Expressivism about explanatory relevance

Josh Hunt[1]

Abstract
Accounts of scientific explanation disagree about what's required for a cause, law, or other fact to be a reason why an event occurs. In short, they disagree about the conditions for explanatory relevance. Nonetheless, most accounts presuppose that claims about explanatory relevance play a descriptive role in tracking reality. By rejecting the need for this descriptivist assumption, I develop an expressivist account of explanatory relevance and explanation: to judge that an answer is explanatory is to express an attitude of *being for being satisfied by that answer*. I show how expressivism vindicates ordinary scientific discourse about explanation, including claims about the objectivity and mind-independence of explanations. By avoiding commitment to ontic relevance relations, I rehabilitate an irrealist conception of explanation.

Keywords Explanation · Expressivism · Quasi-realism · Relevance · Empiricism

1 Introduction

The puzzling nature of scientific explanation is manifest in a long and ever-growing list of competing philosophical accounts. At its core, philosophical disagreement about explanation concerns (i) which kinds of ontic structures are explanatorily relevant and (ii) under what conditions. Candidates include laws of nature, causal mechanisms, difference-makers, causal patterns, grounds, and many more. Following van Fraassen (1980) and Skow (2016), I will treat (successful) explanations as (correct) answers to why questions. More precisely, a scientific explanation of an event is the *reasons why that event occurs*. These reasons why are the relevant ontic structures, such as causes or mechanisms. They are the states of affairs that in some sense produce or are responsible for the event.

✉ Josh Hunt
    joshhunt@mit.edu

1   Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA, USA

My question concerns the relation between these ontic structures and explanatory relevance. What—if anything—makes an ontic structure explanatorily relevant? Most accounts of explanation presuppose that claims about explanatory relevance are *descriptive* or *representational*: they aim to mirror or represent explanatory relevance relations in reality. On this construal, different accounts of explanation disagree about which explanatory relevance relations obtain in reality. They are engaged in an ontological dispute, arising from the difficulty of knowing which explanatory relevance relations obtain. Section 2 demotivates this descriptivist assumption.

Section 4 advocates an alternative approach to understanding explanatory relevance relations. Rather than treat explanatory relevance descriptively, I will provide a *non-descriptive* analysis. Specifically, I propose to be an expressivist about explanatory relevance. Claims about explanatory relevance express acceptance of a set of norms governing what answers ought to satisfy us. My account is irrealist because it denies that claims about explanatory relevance are necessarily descriptive. At the same time, I agree with realists that an adequate account of scientific explanation should vindicate ordinary scientific discourse about explanation. This makes my account into a form of *quasi-realism* (Blackburn, 1993), as opposed to an error theory. For instance, the following sorts of claims are true on the account that I defend: "lightning explains thunder," "socks do not explain why laptops work," and "the correct explanation does not depend on us." In its contemporary forms, expressivism accomplishes this vindicatory project by embracing minimalism about truth (Sect. 5). This deflationary approach to truth allows non-descriptive claims to be true or false, without requiring commitment to underlying ontic truth-makers.[1]

I deny that we need to interpret claims about explanatory relevance as representing states of affairs. Instead, it suffices to focus on the functional roles that scientific explanatory discourse plays, including its roles in coordinating prediction and control. Despite being irrealist, my account is not against either realism or metaphysics. If my account succeeds, then robust metaphysical commitment to explanatory relevance relations is simply *orthogonal* to the aim of vindicating explanatory reasoning. Such commitments might be necessary for other philosophical aims, but not this one.[2]

2 Explanatory relevance relations

Wherever its methods have been applied, science has enjoyed unparalleled success in predicting and controlling natural phenomena. Given this success, we ought to view science as a "paradigm of human rationality," to borrow a phrase from van Fraassen (1994, p. 328). Barring countervailing epistemic reasons, our goal should not be to revise ordinary scientific discourse but to vindicate it, in the hopes that we

---

[1] My use of 'irrealism' is inspired by Wright (1988, p. 30). In my sense, irrealism about a particular discourse permits us to remain agnostic about whether this discourse *also* represents features of the world. Irrealism focuses on the discourse or practice itself rather than claims about what there is in reality. I thank a referee for suggesting this helpful characterization.

[2] I thank Maegan Fairchild for suggesting this notion of the 'orthogonality' of philosophical aims.

can systematically understand how it works so well. To vindicate a discourse, one aims for what Gibbard calls an *internally adequate* philosophical theory. A theory is internally adequate provided that it accounts for the claims and judgments that are internal to a given discourse, at least those claims that are intelligible (2003, p. 186). In the case of ordinary moral judgments, this amounts to vindicating claims such as "murder is wrong" and "keeping promises is good." Moral claims like this serve us well in coordinating human action and affairs. Given their success for this aim, we are interested in vindicating them.

Here, my goal is to vindicate a subset of ordinary scientific discourse: explanatory claims and judgments. These judgments play a central role in our capacities for prediction and control. For instance, scientists have recently confirmed that human-caused climate change is responsible for an increase in extreme weather events.[3] As is well known, climate change is itself the result of greenhouse gas emissions from human activities. These inferences constitute triumphs of scientific reasoning that any philosophical account of explanation ought to vindicate. To do so, we must consider how scientists discern which events are explanatorily relevant for others, and which are irrelevant.

Most philosophical accounts of scientific explanation presuppose that scientists discover explanatory relevance relations. An *explanatory relevance relation R* relates an explanandum *P* to the set of facts that constitute the explanans for *P*. A putative explanans is explanatorily relevant provided that it bears the proper relation *R* to the explanandum, such as the relation of being a cause or a ground.[4] Different philosophical accounts of explanation disagree about these relations: they disagree about what makes any given fact explanatorily relevant to a given explanandum. As Skow puts it, this amounts to a disagreement "over what it takes to be a reason why" (2016, p. 42).[5]

It is convenient to speak uniformly of explanatory relevance relations as relations between facts, but other regimentations are available. Some will prefer to treat at least some explanatory relevance relations as relations between events, or even as sentential connectives rather than relations. For instance, if facts are abstract entities while causation is a relation between concreta, then causation plausibly relates events. Partly to avoid this worry while still speaking in terms of facts, Skow expresses sympathy for treating causation as a sentential connective (2016, p. 34). This would relocate causation from a theory's ontology to its 'structure' or

---

[3] See the recent IPCC report on climate change for details (2021, Ch. 11).

[4] In contrast, van Fraassen introduces a relevance relation *R* as part of the definition of an explanatory why-question *Q* (1980, p. 143). This has the unfortunate consequence that people who disagree about the proper relevance relation do not ask the same explanatory why-question. Yet often, we ask the same why-question while disagreeing about what is—or could be—a legitimate answer, owing to a disagreement about the relevance relation.

[5] Similarly, Potochnik claims that "the kind of responsibility that is explanatory is primarily what is at issue among different accounts of explanation," which she calls "explanatory dependence relations" (2017, pp. 125, 127). Craver claims that "The philosophical dispute about explanation...is about which kinds of ontic structure properly count as explanatory and which do not" (2014, p. 29).

ideology.[6] Nonetheless, for convenience, I will speak uniformly in terms of 'relations,' where I intend this to include the possibility of corresponding sentential connectives instead.[7] Likewise, I will have in mind a broader notion of the 'ontic' that includes claims about metaphysical structure.

What I will call *descriptivist realism* takes explanatory relevance relations to exist as part of the world's ontology, in this broad sense of 'ontology.' It is descriptivist because it takes claims about explanatory relevance to track or represent features of reality. It is realist because it views these relations as being objective and mind-independent (at least when a given explanation does not involve consciousness). So construed, explanatory relevance relations qualify as *ontic relevance relations*.

If we assume descriptivist realism, then different accounts of scientific explanation are engaged in an ontological dispute about which ontic relevance relations actually obtain. Whereas Hempel's (1965) deductive-nomological model treats laws as explanatorily relevant (indeed, explanatorily required), Skow denies that laws are ever reasons why (2016, p. 84). Instead, Skow holds that only causes and grounds of an event can be reasons why for that event (2016, p. 124). For those who restrict the relevance relation to causes, there remains the question of *which causes* count as explanatorily relevant; Strevens calls this "the problem of explanatory relevance" (2008, p. 49). Strevens' account identifies the *difference-makers* as a special, explanatorily relevant subset of the explanandum's causes. In cases of "distinctively mathematical explanations," Lange's (2017) account is even more restrictive, denying that *any causes* of an event are explanatorily relevant. At the other extreme, both Railton (1981) and Lewis (1986) classify *any* information about the event's causal history as explanatorily relevant.

This rampant disagreement indicates that descriptivist realism faces at least a serious epistemic problem: ontic relevance relations appear to be relatively difficult to know about. Nonetheless, it remains clear how descriptivist realism could in principle vindicate ordinary scientific discourse about explanation. Scientists would succeed when their explanatory reasoning correctly represents the explanatory structure of the world. The objectivity and mind-independence of this structure would ground the objectivity and mind-independence of scientists' veridical explanatory claims. Sections 2.1–2.2 question whether this vindicatory story is so straightforward, noting how descriptivist realism faces at least two more worries, besides the epistemic one.

