Explanatory Asymmetry and Inferential Practice

September 5, 2016

Abstract

In this paper, we offer a new account of explanation. Its guiding idea is that explanation is an inferential practice. While some have argued that explanations are inferences, and others have argued that explanations are practices, our view provides a unique combination of the two. Arguably, the most significant challenge to both inferential and pragmatic accounts of explanation is the problem of asymmetry. In short, both inferential and pragmatic approaches to explanation fail to capture the idea that, in general, if *A* explains *B*, then *B* does not explain *A*. We review how this problem has plagued our predecessors, while showing that our own view readily solves this 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 fable almost always includes reference to a flagpole's height and its shadow's length.

The symmetry problem serves as an expedient way to disqualify two general approaches to explanation. The first holds that explanations are inferences (Friedman 1974; Hempel 1965; Kitcher 1989; Schurz 1999; Schurz and Lambert 1994). Indeed, it was Bromberger's (1965) critique of Hempel that first launched the cottage industry of symmetry problems. While other inferential approaches (e.g. Kitcher 1989) purported to restore explanation's asymmetry, they too faced searching counterexamples (see Barnes 1992). The symmetry problem also has posed problems for a second, pragmatic approach to explanation (Achinstein 1983; Faye 2007; Garfinkel 1981; Risjord 2000; Fraassen 1980), with Kitcher and Salmon's (1987) critique of van Fraassen standing as a challenge to others who would claim that explanation is a speech-act or practice.

By contrast, the symmetry problem has been a great advertisement for causal approaches to explanation. On this view, 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).

But perhaps the symmetry problem is not inherent to inferential or pragmatic approaches. Perhaps the problem stems from our misunderstandings of inference, practice, and their connection. In particular, inferential approaches to explanation have assumed that inference is a formal relationship between sentences in a language. It is time to rid ourselves of this positivist hangover: first and foremost, scientific inference should be thought of as a kind of scientific *practice*—something that scientists *do*. Similarly, earlier pragmatic approaches to explanation have assumed that explanatory practice is mostly driven by interests and presuppositions, but have given short shrift to the important ways in which *reasoning* (i.e. inference) also figures in these practices.

In this paper, we offer a new account of explanation, guided by the idea that explanation is an *inferential practice*. Our view provides a unique combination of both inferential and pragmatic approaches, as we argue in Section 2. In Section 3, we review how the symmetry problem has plagued our predecessors, while showing that our own view readily solves this problem.

2   The Inferential Practice Model of Explanation

As just noted, we take explanation to be an inferential practice. In this section, we unpack this slogan. Section 2.1 discusses general features that all inferential practices share. Section 2.2 then distinguishes explanations from other kinds of inferential practices, by appeal to the stability or "sturdiness" of the inferences in which explanations trade. Thus, programmatically, our view states that A explains B if and only if:

1. B is a sturdy consequence of A, and

2. Both A and B are true.

Before discussing inferential practices more broadly, and the particular ones involving sturdy inferences, two preliminary clarifications are in order. First, in this paper, we only provide necessary conditions for when B is a sturdy consequence of A; hence, our analysis is only partial. This will not matter in what follows, since we have provided enough of an analysis to solve the symmetry problem. In future work, we intend to complete this analysis by supplementing our account of sturdy inference accordingly.

Second, for many explananda B, there are multiple propositions $A_1, \ldots A_n$ such that it is natural to say that A_1 explains B, A_2 explains B, etc., and that these explanantia are not in competition. While the arguments are outside of the scope of this essay, it is a consequence of our view that when A explains B, A is the "exhaustive explanation" $A_1, \ldots A_n$. The pragmatics of explanation permit an element of the exhaustive explanation, A_i, to be treated as "the" explanation of B. For ease of exposition, we will not enumerate all of these different components of this exhaustive explanation when we provide examples of sturdy inferences. For the purposes of this paper, what matters most is that our view entails that for any A and B that raises a symmetry problem (i.e. such that A explains B, but B does not explain A), B can be shown not to be part of the sturdy inference that has A as its conclusion, and hence can be excluded from the exhaustive explanation of A.

2.1   Inferential Practice

With these clarifications in hand, let us discuss some general features of inferential practices. Conventional understandings of inference hold that it is fundamentally a relation among sentences in a language. By contrast, for us, inference is first and foremost a *practice* or *activity*; its conception as a sentential relation or function is derivative. Linguistic practices—what Wittgenstein called "language games"—qualify as inferential if the norms that determine the appropriateness of their speech acts can be specified as rules of inference. The most basic of these rules cannot, on pain of infinite regress, be thought of as *explicitly* articulated statements. Rather, the propriety of inferential moves is exhibited by the way other participants respond to those who make them. In this sense, the rules governing the inference-game (i.e. the rules of inference) are *implicit* in the practice of playing it.[1]

---

[1] The idea that drawing inferences is one of, if not *the* essential way in which we use language is central to the approach of semantic inferentialism pioneered by Sellars 1953, 1954 and developed in detail by Brandom 1994, 2008 and Peregrin 2014. Most recently, Hlobil 2016 has established a sequent calculus for material consequence relations, which has been invaluable to our formalization of sturdy inferences; see Appendix below.

Since inferential practices are understood as a type of rule-governed social activity, we may reasonably specify their varieties in explicitly normative terms. Inferences are correct (deductively valid) if commitment to the premises commits the speaker to the conclusion or (inductively good) if commitment and entitlement to the premises entitles the speaker to the conclusion. Speakers may also come to have incompatible commitments when one of their claims prohibits entitlement to another. We thus understand inference and assertion in terms of the normative concepts of *commitment* and *entitlement*.

Thus, the most basic rules of inference are commitments and entitlements that are implicit in our inferential practices. We can sharpen this point by highlighting that inferential practices are rooted in *material* inferences.[2] Unlike the inferences codified by formal logic, material inferences are those whose correctness is *not* determined by the meaning and arrangement of logical terms—i.e. their 'logical form.' The inference from "Boston is north of Atlanta" to "Atlanta is south of Boston" is an example of a good material inference. The formalist's insistence that this inference is enthymematic until the suppressed premise "If x is north of y, then y is south of x" is added simply trades the goodness of the material inference for the truth of the conditional. On our view, the inference is good in virtue of the commitments and entitlements governing the use of its non-logical vocabulary, e.g. "Atlanta," "Boston," "north," and "south."[3] The supposedly 'suppressed' conditional—and logical vocabulary more generally—is a way of making commitment to that inference explicit.[4]

To summarize, we take explanations to be a kind of inferential practice. Inferential practices are linguistic practices governed by implicit commitments and entitlements that specify the ways in which non-logical or material vocabulary can be used in inferences. Traditional logical and semantic vocabulary then makes these norms—which would otherwise be implicit in practice—explicit in speech.

2.2 Explanation as Sturdy Inferential Practice

Our goal is to treat explanatory vocabulary as a kind of logical or semantic vocabulary: as an expressive device that makes commitments and entitlements that would otherwise remain implicit in certain inferential practices explicit in discourse.[5] Thus, just as the conditional statement, "If x is north of y, then y is south of x" makes explicit a norm implicit in the material inference from "Atlanta is south of Boston" to "Boston is north of Atlanta," we claim that "A (exhaustively) explains B" makes explicit certain norms in material inferences involving A and B. However, as the example involving Boston and Atlanta illustrates, not all inferential practices are explanations. Therefore, we must circumscribe those inferential practices that will count as explanatory.

In the programmatic characterization of our view, above, we offered two requirements intended to effect this circumscription. The less remarkable of the two 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 our other condition (sturdiness), and still furnish similar solutions to the symmetry and relevance problems, so we will mostly take this requirement for granted in what follows.[6]

---

[2]Norton 2003 endorses a slightly different material theory of induction. With modest modifications, his account of material inference dovetails with our material account of explanation.

[3]This conception of material inferences can be traced back to (Sellars 1953).

