typically includes an expedient inference defeater. Arguably, idealizations allow for expedient premise defeaters. Thus, both premise and inference defeat can be achieved by either veridical or expedient defeaters.

The pragmatics of explanation is a central gatekeeper of the defeaters that get placed in the background set. An inquirer's interests in specific aspects of a phenomenon dictate her bracketing policies, i.e., the inferences she ought to bracket. For instance, one might be interested in a system's causal properties, in which case, many potential causes of the phenomenon of interest must be bracketed, so that the intervention is not confounded by other variables. Frequently, bracketing inferences in these contexts is achieved via good experimental design and careful statistical analysis. As the

[7] Inference defeaters admit of a further subdivision between what Pollock (2015) calls *rebutting* defeaters (which provide reasons for believing the negation of the conclusion of a given inference) and *undercutting* defeaters (which challenge the support provided by the premises of a given inference). The distinction between rebutting and undercutting defeaters will not be relevant here.

bracketing policies associated with causation prove central to our solution of certain symmetry problems, we provide a more detailed discussion of them in Sect. 5.2.

DIME suggests a much richer pragmatics of explanation than we will explore here. Even within a causal framework, one's interests in certain kinds of causes—for instance, at specific levels of analysis, at more distal or more proximal points in a given causal history, etc.—will demand that different inferences are bracketed. Moreover, interests in noncausal properties will require different bracketing policies. For instance, suppose that Lange (2016) is correct, and that many noncausal explanations (such as the equatorial example above) confer necessities upon their explananda that are stronger than the necessities in which the laws of nature traffic. Then one's explanatory interests should be rightly attuned to those modal properties, and this will mean that inferences that fall short of the relevant modal threshold should be bracketed.[8]

Historically, both causal and noncausal theorists have tended to accord investigators' interests a healthy role in explanatory inquiry (Lipton 2004; van Fraassen 1980). However, whereas causal theorists often take the causal structure of the world to constrain runaway explanatory interests, it has been less clear how those of a noncausal bent can constrain the interest relativity of explanation, particularly in a way that will preserve its asymmetry.[9] DIME offers four objective constraints on explanatory interests, while maintaining the breadth needed to outperform the causal model.

First, the system must actually *have* the properties that interest the inquirer. This is a factual matter, not a pragmatic one. For instance, as we argue below, an interest in intervening on billiard balls requires that the balls have *causal* properties. Similarly, the equatorial temperature explanation depends on the system having the requisite mathematical properties. Because inquirers' interests can "misfire" if the systems lack the properties of interest—as might be the case with explanatory interests in certain quantum systems' deterministic properties or in comets' mental properties—this is one important constraint on our pragmatics of explanation.

Second, recall that bracketing amounts to assuming that certain nontrivial inferences are defeated. Once again, this involves several factual considerations. Most obviously, it is an objective matter as to whether or not a given body of information would defeat a given inference. In the case of veridical defeaters, there is a further question as to whether this information is true. In the case of expedient defeaters, there is an analogous question as to whether bracketing an inference is conducive—if not indispensable—to making correct inferences about the target property. Closely related, bracketing must not result in self-sabotage. For instance, one cannot be interested in the causes of the shadow's length and hold all such causes fixed. Hence, the concept of defeasible inference objectively constrains the pragmatics of explanation in a number of ways.

Third, many vague or gerrymandered properties of explanatory interest might be true of a system, and have a clearly delimited set of facts about defeaters. Hence, our factual constraints on explanatory interests are not enough. Additionally, such interests

[8] One way to do so would be to explicitly add the relevant modal operators into the content of the explanandum, but another way to do so is to keep the explanandum fixed and add modal information to both the defeater set Θ and background set Σ.

[9] For example, see Kitcher and Salmon's Kitcher and Salmon (1987) critique of van Fraassen's van Fraassen (1980) pragmatic approach to explanation.

must be *defensible*, in the sense that they are recognizable scientific or practical goals (such as intervening on a physical system). Otherwise, one may be criticized: "Why are you interested in *that*?" Insofar as anything is genuinely pragmatic, it is this "defensibility constraint," but for the reasons just sketched, it complements, rather than undermines, our two factual constraints.

Finally, after clearing all of these hurdles, there is a further question: given our bracketing policies, which inferences succeed where others fail? This, too, is an objective matter, and brings us to DIME's fourth, and most important, condition.

3.4 What is inferential success and failure?

To solve the symmetry problem, we must introduce a new basis for comparing explanations. In slogan form: only inferences that succeed where all others fail can be candidates for explanation. But what is meant by "success" and "failure" in such a slogan? The defeasibility of explanatory arguments provides an important clue. To say that one inference succeeds when another fails, we can imagine explanations being evaluated according to the following comparative procedure:[10]

Step 1: Consider all of the unbracketed nontrivial inferences that have the explanandum, B, as their conclusion. For each of these inferences, all other unbracketed inferences leading to B are its "competitors."

Step 2: For each A that has B as a nontrivial consequence in Step 1, suppose that all of A's competitors' premises are false.

Step 3: If the falsehood of any of these competitors defeats the inference from A to B, then the latter is not sturdy; otherwise, it is sturdy.[11]

Failure is defeat under these conditions; success is the absence of failure, i.e., sturdiness.

For example, assume that Tower is a correct explanation. Step 1 requires us to consider other ways of inferring the shadow's length. For instance, one competitor might be that a reliable instrument measured the shadow at 839.5 feet. (Call this the Measurement inference.) Step 2 then requires that we suppose the following: there is no such measurement of the shadow. Step 3 then asks whether the claim that no reliable instrument measured the shadow's length defeats the inference from the height of the Empire State Building to the length of the shadow. But, of course, the absence of a measurement will not disrupt the Empire State Building's ability to cast the shadow. Assuming that this could be done with all other competitors, Tower is sturdy.