It is much less clear how descriptivist anti-realisms could succeed at vindicating scientific explanatory discourse. An explanatory anti-realist would deny that explanatory relevance relations are objective or mind-independent. Perhaps most simply, this denial would take the form of an individual or community-based subjectivism,

---

[6] For this notion of structure, see Sider (2011, pp. 5, 12). It generalizes the notion of a perfectly natural property. I thank a referee for suggesting I broaden my discussion in this way to include the possibility that causation is a sentential connective.

[7] Indeed, by the lights of the irrealist position I ultimately defend, it is implausible that determining whether causation is a (i) relation between facts, (ii) relation between events, or (iii) sentential connective matters for the project of vindicating scientific explanatory discourse.

on which explanatory relevance relations depend on features of agents.[8] For convenience, I will refer collectively to these kinds of descriptivist anti-realisms as 'relativism.' At least in its basic forms, relativism seems unable to vindicate ordinary scientific discourse about explanation. Causal, mechanistic, and nomological explanations do not depend on features of agents (at least when the phenomena-to-be-explained don't involve people). It seems central to the concept of explanatory relevance that *what counts as relevant* is objective and mind-independent. Explanatory relevance does not depend on our interests or goals.[9] These considerations dampen the prospects of descriptivist relativism about explanation.

If we treat explanatory relevance descriptively, the constraints of objectivity and mind-independence speak in favor of realism. Hence, I set aside anti-realism—particularly relativism—as a viable vindicatory strategy. We will see in Sect. 3 how my rejection of relativism initially seems to pose a dilemma for non-descriptivism about explanatory relevance. My goal in the rest of this section is to demotivate the need for descriptivist realism, or 'realism' for short. As stated before, I am not arguing against realism *per se*. Rather, I aim to show that realism faces unexpected problems when it comes to vindicating scientific explanatory discourse. By demotivating descriptivism, I take this section to motivate a non-descriptivist approach to explanation. Section 4 develops one such approach using expressivism. Sections 5 and 6 describe how expressivism can vindicate ordinary scientific discourse about explanation, including its objectivity and mind-independence.[10]

2.1 A problem for descriptivism about causal relevance relations

Descriptivist realism faces a special problem when it comes to causal explanations. Admittedly, this problem has restricted scope, since causal explanatory relevance relations are a specific class of explanatory relevance relations. Nevertheless, due to the importance of causal relevance relations for making sense of scientific explanatory practice, the problem is serious. It shows that descriptivism has difficulties vindicating ordinary scientific discourse about causal explanation. For many, this may be sufficient motivation to consider a non-descriptivist approach to explanatory relevance relations.

Consider the following oft-discussed scenario.[11] Thomas throws a rock at his window hard enough to break it, but Catherine catches the rock before it hits. What explains why Thomas's window did not break? We take for granted that Catherine's catch is part of the explanation. Her catch is explanatorily relevant: it prevented the

---

[8] I lack space to discuss other forms of anti-realism, such as fictionalism. Blackburn (2005) provides reasons to think that expressivism is preferable to fictionalism.

[9] Of course, context can determine which facts are conversationally or pedagogically relevant to include while *giving* or *representing* an explanation, based on the interests of the audience or explainer. But this is a point about the pragmatics of explanation, not about the ontic reasons why an event occurred. See Hempel (1965, p. 426) or Skow (2016, p. 19) for similar points.

[10] There remain subtle questions regarding the relationship between expressivism and certain kinds of sophisticated relativism, such as MacFarlane's (2014). For discussion, see Field (2009, pp. 252, 272ff., 2018, pp. 3, 15, especially n. 20) and Beddor (2019, p. 533). I set these questions aside here.

[11] See e.g. Hitchcock (2001, pp. 276–277) or Skow (2016, p. 76).

rock from breaking the window. Less clear is whether Thomas's throw is explanatorily relevant. No doubt, Thomas's throw helped cause Catherine to catch the rock, which caused the window not to break. If causation is transitive, then Thomas's throw is a cause of the window's not breaking. Lewis's (1986, pp. 216–217) account of explanation would endorse this verdict. Yet, for many, it sounds wrong to say that Thomas's throw helps explain why the window did not break. Thomas's throw nearly broke the window! Indeed, in many ordinary and scientific cases, we deny that causation is transitive (Hitchcock, 2001).

One argument for demotivating explanatory descriptivism goes as follows. First, recall my methodological commitment that we ought to vindicate and respect ordinary scientific discourse about causal explanatory relevance, insofar as possible. Next, we notice that scientific practice relies on *selective causal relevance relations*: these pick out only a subset of the events in an explanandum's backwards light cone. Only this privileged subset is causally or explanatorily relevant. In the scenario with the unbroken window, selective relevance relations characterize Thomas's throw as being neither causally nor explanatorily relevant to his window's not-breaking.[12]

Perhaps surprisingly, descriptivist realism stands in tension with the selective causal relevance relations ubiquitous in science. As Neo-Russellians about causation have argued, there is only one mind-independent and objective causal relevance relation. Call this a *complete history relevance relation*: it counts all events in an explanandum's backwards light cone as causally relevant (Field, 2003, p. 439). So the argument goes, any more selective relevance relation constitutes an anthropomorphic carving up of these events in the backwards light cone.[13] Returning to Thomas and his window, the complete history relevance relation counts Thomas's throw as a cause of his window's not breaking. Yet, as we have seen, this verdict is incompatible with ordinary scientific practice. Consequently, it appears that descriptivist realism faces a challenge when it comes to vindicating ordinary scientific discourse about causal explanatory relevance relations.[14]

Perhaps there are other forms of descriptivist realism that avoid this problem. Regardless, I take it as one motivation for considering an alternative approach. As we will see in Sect. 4, expressivism provides a stratagem for avoiding this problem. By denying that we need to assume descriptivism, we can satisfy the constraints of objectivity and mind-independence while respecting the use of selective relevance relations within science.

---

[12] Even on causal accounts of explanation, causal relevance can come apart from explanatory relevance: an explanatory relevance relation can select a subset of an event's causes (Potochnik, 2017, p. 153).

[13] Frisch (2022, p. 459) summarizes this argument using Field's example of a man praying to put out a fire while someone else sprays it with a hose. The praying is in the backwards light cone of the fire going out, but intuitively it is not a cause.

[14] In response to Field's Neo-Russellian argument, Woodward (2007) relies on a notion of coarse-graining that I believe is agent-dependent. In general, claims about causal relations depend on which variables agents choose to use in their causal models (Woodward, 2016).

2.2 Is descriptivism necessary?

Of course, not all accounts of explanation endorse causal relevance relations. Some endorse nomological or mathematical relevance. To demotivate descriptivist approaches to these kinds of relevance relations, I turn to a second type of argument. This argument begins by recalling philosophers' extensive disagreements about explanatory relevance. If descriptivism is required to make sense of explanatory discourse, then settling these ontological disagreements should have implications for vindicating scientists' ordinary explanatory claims. Yet, as I illustrate below, many competing realist accounts are equally capable of vindicating explanatory reasoning. We would not expect this result if descriptivism were necessary to vindicate explanatory discourse. I therefore contend that the central debates inspired by descriptivism are orthogonal to the aim of vindicating explanatory reasoning. To support this diagnosis, I will consider two paradigmatic disputes: (i) the debate between Lange vs. Skow on non-causal explanations and (ii) the debate between Hempel vs. Skow on the explanatory relevance of laws of nature.

Consider first the debate between Lange and Skow, using a case called 'STRAWBERRIES': Jane has 23 strawberries and is trying to divide them evenly among her three children. This is an impossible task: three does not divide 23. Why does Jane fail? According to Lange (2013), STRAWBERRIES is an example of a distinctively mathematical explanation of a physical fact. The mathematical fact that three does not divide 23 explains why Jane cannot succeed. Lange argues that due to the necessity of this mathematical fact, the actual causal mechanism for her failure is explanatorily irrelevant. Although the causal structure of the world figures in Jane's failure, what matters is how the given mathematical fact constrains this causal structure, not the causes themselves (Lange, 2017, p. 20).[15]

In contrast, Skow's view privileges causation. He argues that this mathematical fact is not a reason why Jane fails. Instead, only causes (and grounds) can be reasons why, including that she has three children and that she has 23 strawberries. These are causes of her inability to divide the strawberries evenly. Skow interprets the relevant mathematical fact as a reason why these causes are reasons (2016, p. 114). More explicitly, the fact that three does not divide 23 is a reason why ⟨*the fact that Jane has three children*⟩ is a cause of her failure. On Skow's view, this mathematical fact is a higher-level reason why for Jane's failure but still not a reason why.