[4]This claim expresses one of the central tenets of *logical expressivism*: a view about the role or function of logical vocabulary articulated by Brandom 1994.

[5]Recall that our analysis is only partial. Hence, this particular aspect of our view is not fully developed here.

[6]Below, we introduce a contextual element, Γ, which we also assume is true. Parallel points about alternative notions of quality control apply to Γ.

Our other, more interesting requirement restricts the kinds of material inferences underwriting explanation to what we called "sturdy" inferences above. Sturdy inferences are characterized by three features: nonmonotonicity, modal robustness, and superlative stability. Let us address each in turn.

First, the material inferences with which we are concerned are *nonmonotonic*. In the normative idiom introduced above, they preserve *entitlement*, such that if one is entitled to (assert) the premises, then one is entitled to (assert) the conclusion. Nonmonotonic consequence relations are disrupted when certain additional premises are added to the inference. Ordinary scientific practice is replete with them: e.g. "The liquid is acidic. So, it will turn blue litmus paper red." A good material inference of this sort might turn bad if additional premises or auxiliary hypotheses were added—e.g. "Chlorine gas is present" (since chlorine gas bleaches damp litmus paper).

Arguably, the turn to nonmonotonic inference is our most important departure from the prominent examples of previous inferential approaches to explanation, which take explanations to be classically deductive (and hence monotonic) inferences (Hempel 1965; Kitcher 1989). However, if the explanatory relation is inferential, then it is almost certainly nonmonotonic. For instance, a person's disease explains her symptoms, but a person's disease in conjunction with her taking an effective treatment does not.

The nonmonotonicity of explanation becomes even more striking when modeling practices—such as idealization, abstraction, the use of *ceteris paribus* clauses, and approximation—are brought to bear on explanation. Consider approximations as an example. While a model may claim that a particular explanatory generalization can be expressed using identity, e.g. $y = f(x)$, the truth about the target system is usually only that $y \approx f(x)$. The latter expression does not license much in the way of classical deductive inference. So if all explanations are deductive, then explanatory models could rarely be applied to their target systems. By contrast, approximations do license nonmonotonic inferences, given the purposes at hand and within the limits of the target system's characteristics. Thus, the inferences that underwrite scientific explanations are best understood as nonmonotonic inferences.

To represent the nonmonotonic material inferences that serve as the rudiments of explanatory practice, we will use the symbol $\mid\!\sim$. The most basic inferences in which we are interested have the form: $\Gamma, A \mid\!\sim B$. We read this expression as stating that anyone who is committed and entitled to A in context Γ is entitled to B. Γ represents the set of collateral commitments that form the epistemic context of the inference. As such, we assume that Γ contains the relevant background knowledge that the agent ought to possess—e.g. information about the experimental set-up and background knowledge relevant to the system in question—but does not contain an agent's total knowledge of the world.[7] Consider, for example, this inference:

Acid Γ, The liquid is acidic $\mid\!\sim$ The litmus paper turned red

The nonmonotonicity of inferences such as **Acid** leads naturally to the second characteristic of the material inferences underwriting explanation: their *modal robustness*. By definition, some new information can impugn a nonmonotonic inference. Conversely, other information does not affect these inferences at all. For instance, many pieces of additional information would leave the propriety of **Acid** untouched—e.g. if the litmus paper *were* soaked in acid for two minutes, if the person dipping the paper *were* wearing an orange shirt, if the test *were* conducted in Romania, etc. As our emphasis underscores, the subjunctive form of the verb *to be* naturally expresses this particular aspect of inferential practice. Consequently, the addition of this new information that does not affect a nonmonotonic inference's probity involves the supposition of *possible* states-of-affairs.[8] Thus, the material inferences with which we are presently concerned exhibit some degree of *modal robustness* in the sense that for any particular inference, there are sets of "subjunctive suppositions"

---

[7]In order to perspicuously represent material inferences, we introduce a formal structure with a nonmonotonic material consequence relation defined over a finite language. Partial specification of this formal structure is in the Appendix, below.

under which its propriety would remain unchanged and complementary sets of suppositions under which it would not.

To capture this modal feature of nonmonotonic material inference, we add a variable to material consequence relation, $\mid\!\sim$:

Modally Robust Acid Γ, The liquid is acidic $\mid\!\sim^{\Updownarrow W}$ The litmus paper turned red

One can think of this as a kind of quantification over the consequence relation. Where **Acid** depicted one inference, **Modally Robust Acid** captures a number of inferences. They share premises (Γ, "The liquid is acidic") and a conclusion ("The litmus paper turned red"), but differ in the possible auxiliary premises that are added. The notation $A \mid\!\sim^{\Updownarrow W} B$ means that the material inference from A to B remains good when any of the various suppositions comprising W are added as auxiliary premises.

The variable W captures the range of possible variation within which the inference remains good, what we will call an inference's "island of monotonicity." It could have been the case that the litmus test was conducted in Romania. It also could have been the case that the test was conducted in an atmosphere of pure oxygen. These possibilities do not impeach the inference. In effect, the inference from A to B functions monotonically on the "island" of W. If we think of W as including all of the possible variations in premises consistent with the inference from A to B, the notation $\mid\!\sim^{\Updownarrow W}$ captures the sense in which an inference is modally robust.[9] While we will sometimes write as if a sentence is or is not in W, strictly speaking W is a set of sets. This is because a statement that would defeat an inference by itself might not do so if further premises were added. For example, "the test was conducted in a chlorine atmosphere" would defeat **Modally Robust Acid**, but the pair of sentences, "the test was conducted in a chlorine atmosphere" and "the paper was shielded from the chlorine" would not. For convenience, we have and will continue to refer to these sets of sentences as *suppositions* and to W as a *set of suppositions*. Notice that the consequence relation expressed by classical deductive validity is a limiting case: when *any* supposition can be added to the premises of an inference, it is globally monotonic.

Importantly, all suppositions within a modally robust inference's island of monotonicity must be consistent with its premises. The classical conception of validity has the notorious consequence that any argument with inconsistent premises is valid. While such arguments do have a use in some contexts, inference rules like *Ex Falso Quodlibet* are controversial because they do not always support convincing inferences. This is particularly true of the inferences that underwrite explanations. For instance, suppose that the presence of acid explains why the litmus paper turned red. If suppositions are allowed to contradict premises, then this explanation could be correct even if the litmus paper still would have turned red had the acid been absent. However, this is generally regarded as a sign that the acid's presence does *not* explain why the litmus paper turned red. Hence, to preserve this intuitive feature of explanations, we assume that suppositions must be consistent with the premises of a modally robust inference.

This leads to the last distinguishing feature of the material inferences underlying explanation: their *superlative stability*. Many philosophers take stability of one sort or another to be a feature of explanations (Mitchell 2003; Skyrms 1980; Woodward 2003). While different philosophical discussions of explanation

---

[8]The English conditional "If P were the case, then Q would be the case" is formed with the subjunctive mood. Such sentences are sometimes called *counterfactual* conditionals. Other times, they are *distinguished* from counterfactual conditionals, which are identified with sentences of the form 'If P had been the case, then Q would have been the case.' In many languages, the antecedent of the latter sentence would be formulated in the past subjunctive, however, since English lacks a past subjunctive, the antecedent appears in the indicative past perfect. We will use the term *subjunctive* to pick out clauses of either type. In doing so, we reject the putative contrast between the two wherein properly 'counterfactual' conditionals are said to presuppose that their antecedent is false, while simple 'subjunctive' conditionals do not. As the following sentence shows, this is not always the case: "If Susan had contracted influenza, then she would have presented exactly the symptoms that she did."

[9]See the Appendix for a technical definition of the $\mid\!\sim^{\Updownarrow W}$ notation.

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*. Unsurprisingly, philosophers disagree about the proper characterization of these two elements of stability. Since it will prove useful, we shall first consider how Hempel fits within this bipartite framework, before presenting our own account of stability that we call "sturdiness."