Furthermore, suppose that we flip the script, such that Tower's premises are assumed to be false and treated as potential defeaters to Measurement. Assuming that the angle of the sun stays fixed (as would typically be mandated through

[10] To be clear: we are not claiming that explanations must be the products of this procedure. Indeed, we make no claims about the "production" of explanations whatsoever. Rather, this three-step process is simply a useful heuristic for the reader to identify the relevant inferential properties that distinguish explanations from other nontrivial inferences.

[11] The third step in the sturdiness test is represented formally as follows. If $\Sigma \mid C \mathrel{|\!\!\sim}_{\Theta} B$ is the only competitor to $\Sigma \mid A \mathrel{|\!\!\sim}_{\Theta} B$, then Step 3 consists in determining whether $\Sigma, \neg C \mid A \mathrel{|\!\!\sim}_{\Theta} B$ is defeated.

the bracketing process), then the shadow's length will change. Hence, if the reading of the instrument remains stuck at 839.5 feet (say because of some technological glitch) then MEASUREMENT will yield a false conclusion and thereby be defeated. If, on other hand, the instrument is sensitive to the changes in the shadow's length, then MEASUREMENT'S central premise—that the instrument read 839.5 feet—will be contradicted by its new reading, and hence will also be defeated. So, regardless of how sensitive the instrument is to (hypothetical) changes in the shadow, MEASUREMENT fails to be sturdy.[12]

As we shall argue in Sect. 4, sturdiness is what distinguishes explanations from their symmetry-mongering counterparts. However, before doing so, we should explain why we take sturdiness to be a characteristic feature of the inferences that constitute explanations. First, the ideas underwriting sturdiness capture the comparative evaluation that characterizes much explanatory reasoning. A paradigmatic way of evaluating candidate explanations is to see whether the explanandum still holds when one of the competing explanantia is false while the other is true. For instance, to determine whether Chemical X or Chemical Y caused a reaction, we would hold all other conditions fixed. If the reaction occurred when X was present and Y was absent, but not *vice versa*, then X would be a better explanation of why the reaction occurred than Y. Upon iteration, we would then arrive at the best explanation. Indeed, reasoning such as this bears strong structural affinities with, *inter alia*, controlled experiments, Mill's Method of Difference, and Lipton's (2004) well-known approach to Inference to the Best Explanation.

Second, sturdiness is a form of "stability" that many philosophers take to be a central feature of explanations. While different philosophical discussions of explanation and laws use different terminologies, in its most general form, X is said to be stable if X remains unchanged as other conditions C change. Call X the *stability-bearer*, and C the set of *stability-conditions*. For instance, Hempel's stability-bearers are so-called "lawlike generalizations," and his stability-conditions include spatiotemporal changes among other things. Similarly, Woodward's (2003) notion of "invariance" is a kind of stability, its bearers are generalizations, and its conditions are various kinds of interventions.[13] Our brand of stability, sturdiness, attaches first and foremost to *inferences*, and its stability-conditions are, at root, *competitors*. Our view is compatible with there being a derivative sense in which generalizations and laws are stability-bearers, and in which spatiotemporal changes and interventions are stability-conditions.[14]

For these reasons, sturdiness is the trademark feature of DIME. Sturdiness is tied to explanation in two ways: it dovetails with the ways in which scientists compare explanations, and it exhibits a kind of stability that is characteristic of good explanations.

[12] Strictly speaking, some clever counterexamples might appear to spoil this result. We address them in Sect. 4.1.

[13] Other accounts of stability include Lange (2009), Mitchell (2003) and Skyrms (1980).

[14] Indeed, while we will not argue for it here, Woodward's account of interventions can be seen as exhibiting the kind of inferential sturdiness we describe. Section 5 provides some clues as to how this argument would proceed.

4 Back to the symmetry problem

DIME's animating idea is that an explanation is a nontrivial inference that succeeds where its competitors fail. With this idea in hand, let us return to the symmetry problem. Our argument will proceed as follows:

1. An inference is an explanation only if it is sturdy.
2. The inferences that beget the symmetry problem are not sturdy.

∴ The inferences that beget the symmetry problem are not explanations.

Section 3 motivates the first premise of this argument. The second premise is carrying most of the dialectical load, for our goal in this essay is to show how our version of EAI can resolve the symmetry problem(s). The remainder of this essay supports the second premise by applying DIME to each of the three symmetry problems discussed in Sect. 2. For each problem, we will show that the symmetry-mongering inference is not sturdy, and also sketch how the correct explanation is a viable candidate for sturdiness. In this section, we address Bromberger's classic symmetry problem and the noncausal asymmetry problem involving equatorial temperature. Since it requires a more involved discussion, we postpone our solution to Barnes' symmetry problem until Sect. 5.

4.1 The classic symmetry problem

To show that the shadow's length does not explain the tower's height, we must find some alternative, nontrivial way of inferring the tower's height that succeeds where SHADOW fails. One obvious candidate is:

$$\begin{array}{ll} \Sigma \mid \text{The Empire State Building was designed to be 1,454 feet} & \text{(DESIGN)} \\ \mid\!\!\underline{}_{\Theta}\ H = 1,454\,\text{ft} & \end{array}$$

If the negated premises of just one would-be competitor defeat SHADOW, then the latter is not sturdy. So, let us assume that DESIGN and SHADOW are the only nontrivial inferences that have the tower's height as their conclusion.[15] Thus, for our purposes, Step 1 of 3 in our comparative procedure is complete.