If there really are ontic explanatory relevance relations, then at most one of Lange or Skow can be right. An omniscient agent could tell us whether we are in a Lange-world, a Skow-world, or neither. Nonetheless, a scientist following Lange's account would engage in the same activities as a scientist following Skow's: they would make the same predictions and have the same level of control over dividing strawberries. Either ontological picture suffices for vindicating the use of mathematical facts in giving explanations of physical phenomena. If vindicating explanatory reasoning

---

[15] As Lange puts it, "the distinctively mathematical explanation does not exploit what the world's *causal structure* is like as a matter of mathematical necessity. Rather, it exploits what *the world* is like as a matter of mathematical necessity," namely that three does not divide 23 evenly (2017, p. 20). The actual causal mechanisms are explanatorily irrelevant when there is a distinctively mathematical explanation.

requires discerning ontic relevance relations, then settling the debate between Lange and Skow should matter for making sense of explanatory reasoning. The fact that it does not matter counts as evidence against the need for descriptivism.

As a second example, what could matter more than figuring out whether laws of nature are explanatorily relevant? Asked in a scientific context, this question seems of vital importance. *Prima facie*, to determine whether climate change ever explains severe local weather, we seemingly need to know whether laws of nature are explanatorily relevant. Soberingly, when we get down to philosophical brass tacks, foundational disagreements on this issue again make no difference to how we should act and plan.

Borrowing a scenario from Skow, imagine dropping a rock from a height of one meter. Before hitting the ground, the rock's final speed reaches 4.4 meters per second. Why does it reach this speed? According to Skow, one reason why is simply the fact that the rock was dropped from one meter, since this fact is a cause of the rock's speed reaching 4.4 m/s. On Hempel's account, we need to cite not only this initial condition but also a law of nature. In a Newtonian world, we would cite the law relating the rock's speed to the distance it falls. However, Skow denies that this law is a reason why the rock reaches 4.4 m/s. Similar to his debate with Lange, Skow claims that this law instead provides a reason why ⟨*falling one meter*⟩ is a reason why the rock reaches 4.4 m/s (2016, p. 75). Whereas Hempel views laws as explanatorily relevant, Skow denies that laws are ever reasons why for physical events.

As before, it is unclear how resolving this debate could make any difference to the predictions we make about falling objects, or our strategies for controlling them. What difference would it make whether an implicated law of nature is a genuine reason why (as Hempel's account suggests) rather than at best a higher-order reason why (as Skow claims)? In either a Skow-world or a Hempel-world, scientists could justifiably use laws of nature in their explanatory reasoning and use them *in the same way*. It is this way of using laws that we are interested in vindicating, but this issue floats free from Skow's debate with Hempel.

I take these two examples to further demotivate the presumed need for descriptivism about explanatory relevance relations. Insofar as our primary goal is to vindicate scientific explanatory reasoning, the ontological commitments of descriptivism take us no further. If a non-descriptivist approach succeeds at vindicating explanatory reasoning, it will have the advantage of requiring fewer ontological (or metaphysical) commitments. For it will have succeeded without presupposing ontological commitment to explanatory relevance relations.[16] It therefore seems worthwhile to seriously consider a non-descriptivist approach to explanatory relevance.

---

[16] Recall that I intend a broad notion of 'ontology' that includes Sider's notion of metaphysical structure, which he classifies as part of a theory's ideology. As Sider notes, parsimony considerations also apply to metaphysical structure (2011, p. 14).

3 A challenge for non-descriptivism

Unadorned, non-descriptivism about explanatory relevance relations faces an immediate problem. If we don't posit ontic relevance relations, how can we reject outlandish-sounding explanatory claims? We would seemingly have just as much reason to believe that the cheers of online fans explain a homerun as we do the hitter's successful swing. If we eschew appeals to ontic relevance relations, how can we reject the former explanatory claim while endorsing the latter? Scientists' explanatory reasoning is clearly not a matter of "anything goes," so the non-descriptivist has to say more. We must combine non-descriptivism with additional criteria to distinguish legitimate explanatory claims from illegitimate ones.

Indeed, van Fraassen's pragmatic account of explanation runs into precisely this problem, which some have thought fatal to explanatory irrealism. According to van Fraassen, context determines what counts as explanatorily relevant (1980, pp. 141ff.).[17] His account places no ontological constraints on which explanatory relevance relations are appropriate, denying that we must appeal to "inextricably modal or counterfactual" constraints (1980, p. 143).[18] Kitcher and Salmon (1987) show that by not placing constraints on explanatory relevance relations, van Fraassen's account lacks the resources to rule out an "anything goes" attitude toward explanation. They show that within van Fraassen's framework, *any* true proposition can be made into a maximally good explanation of *any explanandum*. All one needs to do is select an appropriately bizarre explanatory relevance relation. For instance, there is a context in which a triple conjunction of the Sun, Mars, and Mercury explains the assassination of JFK. Kitcher and Salmon conclude that unless van Fraassen "imposes some conditions on relevance relations, his theory is committed to the result that almost anything can explain almost anything" (1987, p. 322).

As Kitcher and Salmon note, a defender of van Fraassen's account might try to rule out these bizarre relevance relations by appealing to ordinary scientific practice. They quote van Fraassen's claim that "to ask that...explanations be scientific is only to ask that they rely on scientific theories and experimentation, not on old wives' tales" (1980, p. 129). Likewise, van Fraassen says that "no factor is explanatorily relevant unless it is scientifically relevant; and among the scientifically relevant factors, context determines explanatorily relevant ones" (1980, p. 126). Thus, van Fraassen takes for granted that certain relevance relations are out of bounds, simply because science says they are out of bounds. Yet, there is something unsatisfying about this response. What should count as legitimate explanatory practice is precisely what's in question here. In other words, on what grounds can we (or scientists) exclude pseudoscientific relevance relations? Kitcher and Salmon conclude that "if van Fraassen is serious in his idea that genuine explanations must not make

---

[17] Although traditionally called a 'pragmatic account' of explanation, van Fraassen's account is also 'contextualist.' His framework adjoins the context of asking and answering questions to the traditional focus on non-contextual world–theory relations.

[18] Elsewhere, I argue that invoking such constraints does not necessarily threaten irrealism about explanation, provided that we interpret them in a deflationary, non-descriptivist fashion. An expressivist about modality and counterfactuals could construe these constraints as being normative as opposed to descriptive or ontological. See Thomasson (2020) for a normativist approach to metaphysical modality.

appeal to 'old wives' tales,' then he ought to be equally serious about showing that relevance is not completely determined by subjective factors" (1987, p. 324).

Kitcher and Salmon argue that any way of meeting this challenge will undermine a key tenet of van Fraassen's constructive empiricism, namely that explanation is a pragmatic virtue of theories, rather than an objective, non-pragmatic virtue. According to them, placing objective constraints on explanatory relevance presupposes that explanatory power is an objective virtue of a scientific theory (1987, pp. 329–330). They take this to provide an "entering wedge" for scientific realism. What matters here is that most forms of scientific realism entail descriptivist realism about explanatory relevance relations. The challenge for explanatory non-descriptivism is to place constraints on explanatory relevance without illicitly appealing to ontic relevance relations.

Fortunately, there are now well-known non-descriptivist strategies for meeting this challenge. In metaethics, expressivism provides a non-descriptivist approach to morality that nonetheless aims to vindicate the objectivity of many moral claims.[19] By adapting expressivism to explanatory relevance relations, I will recover a sense in which relevance relations are objective and mind-independent, without positing ontic relevance relations. When combined with expressivism, explanatory non-descriptivism successfully avoids both descriptivist realism and "anything goes" relativism. The expressivist can thereby agree with Kitcher and Salmon that we ought to endorse objective constraints on explanatory relevance, where the relevant notion of objectivity does not require explanatory realism. Instead, the expressivist account I develop complements van Fraassen's constructive empiricism. I will argue that we can recover objective constraints on explanatory relevance by focusing on the instrumental value of explanatory reasoning for prediction, control, and empirical adequacy. We can meet the challenge for non-descriptivism without viewing explanation as an end in itself.[20]

4 Expressivism about explanation

Traditionally, expressivists distinguish between descriptive and non-descriptive claims or thoughts. Descriptive claims are what Field (2009) describes as *straightforwardly factual*. They represent states of affairs, thereby mirroring reality (be it physical, platonic, divine, etc.). In contrast, non-descriptive claims perform functional roles that do not represent reality. Paradigmatically, these include evaluative and normative roles, such as expressing action-directed states of mind (Chrisman

---

[19] Like me, Franklin-Hall (2014) considers how resources from metaethics might clarify debates about scientific explanation, but she does not consider expressivism as a strategy.

[20] Brandom's inferentialism (2001) provides another prominent non-descriptivist approach. I lack the space here to compare these various frameworks.