Hempel's stability-bearers are so-called "lawlike generalizations," which must be expressible as universally quantified statements that trade exclusively in so-called qualitative predicates. This restriction on predicates guarantees that Hempel's stability-conditions include, among other things, different spatiotemporal regions. In other words, Hempel regarded laws of nature as stable across all times and places.

However, Hempel's account of stability faces two well-known difficulties. First, even by Hempel's own admission, his success in distinguishing laws of nature from accidental generalizations—particularly given his well-worn empiricist straitjacket—was limited at best. Second, by viewing laws of nature as universal statements, and also requiring such laws to figure in all explanations, Hempel's ability to capture biological and social scientific explanations is greatly imperiled.[10] Within the context of identifying the material inferences that underwrite explanations, we can see these two challenges as pulling in opposite directions. On the one hand, if an account is so lax as to count accidental generalizations as stable, then the resulting account of explanation will be too permissive. If, on the other hand, it does not count the generalizations of the special sciences as stable, then the resulting account of explanation will be too restrictive. Call this *Hempel's tightrope*.

To traverse this tightrope, we shall offer a markedly different approach to stability. Hempel's stability-bearer is a kind of *generalization* that then figures in an explanation as a premise in a covering law argument. By contrast, our stability-bearers belong to a class of *material inferences*. It is not the premises, on our view, that generate stability, but the relationship between the premises and their respective conclusions. Specifically, a sturdy inference must satisfy two conditions:

1. It is *modally robust* in the sense articulated above: the inference remains good under some range of suppositions consistent with the premises, and

2. Its range of suppositions is more inclusive than any other inference that could have the explanandum as conclusion.[11]

Inferences that meet these conditions are "sturdy consequences."[12]

All alone, condition (1) does not generate enough stability to keep us balanced on Hempel's tightrope. In particular, some inferences involving accidental generalizations are modally robust. For instance, consider a well-known counterexample to Hempel's DN model (Salmon 1977):[13]

John Jones   Γ, No males who consume birth control pills get pregnant,
John Jones is a male who consumes birth control pills
$\models^{\Uparrow W}$ John Jones does not get pregnant

---

[10]In fairness to Hempel, he had a variety of devices by which he sought to recover explanations in the special sciences: inductive-statistical explanations, explanation sketches, partial explanations, etc. However, many regard these maneuvers as having the untoward result that the special sciences are inferior simply for failing to emulate the physical sciences.

[11]Because we are analyzing exhaustive explanations, this requirement is apt. If we were analyzing a minimal criterion of explanatory relevance, this requirement would be too strong. Recall that we take the pragmatics of explanation to select particular elements of this exhaustive explanation as "the" explanation in a given context.

[12]See the Appendix for a formal definition of sturdy consequence.

[13]A structurally analogous problem arises if we use Henry Kyburg's (1965) hexed salt example.

In the original problem, Hempel's stability-bearers are the generalizations, and so the inference was held to be non-explanatory on the grounds that "No males who consume birth control pills get pregnant" is an accidental generalization, not a law. A similar problem arises if our view is restricted to condition (1). In **John Jones**, the island of monotonicity, W, cannot include any suppositions inconsistent with the premises. Nonetheless, the island remains very large, since it will include every other possibility. Thus, **John Jones** is modally robust, but not an explanation.

To see why **John Jones** is not an explanation, we need to bring the second condition of a sturdy consequence into play. This condition requires a comparison between one modally robust inference and all other inferences with the same conclusion. This is how we end up with *superlative* stability, rather than mere modal robustness. Intuitively, if we think of A as providing the best explanation of B, then we would expect A to be stable under a more inclusive set of suppositions than any less-than-best explanation C of B. More precisely, suppose:

$$\Gamma, A \mathrel{\vert\!\approx^{\Uparrow W}} B$$

and

$$\Gamma, C \mathrel{\vert\!\approx^{\Uparrow W'}} B$$

If $W' \subset W$, then the inference with island W would be more inclusive. To satisfy the second condition for sturdiness, an inference must have an island of monotonicity that includes the islands of all its competitors.

The interaction of the two conditions for a sturdy consequence captures the comparative evaluation that characterizes much explanatory reasoning. A paradigmatic way of evaluating candidate explanations is to see if 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 (as represented by Γ) and see if the reaction occurs when X but not Y is introduced into the experimental setting, and then do the converse. If the reaction occurs when X is present and Y is absent, but not when the chemicals' roles are reversed, then X is a better explanation of why the reaction occurred than Y. Upon iteration, we then arrive at the best explanation. Indeed, reasoning such as this animates, *inter alia*, controlled experiments, Mill's Method of Difference, and Lipton's (2004) well-known approach to Inference to the Best Explanation. Analogously, since the first condition prevents a modally robust inference from remaining good under suppositions inconsistent with its premises, one way to determine whether a particular inference satisfies the second condition is by seeing whether it holds under suppositions inconsistent with its competitors' premises. As we shall see below, this proves pivotal in our solution to the symmetry problem.

Following this approach, we can see that **John Jones** fails as an explanation because it's not sturdy. To appreciate this point, consider the following as an alternative explanation of John Jones' lack of pregnancy:

John Jones* Γ, No males get pregnant, John Jones is a male
                   $\mathrel{\vert\!\approx^{\Uparrow W'}}$ John Jones does not get pregnant

Next, consider the supposition that John Jones does not consume birth control pills. Quite clearly, adding this supposition to the premises of **John Jones*** does not impugn it. Hence it is part of the island of monotonicity for **John Jones***. However, since this supposition directly contradicts the premises of **John Jones**, this supposition cannot inhabit **John Jones**'s island of monotonicity. So, $W' \not\subseteq W$, and **John Jones** is therefore not a sturdy inference. If sturdiness is a necessary condition for explanation, as we propose, it follows that **John Jones** is not an explanation. On our view, statements such as "That John Jones consumed birth control pills does not explain why he is not pregnant" make it explicit that **John Jones** is not a sturdy inference, i.e.

that entitlement to assert that John Jones consumed birth control pills does not thereby entitle one to assert that he is not pregnant under the most inclusive range of suppositions.

The foregoing argument suggests that we have steered clear of at least one of the two dangers associated with Hempel's tightrope: namely that the view is so permissive as to incorrectly label inferences with accidental generalizations in their premises as explanations. Indeed, the second condition for a sturdy consequence will screen off explanatorily irrelevant information—whether it takes the form of an accidental generalization or not. Thus, as an added bonus, our treatment of this example also shows that our approach solves one problem that has plagued Hempel's approach to explanation.

The other risk for any account of explanation—the danger on the other side of Hempel's tightrope—is to circumscribe the domain of explanation so narrowly as to exclude the special sciences. Hempel 1942 faced this danger squarely, arguing that historical explanations fit a deductive-nomological form. In response, many argued that history provides a vast and robust body of explanations that do not fit Hempel's inference patterns, at least not obviously. In Michael Scriven's (1959) version of the argument, he drew on Roger Bigelow Merriman's (1918) explanation for why Cortés led a third expedition seeking an "island of gold" off the west coast of Mexico, after the previous two had failed.

Merriman explains simply that Cortés anticipated great wealth, and he had high confidence in his ability to lead the voyage better than the previous two captains he sent. Understood as a nonmonotonic inference, Merriman's inference was:[14]

Cortés   Cortés was confident and anticipated great wealth from the expedition
           $\models^{\Uparrow W}$ Cortés embarked on the expedition

As is typical in the inferential practice of historians, Merriman does not embellish his account with laws or other generalizations. At issue in the literature replying to Hempel is whether he should. Do historical explanations presuppose implicit lawlike generalizations? From the perspective of the foregoing discussion, we may see this debate as fundamentally about the stability of explanations in the special sciences. Lawlike generalizations were Hempel's means of conferring stability on explanations. If laws do not provide stability, something else needs to do so.