Turning to Step 2, suppose that the Empire State Building was designed to be a different height. That is, we suppose that the premise of DESIGN is false. Such a supposition defeats SHADOW in two ways. First, a different design will imply that, when given the same angle of incidence (θ), the shadow's length will be different (i.e. $M \neq 839.5$ feet). However, this contradicts one of SHADOW's premises, and, as discussed above, inconsistent premises are *verboten* on pain of defeat. Second, a different design implies that the height of the tower will be different (i.e. $H \neq 1454$ feet), so when the different design is added to the premises, we do not get the desired conclusion—also a mark of defeat. Per Step 3, both of these considerations imply

[15] We also assume that Σ is typical in what it brackets, so as to block recherché counterexamples to DESIGN.

that Shadow is not sturdy. Hence, according to DIME, the shadow's length does not explain the tower's height. The symmetry is blocked.

In fairness, while the foregoing works in typical cases, there are more exotic counterfactual scenarios in which negating Design's premises will not defeat Shadow. For instance, Shadow would not be defeated by a different design, if, owing to a construction error, the Empire State Building still ended up being 1454 feet high. In response, let us consider a dilemma. First, if the construction errors actually occurred and affected the height of the Empire State Building, then a third inference, in which the same conclusion ($H = 1454$ feet) is drawn from the construction crew's actions, ought to be considered. Call this new inference Construction. This inference competes with both Design and Shadow. Furthermore, for reasons analogous to those just rehearsed, Construction's negated premises will defeat both inferences. Hence, Shadow will still fail to be sturdy.

If, on the other hand, construction errors played no actual role in the tower's height, then, in typical contexts, the background set (Σ) ought to include veridical defeaters indicating as much. In other words, Construction is bracketed. Moreover, this points to a hitherto unappreciated role of Σ. In Sect. 3.3, background conditions ruled out certain inferences as explanatory nonstarters (also see Step 1 of our sturdiness test.) Now we see that this very same information must not vary when making the kind of counterfactual comparisons required to ascertain which inference is sturdy (Step 3).

In effect, the background set functions as a "level playing field" by which to compare competitors when ascertaining their sturdiness. When determining whether an explanandum B could still be inferred from A if the latter's competitor C were negated, we must assume that Σ does not change. Thus, all of the bracketed inferences to B must remain bracketed. Given that bracketed inferences are known to fail prior to any such comparisons, unbracketing these inferences in our comparisons with competitors in no way helps us ascertain if an inference succeeds where all others fail.[16] With this level playing field in place, we return to our original two competitors, and see that if the design were different, Shadow could not snatch victory from the jaws of defeat by fleeing to a nearby possible world in which construction errors occurred. Once again, Shadow will not be sturdy. Hence, whether or not construction errors make a difference to the tower's height, the shadowy inference cannot lay claim to sturdiness. This dilemma applies to many conditions that would otherwise foil a sturdiness test.

4.2 Noncausal explanatory asymmetries

As argued above, causal theories of explanation cannot do justice to the asymmetry of noncausal explanation. Hence, it will be a victory if DIME can outperform the causal theory on this front. Specifically, we will now show that our model captures the

[16] Compare: suppose that we are debating whether LeBron James would beat Michael Jordan in a one-on-one match if each were at their primes at the same time. Neither basketball player shot over 35% from three-point range, so it would make no sense in this hypothetical basketball game to suppose that either player had greater accuracy from downtown. In the scenario we are considering, construction errors are like these high shooting percentages. As we'll see in Sect. 5, what's good for basketball also carries over to causal explanation.

sturdiness of the equatorial temperature example, and that it rules out the converse as non-explanatory.

The discussion of this example from Sect. 2 suggests the following is a correct explanation:[17]

For any real-valued, continuous function f, and any real number u between $f(a)$ and $f(b)$, there exists a value $c \,\varepsilon\, [a, b]$ such that $f(c) = u$.
The difference in temperature $D(x)$ between an equatorial point x and its antipode is a real-valued and continuous function.
For any point x and its antipode x', 0 is between $D(x)$ and $D(x')$.
$\vdash_{\Theta}$ There exists an equatorial point and its antipode with the same temperature.
(Antipode)

As noted in Sect. 2, noncausal explanations exhibit asymmetries. So, suppose we turn this explanation on its head:

For any real-valued, continuous function f, and any real number u between $f(a)$ and $f(b)$, there exists a value $c \,\varepsilon\, [a, b]$ such that $f(c) = u$.
For any point, x and its antipode x', 0 is between $D(x)$ and $D(x')$.
There exists an equatorial point and its antipode with the same temperature.
$\vdash_{\Theta}$ The difference in temperature $D(x)$ between an equatorial point x and its antipode is a real-valued and continuous function. (Temperature)

The existence of antipodal points with identical temperatures does not explain the continuity of temperature distributions around the equator any more than the length of the shadow explains the tower. Hence, we must show that Temperature is not a sturdy inference, and thereby not an explanation. In this regard, we note that switching from a classical to a defeasible consequence relation actually makes our job harder. As briefly discussed in Sect. 2, this asymmetry is readily captured if one is a deductivist: Antipode is valid; Temperature is not. By contrast, because DIME treats the explanatory relation as defeasible, it is at least *possible* that Temperature is explanatory. We shall now extinguish this possibility, and in the process, highlight a consequence that will figure in Sect. 5.3, where we show that DIME provides a unified treatment of both causal and noncausal asymmetries.