2007, p. 236). Expressivism focuses on the attitudes or commitments that non-descriptive claims express or voice.[21]

Expressivism admits many formulations, whose comparative virtues matter mainly downstream from the issues I consider here. Due to its consilience with the dominant framework for theory interpretation in philosophy of science, I begin with Gibbard's (1990) norm-expressivism. The standard account of theory interpretation—endorsed by scientific realists and anti-realists alike—proceeds through possible worlds semantics. One specifies the way the world would have to be for a theory to be true. Conveniently, Gibbard develops his framework as an extension of possible worlds semantics, adjoining a set of norms to each possible world (where the worlds are specified through descriptive claims). This framework allows for expressivism about explanatory relevance relations alongside descriptivism about causation or laws of nature.[22]

We can illustrate Gibbard's framework in the context of judgments about rationality. When a norm-expressivist about rationality says that X is *rational*, they neither ascribe a property to X nor assert a truth-condition for X (at least not directly). Instead, saying that "X is rational" is equivalent to expressing acceptance of a system of norms that, on balance, permit X (Gibbard, 1990, p. 84). In short, to think something rational is to accept a set of norms that permits it. For any given belief, judgment, or action, a complete system of norms renders it either required, permissible, or forbidden. Formally, we specify descriptive–normative worlds, given by a pair $\langle w, n \rangle$. The possible world w specifies the states of affairs (characterized by descriptive claims). The additional component n specifies a normative system:

> A system of norms, recall, is the end result of the ways the various general normative principles a person accepts combine, weigh against each other, and override one another. If it is complete, then for every conceivable fully described occasion governed by norms, the system classifies each alternative as required, optional, or forbidden....Together, w and n entail a normative judgment for every occasion. (Gibbard, 1990, p. 95)

Applying Gibbard's framework to explanation, the descriptive component w captures whether an explanans actually obtains.[23] Explanantia typically include ontic structures such as laws or causes, along with initial and boundary conditions. For instance, if phlogiston does not exist, then we cannot appeal to phlogiston in giving

---

[21] See Kraut (1990, p. 159) for a detailed characterization of this "bifurcation" between descriptive and non-descriptive claims, although minimalism about truth eliminates some of the distinctions therein. Price (2013) has criticized the bifurcation thesis, arguing that expressivists must be global expressivists. However, I interpret Price as proposing a more nuanced bifurcation thesis. For my purposes here, I simply assume that there is some way to save the spirit of this distinction. See Blackburn (2017) for discussion.

[22] Other expressivist frameworks include Gibbard's later plan-expressivism (2003) (which recasts the attitude of norm-acceptance in terms of planning attitudes), Blackburn's quasi-realism Blackburn (1984, 1993, 1998), Schroeder's (2008) account in terms of a pro-attitude of 'being-for,' and Horgan and Timmons' (2006) in terms of 'ought-commitment.'

[23] For those who think that explanation is essentially contrastive, the descriptive component can also include which contrast class members obtain.

explanations of chemical reactions. Such appeals would not constitute bad or sub-optimal explanations: they would be non-explanations.

The normative component *n* settles what it takes for an answer to be a reason why for the explanandum. To do this, *n* specifies permissible explanatory relevance relations *R* and what it takes to ideally satisfy them.[24] This involves at least two kinds of explanatory norms: (i) norms on which relevance relations are appropriate to use in answering scientific questions and (ii) norms governing what it takes to satisfy a given *R* optimally. These latter norms constitute the evaluative component of explanation, i.e. the evaluation of successful explanations as good or bad. In virtue of satisfying a permissible relevance relation *R*, a bad explanation is still minimally successful; it simply violates additional norms on good explanations. This second set of norms captures the idea that successful explanations can be better or worse. Suppressing these evaluative norms results in a simpler framework that treats explanations as binary rather than gradated, developed in Sect. 4.2.

In applying expressivism to explanation, there is a choice about how much to treat non-descriptively. Causation, laws of nature, and other ontic structures feature prominently in many scientific explanations. Coupled with a causal account of explanation, expressivism about causation would lead to expressivism about explanation (and similarly for other putatively ontic structures). Here, I give an expressivist treatment of what seems to be the bare minimum required: explanatory relevance relations. One can be a realist about causation or laws of nature while being an expressivist about explanatory relevance. This is because—as Sect. 2 documented—identifying an event's causes or laws does not necessarily settle when those causes or laws are explanatory. Claims of the latter form require a further commitment to explanatory relevance relations. Of course, arguments similar to those given in Sect. 2 could motivate non-descriptivism about causation and laws of nature. For many scientific irrealists, this combination of views will be attractive. However, defending non-descriptivism about causes and laws lies outside the scope of this paper.[25]

4.1 Expressivism about explanatory relevance

Gibbard's framework yields a simple expressivist treatment of explanatory relevance:

Relevance Expressivism: To judge that a relation *R* is an *explanatory relevance relation* for question-set $\mathscr{Q}$ is to express acceptance of a system of norms that permit using *R* to answer questions from $\mathscr{Q}$.

---

[24] Notice how by focusing on norms, this framework naturally allows for explanatory pluralism or monism. In contrast, realist descriptivism has difficulties accommodating pluralism, since different ontic relevance relations are most naturally viewed as competing rivals.

[25] For an expressivist analysis of causation see Blackburn (1984, pp. 210–212, 1993, pp. 55–57, 103ff., 179–180) or Beebee (2007); for laws of nature, see Ward (2002), who bases his account on the explanatory role of laws. Elsewhere, I develop an expressivist account of laws based on expressivism about counterfactual reasoning.

Relevance Expressivism does not require endorsing norm-expressivism about rationality. Nonetheless, expressivism about rationality does provide a convenient shorthand: a relation R is an explanatory relevance relation for $\mathscr{Q}$ provided it is *rational* to answer $\mathscr{Q}$-questions using relation R. To say that a relevance relation is *not explanatory* would then be to say that it is *irrational* to use this relation to answer a particular kind of why-question. This amounts to expressing one's rejection of any set of norms on explanation that permit this relevance relation. Relevance relations that are not explanatory would be seen as rationally forbidden. But again, we can easily remain neutral on expressivism about rationality by simply eliminating the shorthand it provides: the judgment that a relevance relation R is *not explanatory* expresses rejection of any set of norms on explanation that permit using R to answer $\mathscr{Q}$-questions.

Alternatively, we can formulate Relevance Expressivism using Schroeder's (2008) framework, based on a pro-attitude of *being-for*. To see how this works in the moral case, consider the judgment that murder is wrong. This expresses an attitude of *being for blaming for murder* (or at least, being for disapproving of it). Likewise, the judgment that murder is not wrong expresses an attitude of *being for not blaming for murder*. To formulate Relevance Expressivism within this framework, we replace the action of 'blaming' with the action of 'using to answer.' Then, the judgment that relation R is explanatorily relevant expresses an attitude of *being for using R to answer the given question*. Likewise, the judgment that relation R is not explanatorily relevant (i.e. is explanatorily irrelevant) expresses an attitude of *being for not using R to answer that question*.

Of course, we might wonder why someone favors using R to answer a particular class of questions. Here, Gibbard's appeal to norms is illuminating. We can understand attitudes such as *being for blaming for murder* as elliptical for a more complex attitude such as *being for accepting a set of norms that permit blaming for murder*. Likewise, we can understand *being for using explanatory relevance relation R* as elliptical for *being for accepting a set of norms that permit using R*. At least on a first pass then, we are not forced to choose between Gibbard's and Schroeder's frameworks.[26] Where it seems helpful, I will provide analyses in terms of both.

As a matter of human psychology, Gibbard's account does not require that people conceive of their judgments through a complete set of norms (1990, p. 93). Most of us are only dimly aware of the norms that we follow, and these norms are almost always incomplete. The expressivist about explanation says the same about scientists: scientists need not be explicitly aware of the norms on explanation that they follow. When we interpret an explanatory judgment, we implicitly supply a set of norms. Philosophical accounts of explanation attempt to make these sets of norms complete. They attempt to provide a full specification of when a given relevance relation is explanatorily required, permitted, or forbidden. To do so is to specify the norms of explanatory relevance.

Still, one might object that scientists do not conceive of their explanatory judgments as involving norms of explanatory relevance at all. Ask a scientist what it

---

[26] Arguably, Schroeder's framework provides a more promising strategy for resolving technical difficulties arising from the Frege–Geach embedding problem.

means to explain a phenomenon, and they will not say that it amounts to expressing acceptance of a normative system that permits a certain answer. If they are comfortable entertaining 'explanatory relevance relations,' they will not say that these amount to *being for using a relation R to answer a why-question*. Instead, the objection continues, scientists presumably take their explanatory judgments to express descriptive claims about states of affairs. Hence, expressivism suggests that scientists are wrong about explanation, unless they understand explanatory judgments as non-descriptive claims. According to this objection then, expressivism posits a widespread error theory about the semantic beliefs of ordinary language-users—in this case, scientsts. Chrisman (2007, p. 239) refers to this worry as a posit of semantic blindness. Problematically, semantic blindness is at least in tension with the aim of vindicating scientists' ordinary discourse about explanation.

Fortunately, Chrisman provides a convincing rebuttal to this worry, at least in the ethical context. He notes that most "ordinary speakers do not have opinions about the expressive force of their claims that are fine-grained enough to cut either for or against [ethical] expressivism" (2007, p. 240). I believe that a similar point applies in the context of scientists' judgments about explanatory relevance: most of them do not have detailed-enough opinions regarding the semantic nature of explanatory relevance. I do not think that scientists' understanding of explanatory relevance is wrong. Rather, they have not thought about explanatory relevance at the level of philosophical detail where expressivism even comes into view. Consequently, expressivism posits semantic blindness primarily for a subset of philosophers who have thought about such matters in sufficient detail. As Chrisman notes, this includes especially those "philosophers who are caught up in the dogma of descriptivism in philosophical semantics" (2007, p. 240). But of course, any philosophical view attributes error to its philosophical opponents. So this is no cost for expressivism about explanatory relevance.