The argument against treating historical explanations as requiring laws is that candidate generalizations are either false or trivial. Candidate covering laws are easy enough to invent, and Scriven works through several suggestions, beginning with this one:

Cortés*   All confident wealth-seeking people undertake any venture which offers wealth,
           The third voyage envisioned by Cortés offered wealth,
           Cortés was confident and wealth-seeking
           $\models^{\Uparrow W}$ Cortés embarked on the expedition

As Scriven points out, while this lawlike generalization makes the argument deductively valid, the generalization is obviously false. Two lines of response are open to Hempel. Qualifications may be added in the hope of making the generalization true and preserving the deductive-nomological form. Alternatively, a probabilistic generalization could be used and the explanation treated as inductive-statistical, rather than deductive-nomological. The many essays that have broached this question concur that neither option is viable. Adding qualifications to the generalization makes it tautologous, intolerably vague, or so context-specific to as to apply only to Cortés at that time. And while statistical generalizations play an important role

---

[14]In this example and those that follow, we do not mention the background commitments of the inference, Γ. These commitments should be clear from the context of the example, and nothing in the discussion will turn on their representation.

in many special sciences, treating the generalization as a tendency adds nothing to historical explanation.

The main point of Scriven's critique is not that efforts to massage historical explanation into Hempel's form must always fail. Rather, he argues that none of Hempel's forms of inference conform to the practice of historians. Historians treat **Cortés** as an adequate explanation as it stands, without embellishment. Even if some version of **Cortés*** could be gerrymandered to satisfy philosophers' intuitions, from a historian's point of view, the additions are unnecessary. By contrast, our view of explanation as sturdy inference conforms to historians' inferential practice, at least on this point. Adding generalizations does not increase the modal robustness of **Cortés**. Rather, its robustness—the size of its island of monotonicity—is determined by facts of human psychology and the local factors that could have affected Cortés's decision. Moreover, unlike Hempel's notion of lawlikeness, sturdiness is, at root, a comparative concept. Explanations are sturdy relative to the alternative ways of inferring the explanandum. So long as every other way of inferring Cortés' leadership of the third expedition to find the "golden island" is modally robust up to the same W as **Cortés**, and both the premises and conclusion of **Cortés** are true, then Cortés' cupidity and confidence are explanatory. Exactly what alternative inferences are available is a question of historical scholarship, not philosophical analysis. If **Cortés** fails as an explanation, it is because there is a better explanation (i.e. a sturdier inference) available. Our account, therefore, not only includes the special sciences, it does a better job of conforming with their explanatory practices than alternative inferentialist accounts.

Our proposal is that the relationship between explanans and explanandum is an inferential one. To circumscribe the class of inferences that support explanation, we follow other accounts of explanation by demanding that an explanation be stable across a wide range of possibilities. Our stability-bearers are sturdy inferences. Sturdy inferences are stable in the sense that they still support their conclusions, even as different suppositions (i.e. members of W) change. Hence, these suppositions are our stability-conditions. In this section, we have made at least a *prima facie* case that treating sturdiness as a necessary condition for explanation can walk Hempel's tightrope, being neither so permissive as to count accidental generalizations as explanatory, nor so restrictive as to exclude explanations in the special sciences.

3   Solving The Symmetry Problem

Having presented our view, let us now turn to the problem of explanatory symmetry, which endures as a deep objection to treating explanation either as a form of inference or as a practice. After first presenting the problem and our solution to it (Section 3.1), we anticipate and rebut two objections to that solution (Section 3.2).

3.1   The Problem and Our Solution

Many physical laws are expressed by equations treating one variable as a function of others. These laws are symmetrical in the sense that the value of any variable can be deduced from the laws and values of the other variables. Since not all of these inferences are explanatory, many have concluded that an inferential account of explanation is at best insufficient. Failure to appropriately narrow the scope of explanatory inferences is, perhaps, the most compelling reason to reject an inferential approach to explanation entirely. Since pragmatic approaches tend to be less restrictive than inferential approaches, they appear even worse-equipped to resolve this difficulty.[15] For this reason, we focus on the inferential approach in what follows.

---

[15]Typically, pragmatic approaches purport to dissolve or bite the bullet on the symmetry problem (Richardson 1995; Risjord 2000; Fraassen 1980). Dialectically, this is a tougher sell than the more straightforward solution to the symmetry problem we offer here.

While the canonical version of the symmetry problem involves a flagpole and its shadow, Sylvain Bromberger's original example was 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:

Derived Shadow $\quad \tan\theta = \frac{H}{M}, \ \theta = 60^\circ, \ H = 1,454\,\text{ft} \ \vdash \ M = 839.5\,\text{ft}$

Derived Tower $\quad \tan\theta = \frac{H}{M}, \ \theta = 60^\circ, \ M = 839.5\,\text{ft} \ \vdash \ H = 1,454\,\text{ft}$

Treating the equation as a "law of geometric optics," both inferences count as explanations according to Hempel's criteria. But as Bromberger points out, only the first is plausibly an explanation.

Arguably, Kitcher's (1989) inferential approach resolves this version of the symmetry problem. However, Barnes 1992 offers a version of the symmetry problem that applies directly to Kitcher's unificationism. Barnes imagines a closed system of "Newtonian particles." Each particle has a position and velocity. 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. Newton's laws, after all, are a paradigm of scientific unification. But Newton's laws permit the deduction of past states from the present state as well. So, if Newton's laws fit Kitcher's criteria, both the forward and backward calculations must count as explanatory. But clearly, calculation of past states does not count as an explanation.

For many, inferential and pragmatic views' failure to address the symmetry problem suggests that a satisfactory account of explanation must invoke causal relationships. In Bromberger's example, **Derived Shadow** is explanatory because it tracks the causal relationship between the sunlight and the shadow, while **Derived Tower** does not. Similarly, the use of Newton's laws to calculate future states will follow the causes, while calculation of past states will not. Since there is nothing in the inferences (as understood by traditional inferential approaches) that will make this distinction, causal theorists of explanation have argued that the relationship between explanans and explanandum should not be treated as a form of inference.

Our turn from formal to material inference and the foregoing articulation of *sturdy* material inferences affords us distinctions that were out of our predecessors' reach. The symmetry counterexamples rest on the central intuition that inferring causes from their effects is not explanatory. Both Kitcher and the causal theorists base their accounts of explanation on features distinctive of causal inferences (or relationships). Our analysis takes the opposite approach: explanatory asymmetry arises not because inferences that track causes are especially sturdy, but because inferences from effects to causes are insufficiently sturdy.

To see why, let us develop a more concrete version of Barnes' example. 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 not moving. After the collision, A's velocity has changed, and B takes on a velocity. These motions can be explained through a relatively simple analysis. Assuming the masses of A and B are

equal and that their collision is elastic, conservation of kinetic energy entails that velocities of the balls will be:

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

In this formula, V_{1X} denotes the velocity of one "Newtonian particle" X up to its collision with another Y that is at rest, while V_{2X} and V_{2Y} denote the velocities of the particles at some time after this collision. Thus, to show that our account of explanation does not beget explanatory symmetries, we must establish that the following inference from the post-collision velocities to the pre-collision velocity of A is not sturdy:

Backwards Billiards   Velocity Law, $V_{2B} = .6\,\text{m/s}$, $V_{2A} = .8\,\text{m/s} \mathrel{\vdash^{\Uparrow W}} V_{1A} = 1\,\text{m/s}$

Here, we take A to be the billiard ball that collides at velocity V_{1A} with the resting billiard ball, B, resulting, per the Velocity Law, in the post-collision velocities V_{2A} and V_{2B}. More importantly, if we can show A's pre-collision velocity cannot be sturdily inferred from these post-collision velocities, then it will not be an explanation on our view.

Because we have a richer notion of inference than Hempel, we can offer a more nuanced taxonomy of inferences. For instance, while we will ultimately argue that the inference from post-collision to pre-collision velocities is not sturdy, it is *modally robust*, just as **Backwards Billiards** states. After all, this is a perfectly acceptable inference, and its island of monotonicity includes a variety of suppositions, e.g. that chalk is transfered from ball A to ball B, that the billiards player is left-handed, etc. Similarly, it excludes contingencies that would defeat it, e.g. that the balls are inelastic or that the table is uneven. As such, the first of our requirements for a sturdy inference is satisfied. This helps to diagnose the source of the symmetry problem: the non-explanations it begets share a good deal in common with explanations.