The competitor to Temperature involves a bit of exposition. D's being real-valued and continuous is an artifact of equatorial temperature's being real-valued and continuous. This, in turn, is an instance of a very general feature of the way that temperature distributes itself within a spatiotemporal region, which is described by the heat equation:

[17] In this section, we leave the background set Σ implicit.

$$\frac{\partial u}{\partial t} - \alpha\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = 0 \qquad \text{(Heat Equation)}$$

Here, u denotes variation in temperature (what is called "heat"); t, time; x, y, and z, spatial coordinates; α, a positive constant denoting thermal diffusivity. A viable competitor to TEMPERATURE would be a nontrivial derivation of the heat equation. (Since the heat equation is continuous and real-valued, the more specific claim about D that serves as TEMPERATURE's conclusion can be appended to this derivation.)

One such derivation proceeds from Fourier's law, which states that the rate at which energy flows out of a body is proportional to the area through which the energy flows, as well as the temperature difference at opposing ends of the body. Since temperature is average kinetic energy, it follows from this that heat cannot make discrete jumps without violating the first law of thermodynamics, which states that energy can neither be created nor destroyed. In other words, the heat equation can be inferred nontrivially from the combination of Fourier's law and the conservation of energy. Whether or not antipodal equatorial points have the same temperature would not defeat such an inference. On the other hand, if deep physical principles such as Fourier's law or the law of energy conservation were false, then nothing about equatorial temperatures would spare TEMPERATURE from defeat. Thus, TEMPERATURE is not sturdy. According to DIME, this means that it is not a candidate for explanation. Hence, our account of explanation blocks the problematic inference that gives rise to noncausal symmetries.

Since it will prove useful below, let us also discuss why the correct explanation, ANTIPODE, does not meet a similarly ignominious fate. This inference is nontrivial in the sense characterized by Sect. 3.2. It is irreflexive, premise consistent, and defeasible. To determine its sturdiness, we need to compare it with other inferences that have the same conclusion. One competitor would proceed by identifying an equatorial point, say Nairobi, Kenya, that has the same temperature as its antipode, Fortaleza, Brazil, then citing the meteorological causes Nairobi's temperature, and then finally looking to the opposite side of the world for the meteorological causes of Fortaleza's temperature. Assuming that these temperatures are the same, we can then infer the conclusion of ANTIPODE. Call this METEOROLOGICAL COMPETITOR.

The second step of the sturdiness test requires us to negate the premises of this argument, i.e. to assume that these meteorological causes did not occur. The third step is to determine whether such an assumption defeats ANTIPODE. Combined, this is to say that if the causes of the temperature of Nairobi and Fortaleza had been different, then there would be no antipodal equatorial points with the same temperature. However, this is clearly false, since the intermediate value theorem confers mathematical necessity on ANTIPODE's conclusion. Hence, if Nairobi and Fortaleza do not have the same temperature, then some *other* pair of antipodal equatorial points must have the same temperature. Thus, ANTIPODE is a good candidate for explanation.[18] Moreover, it clearly defeats METEOROLOGICAL COMPETITOR.

[18] In addition to the differences between the intermediate value theorem and the meteorological causes, ANTIPODE is shown to be sturdy if and only if, for its remaining competitors, either ANTIPODE is undefeated on the supposition that their premises are false, or these competitors are bracketed, as discussed in Sect. 3.3.

Note that METEOROLOGICAL COMPETITOR is a causal explanation. While the inference is not sturdy, it certainly provides some causal story about the temperatures along the equator. This illustrates an interesting result: inferences that do not track causes can sometimes trump inferences that do. Hence, we have raised two problems with the causal approach to the symmetry problem. First, some explanatory asymmetries, such as the difference between ANTIPODE and TEMPERATURE, do not appear to rest on causal facts. Second, under certain conditions, some causal explanations, such as the meteorological competitor, are inferior to noncausal explanations. As we shall see, these two points will prove instrumental in drawing the appropriate lessons from our last and most challenging symmetry problem, to which we now turn.

5 Causation and sturdy inference

Thus far, we have seen one explanatory asymmetry—involving the Empire State Building and its shadow—that *need not* be construed as a causal asymmetry, and another—involving equatorial temperature—that *should not* be construed as a causal asymmetry. However, there is a class of explanatory asymmetries where causation plays a starring role. Barnes' critique of Kitcher, briefly discussed in Sect. 2, provides an exemplar of this kind of asymmetry. Our treatment of this species of asymmetry involves three main argumentative maneuvers. First (Sect. 5.1), we motivate the special challenge these examples pose for our view. In particular, appeals to sturdiness without further appeal to causation end up being either too permissive (leaving the symmetry-mongering inference sturdy) or too prohibitive (leaving the correct explanation unsturdy). Then (Sect. 5.2), we show that appealing to causation delivers the correct verdicts as to which inferences ought to be sturdy. Finally, we argue that despite this appeal to causation, sturdiness is still the driving force behind the explanatory asymmetry (Sect. 5.3).

5.1 Permissive and prohibitive sturdiness

Let us begin by showing that, for some examples of explanatory asymmetry, avoiding causation raises certain problems. For example, conservation of kinetic energy entails that any moving particle X that collides elastically with a resting particle Y of equal mass will obey the following law:

$$(V_{1X})^2 = (V_{2X})^2 + (V_{2Y})^2 \tag{Velocity Law}$$

In this formula, V_{1X} denotes X's velocity up to its collision with Y, while V_{2X} and V_{2Y} denote the velocities of the particles at some time after this collision. For concreteness' sake, let us consider a simple system consisting of two billiard balls, A and B, on a standard billiards table. A moves across the table and collides with B, which was at rest. After the collision, A's velocity has changed, and B takes on a velocity.