4.2 An account of binary explanation

Expressivism about explanatory relevance entails expressivism about explanation. When we say that an answer to a why-question is explanatory, we implicitly reference a relation *R* that settles which facts are explanatorily relevant. Hence, giving an expressivist treatment of explanatory relevance yields an expressivist treatment of answers to why-questions. In the simple case of binary explanations, we can put this as follows:

> **Explanatory expressivism**: to judge that an answer *B* to a why-question *Q* is *explanatory* (i.e. a reason why) is to express one's acceptance of a system of norms *n* that permits this answer. Alternatively, it amounts to *being for answering Q with B*

Part of the criteria is descriptive: if we require successful explanations to be veridical, the explanans must actually obtain, which is settled by the way the world is. Another part of the criteria is non-descriptive, namely determining whether or not

the answer *B* is explanatorily relevant. This requires a judgment that some relation *R* is explanatorily relevant, which expressivism understands as expressing acceptance of a system of norms.

When we reject an answer as explanatory, we reject any system of norms that permits this answer. One way for a purported explanation to fail is to rely on a relation *R* that is not explanatorily relevant, i.e. that is explanatorily irrelevant. Consequently, the preceding expressivist account of explanatory irrelevance also leads to an expressivist analysis of failed explanations, i.e. non-explanations:

> **Explanatory Failure**: To say that an answer is *not explanatory* is to express one's rejection of any set of norms on explanation that permits this answer. Alternatively, it amounts to *being for not answering this question with that answer.*

Those willing to endorse expressivism about rationality enjoy the following shorthand: an answer is *not explanatory* provided that it is *irrational* to give this answer in response to the why-question *Q*. Likewise, an answer is explanatory provided that it is rational to give this answer in response to *Q*.

4.3 Gradating explanation

So far, my account applies only to a binary notion of explanation. This is because permissibility is binary: an answer is either permitted or not permitted. Similarly, one can either answer a question with claim *B* or not answer it with *B*. Consequently, Sect. 4.2 treats every answer as being either explanatory or not explanatory. Ideally, we would like to accommodate common ways that a successful explanation can still be better or worse than another, thereby allowing for bad explanations. Among legitimate answers to a why-question, some are more explanatory than others.

To accommodate this gradated notion of explanation, we need a suitably gradated attitude. Clumsily, one could try gradating 'permissibility' by introducing a notion of 'permissible to degree X.' However, it is better to select an attitude that is both naturally gradated and that has natural language connections to explanation. The attitude of satisfaction seems particularly fitting. When an answer is explanatory, one ought to be satisfied by it. Likewise, when searching for an explanation, we are dissatisfied with our current knowledge in some regard. We are puzzled by something. An explanatory answer would provide relief from this puzzlement. It would provide a certain kind of intellectual satisfaction.[27]

These considerations suggest the following account of gradated explanation:

> **Explanatory expressivism**$_{gradated}$: to judge that an answer *B* is explanatory to degree X is to express acceptance of a system of norms that permit being sat-

---

[27] Published remarks linking explanation with satisfaction could be provided *ad nauseam*. Here is one from Potochnik (said in the context of her communicative conception of explanation): "an account of explanation should indicate what representation counts as a satisfactory answer to any given request for explanation" (2017, p. 123).

isfied with this answer to degree X. It amounts to *being for being satisfied by this answer to that degree*.

Invoking expressivism about rationality yields a succinct shorthand: an answer *B* is explanatory to the extent that it is rational to be satisfied with this answer. An answer to a why-question falls short insofar as it is rational to be dissatisfied with this answer. By suppressing degrees of satisfaction, we also gain another binary account of explanation: to judge that an answer is explanatory is to express an attitude of *being for being satisfied by that answer*.

Alongside a quantitative analysis in terms of degrees of satisfaction, we can also characterize qualitative judgments of comparative explanatoriness. For instance, to judge that an answer is very explanatory expresses a mental state of being for being very satisfied by this answer. Likewise, to judge that an answer is somewhat explanatory expresses a mental state in favor of being somewhat satisfied by it. Judging an answer to be maximally explanatory would correspond to being in favor of being maximally satisfied.

As mentioned earlier, norms of explanatory relevance include both (i) norms governing which relevance relations are appropriate and (ii) norms governing what it takes to optimally satisfy a given relation *R*. These latter norms underwrite differences in explanatory quality. A bad explanation satisfies the appropriate relevance relation but does so relatively poorly. As Potochnik notes, "meeting the requirement for an adequate explanation does not guarantee a good explanation" (2017, p. 158). An answer that violates the first set of norms on explanatory relevance ought not satisfy us: its information is irrelevant. Yet among relevant answers, there is room for different degrees of satisfaction.

Undoubtedly, this notion of intellectual satisfaction is connected with the oft-discussed notion of a sense of understanding. Some readers might therefore worry that my expressivist analysis faces similar problems to those that afflict accounts of explanation based on a sense or feeling of understanding. As Trout (2002, 2007) has argued, this sense of understanding is often an unreliable guide to legitimate explanations. However, expressivism denies that the relevant (higher-order) attitudes or sentiments function as sufficient conditions for the concept being analyzed. The fact that people blame each other for murder is not what makes murder wrong. Likewise, whether or not an answer is explanatory is independent of whether or not anyone is actually disposed to find it satisfying.

By appealing to norms, expressivism avoids a bare dispositional account of intellectual satisfaction. Norms enable us to talk about the conditions under which we ought to be satisfied, even if we in fact are not. According to explanatory expressivism, whether or not an answer is explanatory is *not* a matter of whether agents are disposed to feel satisfied with that answer. Students, for instance, may be disposed to remain dissatisfied with answers that they ought to find completely satisfactory, in virtue of these answers being genuine explanations. Instead, what matters is not the attitudes we are disposed to have, but the attitudes that are apt or fitting. The aptness or fittingness of these attitudes is again settled by the relevant norms. The case of humor provides a helpful analogy: it can be fitting to find a joke funny, even if most people are not disposed to laugh at that joke. Likewise, some jokes are not funny,

even if people laugh at them.[28] Section 6 further clarifies how expressivism preserves the objectivity and mind-independence of explanations, which seem central to scientific practice.

5 Vindication through selective minimalism

Like realism, expressivism shares the aim of vindicating a given piece of discourse, be it moral or scientific. In ordinary moral discourse, we take it to be true that mistreating animals is morally wrong. Despite viewing moral claims like this non-descriptively, expressivism aims to vindicate them. In its older incarnations, expressivism was a kind of non-cognitivism, denying truth-values to non-descriptive claims. Yet, it is difficult to see how non-cognitivism can vindicate claims that we standardly take to be true. Largely for this reason, expressivists have—for more than two decades now—embraced minimalism about truth.[29] As a deflationary theory of truth, truth-minimalism relies on a disquotation principle such as "p" is true if and only if p.[30]

As we have seen, expressivism provides a non-metaphysically committal way of understanding claims such as "murder is wrong." For instance, we can understand claims of moral wrongness as expressing an attitude of being for disapproving of some action. Understanding these claims does not require assuming that they track or represent moral facts or properties in reality. We thereby vindicate assertions of the form "murder is wrong." Then, by applying truth-minimalism, we likewise vindicate assertions of the form "it is true that murder is wrong." In this way, moral expressivism recovers a notion of moral truth while sidestepping the ontological commitments characteristic of moral realism.

Some expressivists go further and explicitly deny that there are any non-deflationary truth-makers for moral claims, but in principle expressivists can remain agnostic on this metaphysical question. What matters is that expressivism does not appeal to non-deflationary truth-makers to ground the truth of non-descriptive claims.[31] Additionally, it is not necessary for moral expressivists to endorse minimalism across the board, for any and all subject matters. It is coherent to hold that certain classes of claims are true or false at least in a deflationary sense, while others can be true or false in a more metaphysically robust, non-deflationary sense.

To recover a sense in which scientific explanatory claims are true or false, we can couple selective minimalism about explanatory relevance with Sect. 4's expressivist analysis. When we say that "Relation R is explanatorily relevant for a class of questions $\mathcal{Q}$," the disquotation schema permits us to say "it is true that R is explanatorily

---

[28] On combining expressivism with a fitting attitude theory, see D'Arms and Jacobson (2006, p. 198).

[29] See Blackburn (1996) for a clear statement that expressivists should reject non-cognitivism.

[30] For this use of truth-minimalism, see Gibbard (2003, p. 18), Horgan and Timmons (2006b, p. 88), or Field (2009, p. 267). Horwich (1998) responds to a battery of concerns about minimalism.