However, it is not a sturdy inference, for it will always be trumped by the inference to the velocity of ball A from some prior event. In other words, it fails to satisfy the second requirement of sturdy inferences. Suppose, for example, that, before colliding with B, A was at rest until struck by a third ball C. In that case, the following material inference would be acceptable (where V_{0C} is the velocity of ball C at time interval $t_0 < t_1$):

Billiards Defeater   Velocity Law, $V_{0C} = 1.17\,\text{m/s}$, $V_{1C} = 0.6\,\text{m/s} \mathrel{\vdash^{\Uparrow W'}} V_{1A} = 1\,\text{m/s}$

This inference is sturdy up to a different range of counterfactual conditions than **Backwards Billiards**. Like W in **Backwards Billiards**, W' in **Billiards Defeater** will also exclude suppositions related to the elasticity of the billiard balls and the unevenness of the table, as these are incompatible with the relationship between V_{0C} and V_{1A}. Indeed, W' will include all of the conditions included within **Backwards Billiards**' island of monotonicity, W. Thus $W \subseteq W'$. Unlike W, however, what happens after t_1 is irrelevant to the inference. Any possible event or condition that would change the A or B's velocities after their collision will be within W', but not within W. Suppose, for instance, that immediately after A and B collide, B hits a thick patch of felt and slows down to $0.4\,\text{m/s}$. Since the propositions that $V_{2B} = .6\,\text{m/s}$ and $V_{2B} = 0.4\,\text{m/s}$ are incompatible, this supposition cannot inhabit W.

To summarize, **Backwards Billiards** does not satisfy the second condition for a sturdy consequence. There is another inference, **Billiards Defeater**, with an island of monotonicity, W', and $W' \nsubseteq W$. Thus, pre-collision velocities cannot be sturdily inferred from post-collision velocities. Since we take sturdiness to be a requirement of explanation, our view captures the core intuition driving the symmetry problem: inferences such as **Backwards Billiards** are not explanations.

On our account, then, the inference of a cause from its effects will not, in general, be explanatory. The symmetry counterexamples arise from laws that let us infer the value of any one variable from the others.

The inference of a cause from its effects will always be trumped by another inference to the same cause, *viz.*, the inference from prior causes. The distinction between explanatory and non-explanatory inferences emerges not because there is something special about causality, but because retrodictive inferences are not sturdy.

3.2 Objections and Replies

In this section, we anticipate and rebut two potential objections to our solution of the symmetry problem. First, we tackle the objection that we have thrown out the baby with the bathwater, i.e. that in blocking the inferences that would beget the symmetry problem, we have also illicitly ruled out perfectly good explanations. Second, we consider the objection that our solution to the symmetry problem all too conveniently mislabels deductive inferences as nonmonotonic. Let us address each in turn.

Consider the first objection. In denying the sturdiness of the symmetry-mongering inference **Backwards Billiards**, we must show that analogous concerns will not jeopardize the sturdiness of the inference from A's pre-collision velocity to B's post-collision velocity, as this appears to be a perfectly good explanation. In other words, our solution to the symmetry problem must be consistent with the following being a sturdy inference:

Normal Billiards   Velocity Law, $V_{1A} = 1\,\text{m/s}$, $V_{2A} = .6\,\text{m/s} \mathrel{\vrule height 1.6ex depth 0pt width 0.5pt\kern-0.5pt\raise0.8ex\hbox{$\Uparrow^{W''}$}} V_{2B} = .8\,\text{m/s}$

Since we do not provide sufficient conditions for explanation, we cannot establish that our view *entails* that this will be an explanation. Rather, we aim for the more modest point that while **Billiards Defeater** trumps **Backwards Billiards**, no such analogue trumps **Normal Billiards**. Such an analogue would be:

Pseudo-Defeater   Velocity Law, $V_{0C} = 1.17\,\text{m/s}$, $V_{1C} = .6\,\text{m/s}$, $V_{1A} = 1\,\text{m/s}$,
$V_{2A} = .6\,\text{m/s} \mathrel{\vrule height 1.6ex depth 0pt width 0.5pt\kern-0.5pt\raise0.8ex\hbox{$\Uparrow^{W'''}$}} V_{2B} = .8\,\text{m/s}$

One of **Normal Billiards**' premises is that $V_{1A} = 1\,\text{m/s}$. **Pseudo-Defeater** is not more modally robust than **Normal Billiards** because there are other ways in which the velocity of ball A could have come about. Indeed, any alternative cause of $V_{1A} = 1\,\text{m/s}$ is in W'' of **Normal Billiards**, since they are compatible with it. However, many alternative causes are incompatible with $V_{0C} = 1.17\,\text{m/s}$, so $W'' \not\subseteq W'''$. Hence, the sturdiness of **Normal Billiards** is not threatened by **Pseudo-Defeater**, and our solution to the symmetry problem is consistent with **Normal Billiards** being a sturdy inference. We haven't thrown out the baby with the bathwater.

Turn now to our second objection—that our solution to the symmetry problem rests on mislabeling deductive inferences as nonmonotonic. In Section 2, we motivated the importance of nonmonotonicity with inferences such as **Cortés**. For historical explanations involving human motivation, it is altogether unsurprising that the explanation breaks down under a number of conditions. Hence, these kinds of examples make for easy pickings in building the case that explanations trade in nonmonotonic inference. By contrast, the inferences of this section—which deal with well-established, mathematically formulated physical laws— walk and talk like deductive inferences, so (the objection goes) to insist that they are (merely) sturdy appears ad hoc. It seems as if we have changed the classical turnstile into our fancy one just because it allows us to solve the symmetry problem.

Replying to this important objection requires further unpacking of the variety of inferences that constitute scientific practice. In Bromberger's formulation of the symmetry problem, he calls the following a "law of geometric optics:"

$$\tan \theta = \frac{H}{M} \hspace{4cm} \text{(Tangent Law)}$$

One might be puzzled by Bromberger's treatment of this mathematical relationship as a law of optics. The Tangent Law, after all, is true of any right triangle. It has nothing in particular to do with light as opposed, say, to the rigging on a sailboat. If something can be treated as a right triangle, trigonometry can be used to calculate the magnitude of its angles and sides. Understood in this way, the Tangent Law is uninterpreted in the sense that the variables are only quantitative relationships, not magnitudes of particular physical objects. Treating the Tangent Law as a law of geometric optics, as Bromberger does, presupposes that the variables are interpreted in a particular way.

Geometric optics is a way of modeling the propagation of light. By making the assumption that light travels as "rays," with a linear motion propagated perpendicular to the wave front, geometric relationships like the Tangent Law can be used to model some (but not all) phenomena associated with light. The Tangent Law is a law of geometric optics in the sense that its variables can be interpreted as the magnitudes of light rays. Light rays, however, are an abstraction, a tool used to model some of the features of light propagation. What Bromberger describes in his example, then, is the use of a model in explanation.

Since Hempel and Bromberger's halcyon days, philosophers of science have recognized and explored the importance of modeling in scientific practice. The role of models in the examples that generate the symmetry problem, however, has not been recognized. Earlier inferential approaches, such as Hempel's, treated all equations as interpreted with observable predicates. Doing so obscures the difference between two types of inferential practice: *calculations within* the model and the *applications* of those calculations *to* target systems for the purposes of explanation. To use the Tangent Law to explain the length of a shadow, one must model the sun-building-shadow system as a right triangle. This means treating the building as standing perpendicular to the ground and treating the ground as perfectly flat. Similarly, as we saw above, in order to use Newtonian mechanics to model the motions of billiard balls with the Velocity Law, we must treat the collisions as elastic and the motions as friction-free. To apply the model to explain the actual length of the shadow or the actual velocity of ball *B*, the properties of objects must be "close enough" to the modeled parameters, and the ignored factors cannot have "too much" influence on the objects' behavior.