As with our previous examples, there are two initial inferences to consider:[19]

$$\Sigma \mid \textsf{Velocity Law},\ V_{1A} = 1\,\text{m/s},\ V_{2A} = .6\,\text{m/s} \vdash_{\Theta} V_{2B} = .8\,\text{m/s} \qquad (\text{Ball A})$$

$$\Sigma \mid \textsf{Velocity Law},\ V_{2B} = .8\,\text{m/s},\ V_{2A} = .6\,\text{m/s} \vdash_{\Theta'} V_{1A} = 1\,\text{m/s} \qquad (\text{Ball B})$$

Much like before, we must show that Ball A's prospects for being sturdy are strong, while Ball B is not sturdy. Following the now-familiar template, consider a would-be competitor to Ball B; say one in which the velocity of ball A is inferred from some prior event.[20] Suppose that before colliding with B, A was at rest until struck by a third ball, C. In that case, the following defeasible inference would be acceptable (where V_{0C} is the velocity of ball C at time interval $t_0 < t_1$):

$$\Sigma \mid \textsf{Velocity Law},\ V_{0C} = 1.17\,\text{m/s},\ V_{1C} = 0.6\,\text{m/s} \vdash_{\Theta''} V_{1A} = 1\,\text{m/s} \quad (\text{Ball C})$$

Let us see whether Ball C blocks the sturdiness of Ball B. As before, we negate the premises of Ball C, i.e., we suppose that C's velocity was not 1.17 m/s. If all went according to plan, Ball B would be defeated. But alas, this is not the case: the supposition that C did not collide with A at 1.17 m/s is compatible with Ball B still being a good retrodiction of ball A's velocity. This is because a different set of conditions at time 0 might have brought it about that $V_{1A} = 1$ m/s. C's velocity could have been higher or lower (making suitable adjustments to the angle of impact and post-collision velocity of C), and it would still follow that $V_{1A} = 1$ m/s. The difference between Ball A and Ball B cannot be made out on a straightforward analogy with our treatment of Shadow in Sect. 4.1.

While it is a dead end, this approach to the symmetry problem underscores the importance of which inferences are bracketed, as discussed in Sect. 3.3. In the earlier discussion of the Tower inference, the existence of another building shading out the Empire State Building was merely a *potential* explanation, since the explanans was known to be false. We could then bracket this potential explanation by including one of its veridical defeaters in the background set. For instance, our background set might include information about the absence of a taller tower behind the Empire State Building. Furthermore, in our discussion of construction errors, we saw that if something is bracketed, it must stay bracketed when judging whether one inference would be defeated by the negation of its competitors' premises. Whether or not Tower is defeated thus depends in part on what inferences are bracketed, which in turn dictates what counterfactual suppositions are allowed. In the discussion of the billiard ball example so far, we have permitted any counterfactual supposition about ball C's velocity to serve as a competitor to Ball B. Ball B is not defeated by negating C's velocity because there are too many ways for C's velocity to be different without changing A's velocity. As a result, the bare fact that C is not moving at 1.17 m/s does

[19] Recall from Sect. 3.2 that such inferences are defeasible because of modeling considerations.

[20] To simplify the presentation, we use the ambiguous phrase "ball A's velocity." Unless otherwise specified, "ball A's velocity" denotes V_{1A}, "ball B's velocity" denotes V_{2B}, and "ball C's velocity" denotes V_{0C}.

not defeat Ball B. Since nothing is bracketed, all of the inferences we have been considering remain undefeated. For this reason, we call this approach to this kind of symmetry problem *permissive sturdiness*.

Since permissive sturdiness fails to achieve the desired asymmetry, and the velocities of balls B, and C are known, then perhaps we should bracket any nontrivial inference other than Ball B or Ball C that has $V_{1A} = 1$ m/s as its conclusion, e.g.

$$\textsf{Velocity Law},\ V_{0C} = 1.28\,\text{m/s},\ V_{1C} = 0.8\,\text{m/s} \mathrel{|\!\!\!\underset{\Theta'''}{\rule{1.5em}{0.4pt}}} V_{1A} = 1\,\text{m/s} \qquad (\text{Ball C*})$$

Let us bracket all other inferences to A's velocity. This ensures that, when we get to Step 2, and negate the premises of Ball C, it follows that $V_{1A} \neq 1$ m/s. In other words, inferences such as Ball C* will no longer be able to "take over" for Ball C once its premises are negated in Step 2. In this way, C's velocity acts like a switch for A's velocity, i.e. changes in C's velocity are both necessary and sufficient to change A's velocity. Hence, the negated premises of Ball C defeat Ball B, for we get an inconsistency, $V_{1A} = 1\,\text{m/s} \wedge V_{1A} \neq 1\,\text{m/s}$. The desired result at Step 3 is thereby achieved: Ball B is not sturdy, and hence not an explanation.

Unfortunately, the problem with this bracketing policy is that it throws out the baby with the bathwater. To see why, let us continue to suppose that every other way of inferring ball A's velocity except Ball B and Ball C is defeated. Next, consider whether the negated premises of Ball B defeat Ball C. Since we have ruled out all of the other possible future trajectories of B, then, if $V_{2B} \neq 0.8$ m/s, V_{1A} will not equal 1 m/s.[21] So, although Ball B will not be an explanation, neither will Ball C (nor, presumably, will Ball A be an explanation for that matter.)[22] If everything outside of the two inferences being compared is bracketed, no inferences are sturdy. Call this *prohibitive sturdiness*.