[31] Sinnott-Armstrong (1993, p. 298) argues that it is difficult for expressivist accounts to rule out corresponding descriptivist accounts. Indeed, ruling these out requires making epistemically-risky claims; exclusionary metaphysics is still metaphysics!

relevant." Similarly, when we say that "*B* explains *P*" or "*B* is a reason why *P*," the disquotation schema permits us to say "it is true that *B* explains *P*." In this way, expressivism can vindicate ordinary scientific claims such as "greenhouse gas emissions explain global warming," and also "socks do not explain why laptops work." Of course, these are not the only kinds of explanatory claims internal to ordinary scientific discourse. Scientists also typically believe that correct explanations do not depend on us. Section 6 describes how expressivism can even vindicate such realist-sounding claims concerning the objectivity and mind-independence of scientific explanation.

One might worry that vindicating such realist-sounding claims is a step too far: what is left to distinguish expressivism from descriptivist realism if they both vindicate claims like "scientific explanations are objective and mind-independent"? This is an instance of the much-discussed *problem of creeping minimalism*: how are we to distinguish realist positions from expressivist irrealisms that recover and vindicate ordinary discourse that is realist-sounding in its surface grammar (Dreier, 2004)? Many solutions to this problem appeal to explanatory differences between expressivism and realism. However, since the nature of explanation is exactly what is in question here, I do not believe these solutions are the best way to proceed. Instead, I believe it suffices to recall the aim of internal adequacy, while distinguishing this aim from other philosophical aims that a realist might have.

For a philosophical theory to be internally adequate for a given discourse, it need only vindicate the claims that are *internal* to that discourse. Internal claims take place within what Horgan and Timmons call "engaged contexts" (2015, pp. 207–208). Engaged contexts occur when practitioners are using a given discourse, rather than talking about it from a perspective outside that discourse. Gibbard contrasts internal claims with *external claims* (2003, p. 186). These supply commentary on a certain kind of discourse or mode of thinking, using concepts from outside that discourse. External claims typically take place in what Horgan and Timmons call "detached contexts," such as the proverbial "philosopher's room."[32] Typically, realists are interested not only in vindicating internal claims, but also in vindicating external, realist-sounding claims. These are claims to the effect that morality is objective and mind-independent, as uttered in a detached context. Moral expressivism does not aim to recover or vindicate these external claims made by moral realists. *Mutatis mutandis*, the same holds for expressivism about explanatory relevance. My aim is to vindicate the internal claims about explanation made by scientists, rather than certain external claims made by realist philosophers.[33] In this way,

---

[32] For a defense of (the intelligibility of) a similar distinction between contexts of ordinary vs. ontological existence assertions, see Chalmers (2009). Horgan and Timmons' engaged/detached distinction has some advantages. See also D'Arms and Jacobson (2006, p. 198).

[33] Many claims in the explanation literature are naturally construed as external claims, especially when what's at issue is competing philosophical views (rather than views coming from science). In discussing Coffa's ontic conception of explanation, Salmon says that "Explanations, in his view, are fully objective and, where explanations of nonhuman facts are concerned, they exist whether or not anyone ever discovers or describes them" (1989, p. 133). Said in a context contrasting the ontic conception with other conceptions of explanation, I read this as an external claim that such explanations exist independently of intelligent life.

Horgan and Timmons' distinction between engaged and detached contexts supplies a simple solution to the problem of creeping minimalism.

One might worry that any distinction between engaged and detached contexts is artificial or *ad hoc*. If so, then expressivists seemingly would be unable to isolate the aim of internal adequacy from more expansive philosophical aims.[34] However, this worry is empirically unfounded. We routinely distinguish between engaged and detached contexts. For instance, we do not take what goes on in the mathematics classroom as being (obviously) sufficient grounds for platonism. We naturally switch between ordinary, engaged contexts where we say "the fundamental theorem of calculus is true" and detached, philosophical contexts where we wonder whether the theorems of calculus correspond to a platonic reality. Even though I do not subscribe to mathematical realism, I still believe that it would be intellectually inappropriate to run into mathematics halls proclaiming the death of mathematics. No doubt, it is sometimes ambiguous which kind of context we are in, or we move quickly back and forth between them. The existence of such ambiguities does not undermine the intelligibility of the distinction.[35]

Rather than speak of truth in a deflationary sense, one might object that we should simply introduce an alternative predicate, such as 'apt.' According to this objection, we ought to reserve 'true' for the non-deflationary notion that features primarily in detached contexts. Whereas expressivism is silent on whether scientific explanatory claims are true in a non-deflationary sense, it at least vindicates their aptness for scientific practice.[36] Adhering to this convention might be fine for conversations confined to the philosopher's room, but it introduces problems for vindicating ordinary scientific discourse. Scientists do not merely claim that (i) "it is apt to say that climate change explains severe weather" and (ii) "it is inapt to say that socks explain the internet." Scientists say that the former explanatory claim is true whereas the latter is false. The proposal to use aptness faces an uncomfortable tension: it simultaneously denies that climate change explains severe weather events while asserting it is apt to say that climate change explains severe weather. As Field notes while defending the deflationary use of 'true,' "Not only is it sensible to so apply the notions of truth and falsity to normative claims, but disallowing such applications would defeat the main purposes that the notions of truth and falsity serve" (2018, p. 16). Hence, I believe that with the help of truth-minimalism, expressivism provides a more promising strategy for vindicating scientific explanatory discourse.

---

[34] Szabó criticizes a superficially similar distinction between ordinary vs. philosophical contexts, proposing an alternative explanation of why correctness conditions seem to shift between these contexts (2010, p. 34). However, the more general engaged vs. detached distinction is compatible with Szabó's alternative explanation and hence avoids his criticism.

[35] Indeed, there might be other kinds of contexts as well, with notions of truth that fall between the deflationary and the non-deflationary. I am inclined to think that constructive proofs in mathematics supply an intermediary context, one in which we have something more than deflationary truth but still short of platonic truth. This idea seems connected with Field's suggestion that the notion of 'objectivity' is graded and contextual (2018, p. 17).

[36] I thank Laura Ruetsche for pressing this objection.

6 Objectivity and mind-independence

We have seen how expressivism can vindicate simple explanatory claims such as "lightning explains thunder." In this section, I describe how expressivism can also vindicate more complex explanatory claims concerning the objectivity and mind-independence of explanatory relevance, uttered within engaged contexts. Section 6.1 discusses how higher-order norms on explanatory relevance allow us to understand the objectivity of scientific explanation. Section 6.2 discusses how expressivism can accommodate the possibilities of both mistakes and substantive disagreements about explanatory relevance. To do so, I introduce a notion of improvement. Rather than by more accurately representing relevance relations in reality, explanatory norms improve by approaching a more optimal functional role relative to the aims of science. Finally, Sect. 6.3 considers the implications of a particular first-order account of improvement. Overall, this section further demonstrates how expressivism avoids the challenge of Sect. 3: it does not collapse into a problematic form of relativism.

6.1 Higher-order norms

In the moral domain, expressivists point out that we endorse higher-order norms that prevent first-order norms from changing based on our inclinations or attitudes.[37] According to these higher-order norms, the wrongness of tripping people is unaffected by people's attitudes toward tripping people. Even in a world where most people think it is morally right to indiscriminately trip others, it would not be morally right to do so. And the fact that most people currently think that tripping people is morally wrong is not a truth-maker for its moral wrongness either.

Through these higher-order norms, expressivism recovers a sense in which moral facts are mind-independent and objective.[38] For instance, Gibbard argues that expressivists can vindicate realist-sounding claims such as "it's a normative fact, out there independent of us, that one ought not to kick dogs for fun" (2003, p. 186). On his account, "accepting this might amount to planning to avoid kicking dogs for fun, planning this even for the contingency of being someone who approves of such fun, and who is surrounded by people who approve" (2003, p. 186). Likewise, Blackburn notes that "If everyone comes to think of it as permissible to maltreat animals, this does nothing at all to make it permissible; it just means that everybody has deteriorated" (1985, p. 14). To reject these higher-order norms or higher-order moral attitudes is to commit a moral wrong or to display a reprehensible moral attitude.

Similarly, as part of our usual system(s) of explanatory norms, we endorse higher-order norms that it is not anything goes. These include norms like the following: "matters of explanatory relevance are not settled by scientific opinion," and "what counts as explanatorily relevant is independent of what people think counts as relevant." Hence, even if we were to endorse different views about relevance, this

---

[37] See Blackburn (1984, pp. 217ff.) and Gibbard (1990, p. 165, 2012, p. 233, 2015, p. 172). Blackburn (1993, pp. 127–129) presents an earlier version of this argument (originally from 1973).

[38] Field (2018, p. 16) refers to this as a kind of counterfactual objectivity. For criticism see Street (2011); Gibbard (2011) provides a response.

would not affect what *ought* to count as explanatorily relevant. These higher-order norms build in a form of counterfactual stability, making the correct explanatory judgments robust under perturbation of different attitudes about explanation.