The scare quotes around "close enough" and "too much" flag two features of the use of models in explanation. First, there is a pragmatic element that determines the usefulness of a model to explain a particular situation. How much is too much depends in part on the purposes at hand. Second, the explanation requires a commitment to the applicability of the model to the situation. Equations like the Tangent Law and the Velocity Law can be applied to a wide variety of target systems, but they are not indefinitely flexible. The suitability of a model to a particular kind of system, and the kind of explanations that the model will support, are empirical matters settled in the process of scientific inquiry.

Because of the role that modeling plays in the symmetry counterexamples, the explanations offered therein involve a somewhat more complex system of inference than has been appreciated. The explanation is not simply the deduction or calculation of one value from the law and other values. Rather, the explanation uses the formal relationships specified in the model in order to make inferences about the target system. Using geometric optics to explain the length of a shadow, or the conservation of kinetic energy to explain the velocity of a billiard ball, is not a single inference, but a chain of three interlocking inferential practices.

First, there are inferences that use the generalizations of the model to determine values for the model's variables. These will often be calculations or other monotonic inferences. When the shadow of the Empire State Building is modeled with geometric optics, for instance, a variety of deductive inferences can be made, including inferences such as **Derived Shadow** and **Derived Tower**. These inferences are monotonic, not undermined by the kinds of possible conditions to which the model might be applied. Considered only

as calculations with a model, these inferences are not explanations. Nothing has been explained yet: the calculations are applicable to phenomena as physically different as shadows and ropes. The explanation depends on further inferences that apply the model to a particular physical system.

The second inferential practice applies the model, within limits and to degrees of acceptable approximation, to a particular physical system. Mauricio Suárez 2003, 2004 has provided an inferentialist account of this step of the process. He argues that to interpret a model as representing a particular physical (target) system is to treat the model as providing a rule for inferences about the target system. This depends on known features of the physical system. It also has a pragmatic dimension, sensitive to our interest in particular aspects of the system and to the use to which the information will be put. Calculations with the model license nonmonotonic inferences, which constitute the third step of this process.

In this last inferential practice, inferences have premises and conclusions about the target system. These inferences involve commitments about physical states, properties, and events, and they have islands of monotonicity closely related to the limits within which the model has been applied. Some, but not all of these inferences will be sturdy, and they are the explanations properly so-called. In Bromberger's original example, these inferences should be understood as:

Materially Inferred Shadow $\qquad \tan\theta = \frac{H}{M},\ \theta = 60^{\circ},\ H = 1,454\,\text{ft} \mathrel{\vdash^{\Uparrow W}} M = 839.5\,\text{ft}$

Materially Inferred Tower $\qquad \tan\theta = \frac{H}{M},\ \theta = 60^{\circ},\ M = 839.5\,\text{ft} \mathrel{\vdash^{\Uparrow W}} H = 1,454\,\text{ft}$

The islands of monotonicity for these material inferences exclude the kinds of conditions that would make the model inapplicable. While both inferences are licensed by the model, the arguments of the previous section show that only **Materially Inferred Shadow** is sturdy, and thereby an explanation.

To summarize, the three inferential practices involve inferences between elements of the model, the application of the model to the target system, and finally inferences between claims about the target system itself. By themselves, calculations with a model and mere applications of a model to a target system fail to distinguish explanations from predictions, descriptions, classifications and other uses of models in scientific practice. It is only when such reasoning is supplemented by sturdy target-to-target inferences that we have an explanation.

4   Conclusion

In this paper, we have presented a new, "inferential practice" account of explanation. Its core idea is that explanatory vocabulary plays a role in a wider set of inferential practices. Chief among these roles is making explicit one's commitment to a norm in which the premises entitle one to infer a conclusion under the widest set of suppositions—what we have called a *sturdy inference*. As we saw, some earlier views, such as Hempel's, privileged inference to the detriment of practice. Other views, such as van Fraassen's, privileged practice to the detriment of inference. Both of these views offer unsatisfactory solutions to the problem of explanatory asymmetry. By contrast, we have seen that our view readily solves this problem, and thereby provides a promising alternative to the causal approaches that enjoy a certain privileged status in contemporary discussions of explanation.

This is but an opening salvo in a research program that we hope to develop in greater detail, by solving venerable problems in the explanation literature. Earlier inferential and pragmatic approaches to explanation

---

[0]Of course, models need not be constructed in a way that permits the kind of computation described here. Inferences within a model may also be nonmonotonic.

face a variety of problems.[16] 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 inferential and pragmatic views enjoy over causal approaches, which we also intend to develop in future work. First, inferential and pragmatic approaches have subsumed a wide variety of *non-causal* explanations within their framework. This suggests that either inferential or pragmatic considerations are latching onto explanation's deeper structures, and is further buttressed by the growing stockpile of examples of non-causal explanations arising from careful examination of scientific practice (Baker 2005; Batterman 2002; Bokulich 2011; Huneman 2010; Irvine 2015; Lange 2013a, 2013b; Rice 2015; Risjord 2005).

Second, in comparison to causal approaches, inferential and pragmatic approaches enjoy 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. Similarly, pragmatic approaches hold that competent language users can explain simply by correctly answering why-questions, which, in some cases, involves little more than fluency in a scientifically respectable theory.[17] As a result, both approaches avoid 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.).[18]

Finally, we intend to develop a richer pragmatics of explanation than those currently on tap,[19] to engage in more detailed comparisons with other leading accounts of explanation, and to apply our account to a number of problems in which the concept of explanation has figured prominently: e.g. inference to the best explanation, scientific realism, reductionism, and the science and values debates. By solving the problem of explanatory asymmetry, we have thereby cleared an important hurdle in addressing these larger concerns.

5 Appendix

In this appendix, we provide a formal account of the conditions necessary for a consequence relation to qualify as *sturdy*. Since sturdy inferences are a species of material inferences, we begin with a material consequence relation defined over a finite language, $\mathscr{L}_0$, consisting of atomic sentences $p, p_1, p_2, \ldots, p_n$. Let A, B, C, D range over sentences; Γ, Δ over sets of sentences; and W, W', W'' over sets of sets of sentences. In order to represent material incoherence, we extend $\mathscr{L}_0$ to include the constant $\bot$, i.e. $\mathscr{L} = \mathscr{L}_0 \bigcup \{\bot\}$. We stipulate that $\bot$ can neither appear to the left of the turnstile nor be embedded. Thus, $\vdash\!\sim$ maps sets of sets of sentences in $\mathscr{L}_0$ to sentences in $\mathscr{L}$, i.e. $\vdash\!\sim\; \subseteq \mathscr{P}(\mathscr{L}_0) \times \mathscr{L}$. We define $\vdash\!\sim$ as the relation over $\mathscr{L}$ such that $\Gamma \vdash\!\sim A$ iff Γ materially implies A, and $\Gamma \vdash\!\sim \bot$ iff Γ is materially incoherent.

We now have in place an object language $\mathscr{L}$ (an extension of $\mathscr{L}_0$ that includes $\bot$) and a meta-language that consists, *inter alia*, of $\vdash\!\sim$. We use a meta-theoretical conditional, $\Longrightarrow$, to formulate the properties of the structure $\langle \mathscr{L}, \vdash\!\sim \rangle$ as follows:

---

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

[17]Obviously, van Fraassen, whose broader empiricist commitments lead him to this point, is the clearest proponent of this line of thought in pragmatic approaches to explanation.

[18]Indeed, the semantic inferentialist 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).

[19]In particular, we hope to provide a normative-pragmatic account of why-questions that utilizes the framework first suggested by Khalifa 2011 and developed more fully by Millson 2014.