Since sturdiness depends on comparative defeasibility, and defeat depends, in part, on what other inferences are bracketed, sturdiness depends (in part) on the background set Σ. We have seen that if we bracket *nothing*, then *both* the symmetry-mongering inference, Ball B, and the correct explanation, Ball C, are sturdy. That is the problem with being overly permissive. Yet, if we bracket *as much as possible*, then *neither* of these inferences are sturdy. That is the problem with being overly prohibitive. The challenge is to find principled ways to bracket the right sorts of inferences when comparing inferences.

5.2 Causal sturdiness

Let us take stock. When we were permissive about the background, we did not limit what might happen *after* ball A achieved its velocity ($V_{1A} = 1$ m/s). This had the virtue of precluding the negated premises of Ball B from defeating Ball C. Hence,

[21] In effect, this will underwrite a backtracking counterfactual: "Had $V_{2B} \neq 0.8$ m/s, then it would have had to have been the case that $V_{1A} \neq 1$ m/s.

[22] Since the arguments showing that Ball C is sturdy (or not) can be extrapolated to Ball A, we focus on showing that the former is a candidate for sturdiness. Analogous considerations apply to the latter.

permissive sturdiness captures the fact that future events cannot influence past events. Conversely, prohibitive sturdiness brackets a good deal with respect to what happens *before* A gets its velocity of 1 m/s, so that the negated premises of BALL C defeat BALL B. In this way, it captures the idea that future events can be influenced by past events. If we could combine these restrictions, the symmetry of BALL A and BALL B could be broken. But how can this be done nonarbitrarily?

Return to our discussion of the pragmatics of explanation from 3.3. Because billiards is a game predicated on physical interventions, billiards players are interested in the *causal* properties of the billiard balls. As discussed above, this dictates what gets bracketed. Specifically, in a causal system, holding certain causes fixed amounts to bracketing their inferential analogues. This bracketing policy, in turn, renders the temporally forward-looking inferences sturdy while undermining the sturdiness of their temporally backward-looking cousins. Let us call this *Causal Sturdiness*:

If C is an actual cause of A, and B is a later event, then (*Causal Sturdiness*)

$\Sigma \mid B \mathrel{|\!\!\!\sim_{\Theta}} A$ is not sturdy, while $\Sigma \mid C \mathrel{|\!\!\!\sim_{\Theta'}} A$ may be.

Applied to the example at hand, this means that if C's velocity is an actual cause of A's velocity, then BALL B is not sturdy, but this does not prohibit BALL C from being a sturdy inference. To see why *Causal Sturdiness* holds, let us unpack two things. First, we borrow the crucial concept of an "actual cause" from Woodward (2003):[23]

(AC1) The actual value of $X = x$ and the actual value of $Y = y$.

(AC2) There is at least one route R from X to Y for which an intervention on X will change the value of Y, given that other direct causes Z_i of Y that are not on this route have been fixed at their actual values. (It is assumed that all direct causes of Y that are not on any route from X to Y remain at their actual values under the intervention on X.)

Second, let us now consider the bracketing policy of an inquirer interested in a system's actual causes. *Ex hypothesi*, the actual value of $V_{0C} = 1.17$ m/s and the actual value of $V_{1A} = 1$ m/s. Hence, (AC1) is satisfied. In our idiom, this also provides veridical premise defeaters for every nontrivial inference in which V_{0C} assumes some value other than 1.17 m/s. As a result, these inferences are bracketed. Like prohibitive sturdiness, this immediately takes the problematic inference BALL C* out of the competition.

Furthermore, (AC2) requires some "route" or causal chain from V_{0C} to V_{1A} such that at least one change to V_{0C} leads to a change in V_{1A} when all other causes not on this route are held fixed at their actual value. In effect, this means that given an inquirer's interest in determining the actual causal contributions of balls B and C to ball A's velocity, all other inferences wherein the premises cite *causes* of the event described in the conclusion are bracketed. More precisely, assume that the other causes Z_i of ball A's velocity can be represented as inferences of the form $\Sigma \mid Z_i = z_i \mathrel{|\!\!\!\sim_{\Theta_i}} V_{1A} = 1$

[23] We have not used Woodward's ultimate formulation of actual causation, (AC*), as it introduces complexities that are unnecessary for the purposes at hand.

m/s.[24] Then (AC2) requires each candidate for sturdiness to include two different kinds of information in its background set, Σ. First, for all i, $Z_i = z_i$ is in Σ. This defeats any inference that has non-actual values for Z_i in its premises, and thereby captures the idea that the other direct causes of ball A's velocity are held fixed in any comparison performed in the context of our sturdiness test. Second, it is assumed that these alternative routes are not competing with BALL C and BALL B. Hence, even though these inferences describe things that have a causal bearing on ball A, they are treated as defeated for the purposes at hand. Consequently, interest in actual causation requires Σ to contain expedient inference defeaters of the alternative routes.

Summarizing, the veridical premise defeaters for all non-actual values of V_{0C} are in Σ. In other words, merely potential causes cannot be the correct explanation. Furthermore, the expedient inference defeaters for all other causes of V_{1A} are also in Σ; i.e. they are held fixed. These defeaters constitute the level playing field on which actual causal explanations are compared. Once Σ is so structured, it follows that whenever $V_{0C} \neq 1.17$ m/s, then $V_{1A} \neq 1$ m/s, even if all of the other causes of A's velocity remained exactly the same. Hence, *Causal Sturdiness* replicates the desirable feature of prohibitive sturdiness: the negated premises of BALL C defeat BALL B, and the latter is prevented from being sturdy.