As in the ethical domain, higher-order norms make explanation mind-independent in a way that matters of taste are not. When we assert a moral or explanatory claim, we typically want our audience to agree with us. We want them to endorse the moral or intellectual attitudes that we endorse, based on our background set of norms. In contrast, when we assert a mere preference or taste, we often neither desire nor imply that our audience ought to agree with us. We can view our audience as having poor or even tragic tastes, while recognizing that this difference is a matter of taste all the same. It is not the sort of disagreement over which we come to moral or intellectual blows. Regarding matters of taste, we do not endorse corresponding higher-order norms. If one rejected such higher-order norms about explanation, they might come to view judgments of explanatory relevance as matters of taste. The preceding discussion shows that expressivists do not have to view all explanatory disputes this way, even after rejecting descriptivism about explanation. Although an expressivist could be a relativist, they need not be. Embracing relativism involves taking a particular first-order stance on explanatory norms.[39]

6.2 Instability, error, and disagreement

Although higher-order norms recover a sense in which our first-order norms can be objective and mind-independent, at least two substantive worries remain. First, why should we endorse one system of explanatory norms over another? If two systems of norms rule each other out, while neither purports to represent ontic relevance relations, why privilege one over another? I'll call this concern the *instability problem* for expressivism.[40] Second, how can expressivism accommodate the possibility of error about explanatory claims? Scientists often speak as though they could be mistaken about the right explanation, including whether or not particular information is explanatorily relevant. A scientist might say "I believe this explains it, but I could be wrong." Since expressivists seek to preserve ordinary scientific discourse about explanation (rather than undermine it), they must accommodate these ordinary assertions of explanatory fallibilism. Call this the *fallibility problem*.[41]

Both instability and fallibility are closely connected with the possibility of disagreement, either with our temporally-extended selves or others. Two scientists can disagree about what information is explanatorily relevant, while believing that at most one of them can be correct. In such contexts, realist descriptivists have a simple story to tell: at most one of the scientists is tracking reality. Expressivism faces the challenge of giving a non-descriptive construal of explanatory disagreements.

---

[39] Note that expressivism allows interpreting *some* debates over explanatory relevance as being matters of taste or regimentation, such as the characteristically philosophical debates discussed in Sect. 2.2.

[40] See Blackburn (1984, p. 197) for these worries, sometimes called the "schizoid attitude problem" (although this terminology seems infelicitous).

[41] As Gibbard notes in the moral context, "we have to make sense of the possibility that we might feel approval for an action that isn't good" (2015, p. 184).

What is it to be right or wrong about an explanatory judgment if there is no state of affairs that settles the dispute? Gibbard (2015) views the possibility of disagreement as the key to an expressivist account of how we might be wrong about moral claims. These disagreements involve 'objectivizing' moral properties, i.e. speaking as if there are substantive moral properties that we can disagree about.

To solve both the instability and fallibility problems, it suffices to provide an account of how our norms improve. Let's consider fallibility first. Judgments of fallibility arise from the epistemic possibility that our norms could be improved. Horgan and Timmons (2015) use this observation to provide a detailed expressivist treatment of judgments of the possibility of moral error.[42] Simplifying their account, we can analyze assertions of the form "B is explanatory (or explanatorily relevant), but I might be wrong" as expressing the following attitude: it is epistemically possible that on an improved system of norms, B is not explanatory (or not explanatorily relevant).[43] Likewise, we can analyze assertions of the form "I was wrong to say that B is explanatory" as expressing the following attitude: I have since arrived at an improved set of norms on which B is not explanatory.

The improvement of norms also addresses the instability problem. A set of norms is *stable* provided that there are no nearby or obvious improvements. When we do consider alternatives, we should switch only if we view an alternative as an improvement. Provided that we have reason to believe our current explanatory judgments are stable (or close to stable), we should not be overly concerned about alternative systems of norms that we could have endorsed. As with most epistemic defeaters, we may become increasingly concerned about alternatives as they are raised to salience. But that is simply to subject our current system of norms to scrutiny, and scrutiny may itself be conducive to further improvement. In general, improving a system of explanatory norms will not require jettisoning most current beliefs about explanation. By recognizing that some changes are for the better, we can therefore view our commitments as stable yet fallible.

6.3 But what *are* the norms?

So what are the norms on improvement? Or, more interestingly, what should they be? Any answer to this question takes us from a meta-theoretical account of explanatory relevance to a first-order proposal.[44] Although my primary goal is to rehabilitate an irrealist conception of explanation, it is instructive to consider a first-order proposal regarding norms on improvement.

---

[42] Horgan and Timmons relate their framework to Neurath's boat, as do Blackburn (1993, p. 79) and Field (2018, p. 3). However, I believe that an analogy with optimization surfaces is more apt. I plan to develop this approach elsewhere, based on similar models used in complex systems theory.

[43] For their part, Horgan and Timmons are inclined to regard epistemic possibilities as descriptive (2015, p. 198). However, anyone tempted to expressivism about scientific explanation will likely endorse expressivism about epistemic possibility, e.g. along the lines of Yalcin (2007).

[44] See Gibbard (2003, p. 185) for discussion of how expressivism is a metaethical framework, rather than a first-order normative theory.

First, I assume that to count something as an improvement, one must presuppose an aim. For instance, when we count an increase in accuracy as an improvement, we presuppose that belief aims at truth (or at least a subset of truths). Without this aim, it would be unclear why greater accuracy is an improvement rather than neutral or worse. Given this assumption about the concept of improvement, we can specify norms on explanatory improvement by asking what non-descriptive functional roles explanatory judgments perform (or ought to perform).

The discussion of climate change in Sect. 2 suggests at least three related functional roles of explanatory judgments. First, they guide our predictions about what will happen under various actual or hypothetical circumstances. Second, they help us exert control on physical subsystems, by guiding what we should do in order to achieve certain ends. Finally, due to their role in guiding predictions, explanatory judgments also influence whether we think a given theory or model is empirically adequate, i.e. saves the observable phenomena. Part of empirical adequacy involves saving the phenomena not just in the past and present, but also the future. If we think a model has bizarre explanatory relevance relations, we might be motivated to test for a failure of empirical adequacy.[45] Collectively, these three functional roles seem intricately connected to what Woodward (2003) calls *what if things had been different questions*. This form of counterfactual or subjunctive reasoning involves considering how an output variable would change if we were to alter an input variable.

Focusing on these three functional roles leads to an *instrumentalist* interpretation of explanation. According to *explanatory instrumentalism*, explanations do not possess final value; instead, they are instrumentally valuable for the non-explanatory aims of science. These aims are standardly taken to include empirical adequacy, prediction, and control.[46] If we construe explanation as instrumentally valuable for these aims, then changes to our explanatory norms that facilitate these aims count as improvements. For the explanatory instrumentalist, one system of norms is better than another provided that it better facilitates the non-explanatory aims of science, *ceteris paribus*. These aims thereby provide criteria for assessing whether we have arrived at the right system of explanatory norms.

In suggesting that the norms on explanatory improvement are determined by certain non-explanatory aims of science, we need not suppose that scientists think about the value of explanation instrumentally. Scientists need not be aware of what leads them to various assessments of explanatory improvement. Similarly, if our moral system is improved by choices that better facilitate cooperation, we need not be actively thinking about *cooperation as an aim of morality* when we make these choices. Psychologically, many scientists take a realistic outlook, in the same way that many people think of morality objectively. It is part and parcel of irrealist

---

[45] Although Sect. 2.2's argument against (the need for) descriptivism relies on there being no empirical differences between a few pairs of rival philosophical theories, it does not entail that different norms on relevance *never* inspire theories with empirical differences. Consider, for instance, the empirical failures of Aristotelian physics, with its reliance on teleological relevance relations.

[46] For discussion of instrumentalism about explanation, see Lombrozo (2011). Van Fraassen—a chief proponent of explanatory instrumentalism—takes explanations to be instrumentally valuable for greater empirical adequacy (1980, pp. 92ff.). Dorst (2019, p. 2672) argues that they are instrumentally valuable for predictions.

philosophies that what motivates people to act and believe in particular ways can be quite different from what the irrealist takes to be characteristic of their activity. We come to conclusions about a domain of discourse primarily based on how people act, rather than through their first-personal reports of those activities. One cannot establish platonism simply by pointing out that mathematicians routinely express belief in mathematical objects.

No doubt, an explanatory expressivist could endorse a very different account of improvement than that suggested by explanatory instrumentalism. Indeed, an expressivist can always grant that there might be better norms to be had (illustrating another difference with relativism). Proponents of different norms of improvement would then disagree about which system of explanatory norms is better, or at least *why* an agreed-upon system is better. By arguing over norms of improvement, we indirectly argue about the explanatory relevance relations themselves.

7 Conclusion

Many accounts of explanation presuppose that the world comes equipped with a particular set of explanatory relevance relations. This descriptivist assumption frames philosophical investigation about explanation as a search for these facts. However, this search distracts us from the roles that explanatory judgments perform in scientific practice. It leads us to consider questions that—despite appearing to be foundational—are actually idle. For instance, philosophical disputes about non-causal explanations may just be disputes over how to regiment our language. Settling many of these philosophical debates is not necessary to vindicate scientists' ordinary explanatory reasoning.