1. $\mathscr{L}_0 \mathrel{|\!\sim} \perp$

2. $\emptyset \mathrel{|\!\not\sim} \perp$

3. $\forall p, \forall \Delta \subseteq \mathscr{L}_0 (\Gamma, \Delta \mathrel{|\!\sim} \perp \implies \Gamma \mathrel{|\!\sim} p)$ (Ex Falso Fixo Quodlibet)

4. $\forall \Gamma \subseteq \mathscr{L}_0 (\Delta \in \Gamma \implies \Gamma \mathrel{|\!\sim} \Delta)$ (Reflexivity)

5. $\forall \Delta \subseteq \mathscr{L}_0 (\Gamma \mathrel{|\!\sim} p \not\!\!\implies \Gamma, \Delta \mathrel{|\!\sim} p)$ (Nonmonotonicity)

6. $(\Gamma \mathrel{|\!\sim} p_j$ and $\Gamma, p_j \mathrel{|\!\sim} p_k) \not\!\!\implies \Gamma \mathrel{|\!\sim} p_k$ (Non-transitivity)

The first two properties state, respectively, that the totality of $\mathscr{L}$ is incoherent and that the empty set is not. The principle of *Ex Falso Fixo Quodlibet* is a modification of *Ex Falso Quodlibet* that restricts 'explosion' to monotonic contexts only. The rationale for this restriction is that if a set of atomic sentences is nonmonotonically incoherent, then adding additional sentences to it may make it coherent, and therefore we are not licensed to infer an arbitrary atom from it (the original set). Reflexivity is a standard property of consequence relations and nonmonotonicity has been explained above.

The last property of the structure, however, is sure to surprise the reader. Nearly all consequence relations, including the standard nonmonotonic ones, are transitive and include the Cut rule as a structural rule. Nevertheless, as Morgan 2000 has demonstrated, a deduction theorem (i.e. $\Gamma, A \mathrel{|\!\sim} B \iff \Gamma \mathrel{|\!\sim} A \to B$) cannot hold for a nonmonotonic, reflexive, and transitive consequence relation. The provability of a deduction theorem is essential to our inferentialist approach to logical vocabulary, since it establishes that an object-language operator gives expression to, i.e. makes explicit, what would be given as a rule of inference in the meta-language, i.e. using '$\mathrel{|\!\sim}$'. We have already motivated the nonmonotonic character of material inference above. So, we must choose between reflexivity and transitivity.

There are some reasons to prefer a consequence relation that is reflexive but non-transitive. First, logicians of inferentialist and proof-theoretic persuasion have already explored systems in which Cut fails (Ripley 2011; Tennant 2014), whereas none seems inclined to pursue non-reflexive ones. Second, Ulf Hlobil 2016 defines a sequent calculus over structures with roughly the same properties as $\langle \mathscr{L}, \mathrel{|\!\sim} \rangle$ that includes left- and right-rules for several logical operators and proves that they are conservative extensions of an atomic language like our $\mathscr{L}_0$. In subsequent work, we hope to build on this system and to develop calculi with rules for logical and explanatory vocabulary. Finally, sets of explanatory statements often fail to license transitive inferences—e.g. the occurrence of the Big Bang does not explain why Adam ate the apple, even if there are true explanatory statements linking the Big Bang to event E_1, E_1 with E_2, and so on up to Adam's eating of the apple. Of course, reflexivity is not a property of explanations either, but we can capture this fact with appropriate constraints on sturdy inferences.

As noted above, sturdy inferences belong to the class of modally robust inferences—i.e. those material consequence relations that *would* remain good under some range of suppositions consistent with the premises. In order to represent the range of modal robustness exhibited by nonmonotonic material consequence relations we first quantify over $\mathrel{|\!\sim}$.

Quantified Material Consequence (QMC):

$$\Gamma, A \mathrel{|\!\sim^{\Uparrow W}} B \iff_{df} \begin{cases} 1. \, W \subseteq \mathscr{P}(\mathscr{L}) & \text{and} \\ 2. \, \forall \Delta \in W (\Gamma, A, \Delta \mathrel{|\!\sim} B) \end{cases}$$

QMC provides a precise way of talking about *sets* of material inferences. Since $\emptyset \in W$, our set must include $\Gamma, A \mathrel{|\!\sim} B$. Call this the *base inference* of the set. Sets of inferences are delimited by the set of sets of

sentences of $\mathscr{L}$ whose members may be added to the premises without impugning the base inference. The resulting set thus contains $|W|$ many inferences. When W is the set of all sets of the language, i.e. $\mathscr{P}(\mathscr{L})$, the consequence relation is globally monotonic—in which case we simply write $\mathrel{\vdash^{\uparrow}}$.

The reason for circumscribing these sets of inferences by subsets of $\mathscr{P}(\mathscr{L})$ is that atomic sentences that would, by themselves materially defeat an inference if added to the premises, need not defeat that inference if they are added along with other atomic sentences. As noted above, we think of these sets of atomic sentences as *suppositions* and of W as a set of such suppositions. Thus $\Gamma, A \mathrel{\vdash^{\uparrow W}} B$ says that the (base) inference from A to B *would* remain materially good in context Γ, even if the suppositions in W *were* to hold.

However, $\Gamma, A \mathrel{\vdash^{\uparrow W}} B$ does not imply that the base inference would become materially bad if we were to add something from *outside* of W, and so **QMC** does not capture an inference's range of modal robustness. This range consists in the set of *all* suppositions that could be added to its premises without undermining its correctness. The following definition permits us to identify consequence relations by their range of modal robustness. (For the remaining meta-theoretical claims, we omit reference to $\mathscr{L}$ and $\mathscr{P}(\mathscr{L})$.)

Modally Robust Consequence (MRC):

$$\Gamma, A \mathrel{\vdash^{\Uparrow W}} B \iff_{df} \begin{cases} 1.\ \Gamma, A \mathrel{\vdash^{\uparrow W}} B & \text{and} \\ 2.\ \forall W'(\Gamma, A \mathrel{\vdash^{\uparrow W'}} B \Longrightarrow W' \subseteq W) & \text{and} \\ 3.\ \forall \Delta (\Gamma, A, \Delta \mathrel{\vdash^{\uparrow}} \bot \Longrightarrow \Delta \notin W) \end{cases}$$

The second condition of **MRC** says that W contains *all* those sets of sentences whose addition would not defeat the base inference. The third condition prohibits the elements of W from being logically inconsistent with the premises, thereby ensuring that the conclusion does not follow trivially from *Ex Falso Fixo Quodlibet*. When the consequence relation $\mathrel{\vdash^{\Uparrow W}}$ holds, we say that the base inference is *modally robust up to W*.

A sturdy inference must not only be modally robust; it must be superlatively stable. In other words, there must be no material inference with the same conclusion in the same context whose range of modal robustness is strictly greater than that of a sturdy inference. The following meta-theoretical conditional, in which $\mathrel{\vdash^{\updownarrow}}$ denotes sturdy consequence relations, captures these characteristics.

Sturdy Consequence (SC):

$$\Gamma, A \mathrel{\vdash^{\updownarrow}} B \Longrightarrow \begin{cases} 1.\ \Gamma, A \mathrel{\vdash^{\Uparrow W}} B & \text{where } A \neq B, \text{ and} \\ 2.\ \forall C\,(\Gamma, C \mathrel{\vdash^{\Uparrow W'}} B \Longrightarrow W' \subseteq W) & \text{where } B \neq C \end{cases}$$

The first condition states that sturdy inferences are modally robust, with the proviso that their non-contextual premises are distinct from their conclusions. This ensures that unlike other material inferences, sturdy inferences are not reflexive. The second condition of **SC** says that a sturdy inference must be modally robust up to some W such that W includes all those suppositions under which any sentence from $\mathscr{L}$ materially implies B, excluding B itself. Since a sturdy inference's island of monotonicity encompasses all other islands from which the conclusion can be drawn, there is no further need to compare sets of suppositions, and thus the turnstile for the sturdy consequence relation may omit 'W'.