Causal Sturdiness also inherits the good parts of permissive sturdiness. Treating C's velocity as an actual cause does not require us to bracket inferences describing the *effects* of A's velocity. Even if we negate the premises of BALL B, we may still infer A's velocity from C's velocity. The reason for this is that any number of other "retrodictive" inferences leading to A's velocity are compatible with the premises of BALL C. Hence BALL C is not defeated. Consistent with the lessons of Sect. 3.3, these retrodictive inferences *can be* bracketed, but given inquirers' interests in the billiard balls' causal properties, they *are not* and indeed *ought not* be bracketed.

Causal Sturdiness says that when we treat the events under consideration as parts of a system where some events are *actual causes* of others, retrodictive inferences are rendered non-sturdy. This raises a concern: doesn't this solution to the symmetry problem concede too much to causality?

5.3 Causation or inference: which comes first?

Our discussion of the billiards example highlights an underlying tension in the larger debate as to whether causation or inference is more fundamental to explanation. On the one hand, the example shows causal explanations to be *sturdy*, which favors an "inference-first" approach to explanation. On the other hand, it also shows *causal* explanations to be sturdy, which favors a "causation-first" approach to explanation. So what role does causation play in grounding the asymmetry in the billiards example?

On our inference-first approach, the asymmetry of explanation ultimately boils down to differences in sturdiness. Causes turn out to be especially good ways of

[24] Since the well-known syphilis-paresis example, there is a longstanding debate as to whether such an assumption is safe. However, the leading causal theorists, e.g. Strevens (2008) and Woodward (2003), both require every explanation to have some inferential structure, even if they do not require every explanation of be an inference. We hope to link these ideas to DIME in future work.

achieving sturdiness, but they are not the only means for doing so. For instance, even though Bromberger's classic example appears to involve a causal explanation, we managed to preserve its asymmetry without appealing to any causal relationship between the tower and the shadow. More telling, accounting for the asymmetry of explanations involving equatorial temperature cannot advert to causes. Finally, and perhaps most importantly, even when we deploy causes, our argument is altogether different than the typical argument for establishing explanatory asymmetry via causal asymmetry, i.e. we did *not* argue as follows:

1. $V_{1A} = 1$ m/s causes $V_{2B} = 0.8$ m/s.
2. Causation is asymmetrical, i.e. if A causes B, then B does not cause A.
3. All explanations cite causes.

∴ $V_{2B} = 0.8$ m/s does not explain why $V_{1A} = 1$ m/s.

As our discussion of equatorial temperatures makes clear, we reject the third premise, and, for all that we've argued, we can remain agnostic about the second. Indeed, while we accept the first premise, it also played no role in establishing the asymmetry of causal explanation in the billiard ball example. That premise relies on the causal relationship between the velocities of balls A and B to establish the desired asymmetry. By contrast, *Causal Sturdiness* relies on the causal interaction between the velocities of balls C and A to establish that the symmetry-mongering inference BALL B is not sturdy. In short, no part of the argument above is essential to our solution of Barnes' symmetry problem.

Since the argument above is the standard way of arguing for explanatory asymmetry on the basis of causal asymmetry, it is clear that we are up to something else. More precisely, our argument is as follows:

1. All explanations are sturdy inferences.
2. If $V_{0C} = 1.17$ m/s is an actual cause of $V_{1A} = 1$ m/s, then BALL B is not a sturdy inference.
3. $V_{0C} = 1.17$ m/s is an actual cause of $V_{1A} = 1$ m/s.

∴ BALL B is not an explanation.

Furthermore, prioritizing inferences over causes has several advantages when thinking about explanatory asymmetries. First, some asymmetries are not causal, as was seen in the example of ANTIPODE. Second, causes that fail to be sturdy can be trumped by sturdy non-causal competitors, as was the case with METEOROLOGICAL COMPETITOR when compared with ANTIPODE. Furthermore, the billiards example now shows that even the asymmetries that, according to the earlier state of the field, appeared to be the exclusive province of causal approaches also admit of an inferential rendering. So, when compared to our inference-first approach, causation-first approaches suffer several disadvantages, and enjoy no distinct advantages.

6 Conclusion

Our aims in this essay have been both critical and constructive. On the critical side, we have shown that not all explanatory asymmetries are causal asymmetries, and their standard causal diagnosis cannot be right. Indeed, the symmetry problem is really a three-headed monster—at least when viewed against the broader dialectic of causal and inferential approaches to explanation. Some asymmetries have an ineluctable causal element; some are decidedly noncausal; and others are fair game for both parties to the debate.

On the constructive side, we have sketched the broad contours of a new version of EAI—what we have called the defeasible inference model of explanation (DIME). It departs from earlier versions of EAI in its melding of defeasible and comparative components. This duet achieves its denouement in the concept of sturdiness—roughly, the idea that explanations are inferences that succeed where their competitors fail. As we have shown, sturdiness is the common thread that ties together the different kinds of explanatory asymmetries.

This is but an opening salvo in a research program that we hope to develop in greater detail, by extending DIME to solve other venerable problems in the explanation literature. Earlier versions of EAI face a variety of problems.[25] How should laws be characterized? How to make sense of indeterministic explanations of improbable events, such as the fact that a person's untreated syphilis explains his paresis, despite the fact that only 25% of untreated syphilitics suffer from paresis?

Equally importantly, the symmetry problem has overshadowed two *advantages* that early variants of EAI enjoy over causal approaches. First, EAI is a natural way to analyze *non-causal* explanations. However, this paper has only focused on the *asymmetry* of these explanations. In the future, we hope to extend DIME to the growing stockpile of examples of non-causal explanations (Baker 2005; Batterman 2002; Bokulich 2011; Huneman 2010; Irvine 2015; Lange 2016; Rice 2015; Risjord 2005).