The drawbacks of descriptivism motivate a non-descriptivist approach to explanation. On this approach, we focus on the non-descriptive functional roles that explanatory judgments perform. Whereas many have taken non-descriptivism to collapse immediately into conventionalism or relativism, expressivism shows a promising way forward. Judgments of explanatory relevance are not a matter of anything goes, even if they do not necessarily track ontic relevance relations. Instead, explanation is a norm-governed process. Scientists' explanatory judgments express their endorsement of norms on explanatory relevance. These norms are better or worse insofar as they facilitate the aims of science. By specifying a particular account of what it takes to improve explanatory norms, an expressivist can provide a first-order account of explanation that is irrealist yet not relativist.

Here, I have considered an instrumentalist interpretation of explanation. On this view, we see how philosophical debates such as those between Lange and Skow or between Skow and Hempel might be matters of taste. Provided that either option in these debates does not affect the instrumental value of explanatory judgments, they may be merely matters of regimentation. The value of choosing one position over another would amount to the value of having a consistent convention. As with driving on the right-side of the road, such conventions are at most locally important. In contrast, scientifically-motivated disputes about explanatory relevance are typically not like this. The explanatory commitments of Aristotelian teleology are

worse for the aims of science. They inspire less effective plans for problem-solving. It is therefore not a matter of taste that teleological relevance relations are illegitimate for physics. Through a deflationary approach to truth, expressivism recovers the ordinary claim that Aristotelians were wrong about what explains a stone's falling to the earth. Science made progress by rejecting teleological relevance relations. Expressivism about explanation shows that this notion of progress does not require successfully representing unobservable explanatory relevance relations.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article.

References

Beddor, B. (2019). Relativism and expressivism. In M. Kusch (Ed.), *The Routledge handbook of philosophy of relativism* (pp. 528–539). Routledge.

Beebee, H. (2007). Hume on causation: The projectivist interpretation. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality. Russell's republic revisited* (pp. 224–49). Oxford University Press.

Blackburn, S. (1984). *Spreading the word: Grounding in the philosophy of language*. Oxford University Press.

Blackburn, S. (1985). Errors and the phenomenology of value. In T. Honderich (Ed.), *Morality and objectivity* (pp. 1–22). Routledge and Kegan Paul.

Blackburn, S. (1993). *Essays in quasi-realism*. Oxford University Press.

Blackburn, S. (1996). Blackburn Reviews Dworkin. Brown Electronic Article Review Service. J. Dreier & D. Estlund (Eds.).

Blackburn, S. (1998). *Ruling passions*. Oxford University Press.

Blackburn, S. (2005). Quasi-realism no fictionalism. In M. E. Kalderon (Ed.), *Fictionalism in metaphysics* (pp. 322–38). Oxford University Press.

Blackburn, S. (2017). Pragmatism: All or some or all and some? In C. Misak & H. Price (Eds.), *The practical turn: Pragmatism in Britain in the long twentieth century* (Vol. 210, pp. 61–74). Oxford University Press.

Brandom, R. B. (2001). *Articulating reasons: An introduction to inferentialism*. Harvard University Press.

Chalmers, D. J. (2009). Ontological anti-realism. In D. J. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics. New essays on the foundations of ontology* (pp. 77–129). Oxford University Press.

Chrisman, M. (2007). From epistemic contextualism to epistemic expressivism. *Philosophical Studies, 135*(2), 225–54.

Craver, C. F., et al. (2014). The ontic account of scientific explanation. In M. Kaiser (Ed.), *Explanation in the special sciences: The case of biology and history* (pp. 27–52). Springer.

D'Arms, J., & Jacobson, D. (2006). Sensibility theory and projectivism. In D. Copp (Ed.), *The Oxford handbook of ethical theory* (pp. 186–218). Oxford University Press.

Dorst, C. (2019). Humean laws, explanatory circularity, and the aim of scientific explanation. *Philosophical Studies, 176,* 2657–79.

Dreier, J. (2004). Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives, 18*(1), 23–44.

Field, H. (2003). Causation in a physical world. In M. J. Loux & D. W. Zimmerman (Eds.), *The Oxford handbook of metaphysics* (pp. 435–60). Oxford University Press.

Field, H. (2009). Epistemology without metaphysics. *Philosophical Studies, 143,* 249–90.

Field, H. (2018). Epistemology from an evaluativist perspective. *Philosophers' Imprint, 18*(12), 1–23.

Franklin-Hall, L. (2014). The meta-explanatory question (PSA 2014).

Frisch, M. (2022). Causal explanation in physics. In E. Knox & A. Wilson (Eds.), *The Routledge companion to philosophy of physics* (pp. 454–65). Routledge.

Gibbard, A. (1990). *Wise choices, apt feelings*. Harvard University Press.

Gibbard, A. (2003). *Thinking how to live*. Harvard University Press.

Gibbard, A. (2011). How much realism? Evolved thinkers and normative concepts. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 3, pp. 33–51). Oxford University Press.

Gibbard, A. (2012). *Meaning and normativity*. Oxford University Press.

Gibbard, A. (2015). Improving sensibilities. In R. N. Johnson & M. Smith (Eds.), *Passions and projections: themes from the philosophy of Simon Blackburn* (pp. 172–89). Oxford University Press.

Hempel, C. G. (1965). Aspects of scientific explanation. *Aspects of scientific explanation, and other essays in the philosophy of science* (pp. 331–496). Free Press.

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy, 98*(6), 273–99.

Horgan, T., & Timmons, M. (2006). Cognitivist expressivism. In T. Horgan & M. Timmons (Eds.), *Metaethics after Moore* (pp. 255–98). Oxford University Press.

Horgan, T., & Timmons, M. (2006). Expressivism, Yes! Relativism, No! In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 1, pp. 73–98). Oxford University Press.

Horgan, T., & Timmons, M. (2015). Modest quasi-realism and the problem of deep moral error. In R. N. Johnson & M. Smith (Eds.), *Passions and projections: Themes from the philosophy of Simon Blackburn* (pp. 190–209). Oxford University Press.

Horwich, P. (1998). *Truth* (2nd ed.). Oxford University Press.

Kitcher, P., & Salmon, W. C. (1987). Van Fraassen on explanation. *The Journal of Philosophy, 84*(6), 315–30.

Kraut, R. (1990). Varieties of pragmatism. *Mind, 99,* 157–83.

Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *British Journal for the Philosophy of Science, 64,* 485–511.

Lange, M. (2017). *Because Without Cause: Non-causal Explanations in Science and Mathematics*. Oxford University Press.

Lewis, D. (1986). Causal explanation. *His philosophical papers* (Vol. 2, pp. 214–40). Oxford University Press.

Lombrozo, T. (2011). The instrumental value of explanations. *Philosophy Compass, 6*(8), 539–51.

MacFarlane, J. (2014). *Assessment sensitivity: Relative truth and its applications*. Oxford University Press.

Masson-Delmotte, V. (2021). Climate change 2021: The physical science basis. Technical report. Intergovernmental Panel on Climate Change (IPCC).

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Price, H. (2013). The Descartes lectures 2008. *Expressivism, pragmatism and representationalism* (pp. 3–64). Cambridge University Press.

Railton, P. (1981). Probability, explanation, and information. *Synthese, 48,* 233–56.

Salmon, W. C. (1989). *Four decades of scientific explanation*. University of Minnesota Press.

Schroeder, M. (2008). *Being for*. Oxford University Press.

Sider, T. (2011). *Writing the book of the world*. Oxford University Press.

Sinnott-Armstrong, W. (1993). Some Problems for Gibbard's norm-expressivism. *Philosophical Studies, 69,* 297–313.

Skow, B. (2016). *Reasons why*. Oxford University Press.

Street, S. (2011). Mind-Independence without mystery: Why quasi-realists can't have it both ways. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 6, pp. 1–32). Oxford University Press.

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

Szabó, Z. G., et al. (2010). The ontological attitude. In T. Czarnecki (Ed.), *The analytical way. Proceedings of the 6th European congress of analytic philosophy* (Vol. 25, pp. 21–42). College Publications.

Thomasson, A. L. (2020). *Norms and necessity*. Oxford University Press.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science, 69*(2), 212–33.

Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass, 2*(3), 564–91.

van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.

van Fraassen, B. C. (1994). Against transcendental empiricism. In T. J. Stapleton (Ed.), *The Question of Hermeneutics. Essays in Honor of Joseph J. Kockelmans* (pp. 309–35). Springer Science.

Ward, B. (2002). Humeanism without humean supervenience: A projectivist account of laws and possibilities. *Philosophical Studies, 107*(3), 191–218.

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

Woodward, J. (2007). Causation with a human face. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality. Russell's republic revisited* (pp. 66–105). Oxford University Press.

Woodward, J. (2016). The problem of variable choice. *Synthese, 193,* 1047–72.

Wright, C. (1988). Realism, antirealism, irrealism, quasi-realism. *Midwest Studies in Philosophy, 12*(1), 25–49.

Yalcin, S. (2007). Epistemic modals. *Mind, 116*(464), 983–1026.