6   Compliance with Ethical Standards

References

Achinstein, Peter. 1983. *The Nature of Explanation*. ix, 385. New York: Oxford University Press.

Baker, Alan. 2005. "Are There Genuine Mathematical Explanations of Physical Phenomena?" *Mind* 114 (454): 223–238.

Barnes, Eric C. 1992. "Explanatory Unification and the Problem of Asymmetry". *Philosophy of Science* 59 (4): 558–571.

Batterman, Robert W. 2002. *The Devil in the Details : Asymptotic Reasoning in Explanation, Reduction and Emergence*. X–144 p. New York: Oxford University Press.

Bokulich, Alisa. 2011. "How Scientific Models Can Explain". *Synthese* 180 (1): 33–45.

Brandom, Robert. 2008. *Between Saying and Doing : Towards an Analytic Pragmatism*. Oxford ; New York: Oxford University Press.

— . 2015. *From Empiricism to Expressivism*. Cambridge, Mass.: Harvard University Press.

— . 1994. *Making It Explicit : Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.

Bromberger, S. 1966. "Why-Questions". In *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, ed. by R. Colodny, 86–111. Pittsburgh: University of Pittsburgh Press.

Bromberger, Sylvain. 1965. "An Approach to Explanation". In *Studies in analytical philosophy*, ed. by R.J. Butler, 2:72–105. Oxford: Blackwell.

Faye, Jan. 2007. "The Pragmatic-rhetorical Theory of Explanation". In *Rethinking Explanation*, ed. by J. Persson and P. Ylikoski, 252:43–68. Boston studies in the philosophy of science. Dordrecht: Springer.

Fraassen, B.C. van. 1980. *The Scientific Image*. New York: Clarendon Press.

Friedman, Michael. 1974. "Explanation and Scientific Understanding". *Journal of Philosophy* 71 (1): 5–19.

Garfinkel, A. 1981. *Forms of Explanation: Rethinking the Questions in Social Theory*. New Haven: Yale University Press.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*. New York: Free Press.

Hempel, Carl G. 1942. "The Function of General Laws in History". *The Journal of Philosophy* 39 (2): 35–48.

Hlobil, Ulf. 2016. "A Nonmonotonic Sequent Calculus for Inferentialist Expressivists". In *Logica Yearbook 2015*, ed. by Pavel Arazim and Michal Dančák, 87–105. London: College Publications.

Huneman, Philippe. 2010. "Topological Explanations and Robustness in Biological Sciences". *Synthese* 177 (2): 213–245.

Irvine, Elizabeth. 2015. "Models, Robustness, and Non-causal Explanation: A Foray into Cognitive Science and Biology". *Synthese* 192 (12): 3943–3959.

Khalifa, K. 2011. "Contrastive Explanations As Social Accounts". *Social Epistemology* 24 (4): 263–284.

Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World". In *Scientific explanation*, ed. by P. Kitcher and W. C. Salmon, XIII:410–506. Minneapolis: University of Minnesota Press.

Kitcher, Philip, and Wesley C. Salmon. 1987. "Van Fraassen on Explanation". *Journal of philosophy* 84 (6): 315–330.

Kyburg, Henry E. 1965. "Salmon's Paper". *Philosophy of Science* 32 (2): 147–151.

Lange, Marc. 2013a. "Really Statistical Explanations and Genetic Drift". *Philosophy of Science* 80 (2): 169–188.

— . 2013b. "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for the Philosophy of Science* 64 (3): 485–511.

Lipton, P. 2004. *Inference to the Best Explanation*. International library of philosophy and scientific method. Routledge.

Merriman, Roger Bigelow. 1918. *The Rise of the Spanish Empire in the Old World and in the New*. Vol. 1–4. v. 1–4. New York: Macmillan.

Millson, Jared. 2014. "How to Ask a Question in the Space of Reasons: Assertions, Queries, and the Normative Structure of Minimally Discursive Practices". PhD thesis, Emory University.

Mitchell, Sandra D. 2003. *Biological Complexity and Integrative Pluralism*. Cambridge University Press.

Morgan, Charles G. 2000. "The Nature of Nonmonotonic Reasoning". *Minds and Machines* 10 (3): 321–360.

Norton, John D. 2003. "A Material Theory of Induction". *Philosophy of Science* 70:647–670.

Peregrin, Jaroslav. 2014. *Inferentialism: Why Rules Matter*. London: Palgrave Macmillan.

Rice, Collin C. 2015. "Moving beyond Causes: Optimality Models and Scientific Explanation". *Noûs* 49 (3): 589–615.

Richardson, Alan. 1995. "Explanation: Pragmatics and Asymmetry". *Philosophical studies* 80 (2): 109–129.

Ripley, David. 2011. "Paradoxes and Failures of Cut". *Australasian journal of philosophy* 91 (1): 139–164.

Risjord, Mark. 2005. "Reasons, Causes, and Action Explanation". *Philosophy of the Social Sciences* 35 (3): 294–306.

— . 2000. *Woodcutters and Witchcraft: Rationality and Interpretive Change in the Social Sciences*. ix, 201. Albany: State University of New York Press.

Salmon, Wesley C. 1977. "A Third Dogma of Empiricism". In *Basic Problems in Methodology and Linguistics: Part Three of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science*, ed. by Robert Butts and Jaakko Hintikka, ed. by Jaakko Hintikka, ed. by Robert Butts, 149–166. The Western Ontario Series in Philosophy of Science. Dordrecht-Holland: Dordrecht Reidel.

— . 1989. "Four Decades of Scientific Explanation". In *Scientific Explanation*, ed. by Philip Kitcher and Wesley Salmon, 3–219. Minneapolis: University of Minnesota Press.

Schurz, Gerhard. 1999. "Explanation As Unification". *Synthese* 120 (1): 95–114.

Schurz, Gerhard, and Karel Lambert. 1994. "Outline of a Theory of Scientific Understanding". *Synthese* 101 (1): 65–120.

Scriven, Michael. 1959. "Truisms as the Grounds for Historical Explanations". In *Theories of History: Readings from Classical and Contemporary Sources*, ed. by P. Gardiner, 443–475. Glencoe, IL: Free Press.

Sellars, Wilfrid. 1957. "Counterfactuals, Dispositions, and the Causal Modalities". In *Minnesota Studies in The Philosophy of Science, Vol. II*, ed. by Grover Maxwell, ed. by Herbert Feigl, ed. by Michael Scriven, ed. by Grover Maxwell, 225–308. Minneapolis: University of Minnesota Press.

— . 2007. *In the Space of Reasons : Selected Essays of Wilfrid Sellars*. Ed. by K. Scharp and R. Brandom. Cambridge, Mass.: Harvard University Press.

— . 1953. "Inference and Meaning". In Sellars 2007, 3–27.

— . 1954. "Some Reflections on Language Games". In Sellars 2007, 28–56.

Skyrms, Brian. 1980. *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven, CT: Yale University Press.

Stovall, Preston. 2015. "Chemicals, Organisms, and Persons: Modal Expressivism and a Descriptive Metaphysics of Kinds". PhD thesis, University of Pittsburgh.

Strevens, M. 2008. *Depth: An Account of Scientific Explanation*. Cambridge: Harvard University Press.

Suárez, Mauricio. 2004. "An Inferential Conception of Scientific Representation". *Philosophy of Science* 71 (5): 767–779.

— . 2003. "Scientific Representation: Against Similarity and Isomorphism". *International Studies in the Philosophy of Science* 17 (3): 225–244.

Tennant, Neil. 2014. "Natural Deduction and Sequent Calculus for Intuitionistic Relevant Logic". *The Journal of Symbolic Logic* 52 (03): 665–680.

Thomasson, Amie L. 2007. "Modal Normativism and the Methods of Metaphysics". *Philosophical Topics* 35 (1/2): 135–160.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. vi, 410. New York: Oxford University Press.

— . 2014. "Scientific Explanation". In *The Stanford Encyclopedia of Philosophy*, Winter 2014, ed. by Edward N. Zalta.