Second, in comparison with causal approaches, earlier proponents of EAI enjoyed what we might call *Humean modesty*. Inference-based approaches argue that explanations are simply inferential relationships between certain empirical statements. Hence, competent language users can explain by wielding inferences that carry no further commitment to a substantive modal or causal ontology. As a result, EAI often avoids the various placement problems associated with modality and causality (e.g. how modality fits within a naturalistic ontology, how modal and causal claims can be known, etc.).[26]

Since the demise of the covering law model, the symmetry problem hung like an albatross around the neck of proponents of EAI. This paper has sought to loosen that grip, and to allow new approaches to EAI to breathe. By solving the problem of explanatory asymmetry, we have cleared an important barrier to showing how inferential considerations latch onto explanation's deeper structures.

[25] For a review of these challenges, see Salmon (1989) and Woodward (2014).

[26] Indeed, in Millson et al. (2018), we develop a more precise account of explanation using a broadly inferentialist semantics. This approach to modal and explanatory vocabulary gives those of a Humean bent a compelling story about how one can *use* modal vocabulary without having to *represent* or be *ontologically committed* to metaphysically controversial modal entities (see Brandom 2008, 2015). This idea can be traced back to Sellars (1957). For similar approaches to the semantics of modal vocabulary see Thomasson (2007) and Stovall (2015).

Compliance with ethical standards

Conflicts of interest The author declares no potential conflicts of interests with respect to the authorship and/or publication of this article. The author received no financial support for the research and/or authorship of this article.

References

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind*, *114*(454), 223–238.

Bangu, S. (2016). Scientific explanation and understanding: Unificationism reconsidered. *European Journal for Philosophy of Science*, *7*(1), 103–126.

Barnes, E. C. (1992). Explanatory unification and the problem of asymmetry. *Philosophy of Science*, *59*(4), 558–571.

Batterman, R. W. (2002). *The Devil in the details : Asymptotic reasoning in explanation, reduction and emergence*. New York: Oxford University Press.

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*(1), 33–45.

Brandom, R. (2008). *Between saying and doing: Towards an analytic pragmatism*. Oxford: Oxford University Press.

Brandom, R. (2015). *From empiricism to expressivism*. Cambridge: Harvard University Press.

Bromberger, S. (1965). An approach to explanation. In R. Butler (Ed.), *Studies in analytical philosophy* (Vol. 2, pp. 72–105). Oxford: Blackwell.

Bromberger, S. (1966). Why-questions. In R. Colodny (Ed.), *Mind and cosmos: Essays in contemporary science and philosophy* (pp. 86–111). Pittsburgh: University of Pittsburgh Press.

Cartwright, N. (1983). *How the laws of physics lie*. New York: Oxford University Press.

Colyvan, M. (1998). Can the eleatic principle be justified? *Canadian Journal of Philosophy*, *28*(3), 313–335.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, *71*(1), 5–19.

Hempel, C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.

Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese*, *177*(2), 213–245.

Irvine, E. (2015). Models, robustness, and non-causal explanation: A foray into cognitive science and biology. *Synthese*, *192*(12), 3943–3959.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation* (Vol. XIII, pp. 410–506). Minneapolis: University of Minnesota Press.

Kitcher, P., & Salmon, W. C. (1987). Van Fraassen on explanation. *Journal of philosophy*, *84*(6), 315–330.

Lange, M. (2009). Why do the laws explain why? Mind Association Occasional Series. In T. Handfield (Ed.), *Dispositions and Causes*. Oxford: Oxford University Press.

Lange, M. (2016). *Because without cause: Non-causal explanations in science and mathematics*. New York: Oxford University Press.

Lipton, P. (2004). *Inference to the best explanation. International library of philosophy and scientific method*. London: Routledge.

Millson, J., Khalifa, K., & Risjord, M. (2018). Inferentialist expressivism for explanatory vocabulary. In O. Beran, V. Kolman, & M. Koreñ (Eds.), *From rules to meanings: New essays on inferentialism* (pp. 155–178). London: Routledge.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Pollock, J. (2015). *Knowledge and justification*. Princeton: Princeton University Press.

Reutlinger, A. (2017). Explanation Beyond Causation? New directions in the philosophy of scientific explanation. *Philosophy Compass*, *12*(2),

Reutlinger, A., Schurz, G., & Hüttemann, A. (2015). Ceteris paribus laws. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Center for the Study of Language and Information: Stanford University.

Rice, C. C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs*, *49*(3), 589–615.

Risjord, M. (2005). Reasons, causes, and action explanation. *Philosophy of the Social Sciences*, *35*(3), 294–306.

Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 3–219). Minneapolis: University of Minnesota Press.

Schurz, G. (1999). Explanation as unification. *Synthese*, *120*(1), 95–114.

Schurz, G., & Lambert, K. (1994). Outline of a theory of scientific understanding. *Synthese*, *101*(1), 65–120.

Sellars, W. (1957). Counterfactuals, dispositions, and the causal modalities. In G. Maxwell (Ed.), *Minnesota studies in the philosophy of science* (Vol. II, pp. 225–308). Minneapolis: University of Minnesota Press.

Skyrms, B. (1980). *Causal necessity: A pragmatic investigation of the necessity of laws*. New Haven: Yale University Press.

Stovall, P. (2015). *Chemicals, organisms, and persons: Modal expressivism and a descriptive metaphysics of kinds*. Ph.D. thesis, University of Pittsburgh.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Cambridge: Harvard University Press.

Thomasson, A. L. (2007). Modal normativism and the methods of metaphysics. *Philosophical Topics*, *35*(1/2), 135–160.

van Fraassen, B. (1980). *The scientific image*. New York: Clarendon Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2014). Scientific explanation. In E.N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2014 ed.).