

Scientific explanation and disciplinary goals: the case for constrained contextualism

Andrew Wayne
University of Guelph

Abstract

This paper examines the role of context in scientific explanation. It develops an approach, constrained contextualism, wherein what makes an argument genuinely explanatory depends in part on how well it serves other scientific aims. Disciplinary communities prioritize diverse goals in addition to explanation, including prediction, intervention, representation, realism, and understanding. Communities working with the same scientific models and within the same theoretical framework may legitimately maintain different explanatory standards based on their alignment with other disciplinary goals. The paper examines and rejects three alternative approaches to contextualism: traditional accounts, which allow only a very narrow role for context in explanation (Woodward); theory-relative accounts, in which factors to relevant high-level theory are salient to explanation (Jansson, Lange and idealized model explanations); and extreme contextualism about explanation (van Fraassen). It responds to an argument from Potochnik that there is rarely if ever any such alignment between the aim of explanation and other scientific goals. Ultimately, constrained contextualism is able to defend reasonable normative standards for explanation while at the same time recognizing the considerable role contextual factors do and should play in explanatory practice.

Keywords

Scientific explanation, contextualism, scientific goals, physics, idealization

1. Introduction

Contextualism about scientific explanation has something of a bad reputation among philosophers of science. For some, *bona fide* explanations track causal or nomological relations in the world that are objectively relevant to the phenomenon to be explained. For others, explaining a phenomenon involves identifying the mechanisms responsible for it, again a matter of discovering objective features of the world. Even on a unificationist account, explanations aim to unify phenomena using argument patterns that reflect objective similarities in nature. Of course, explanatory practices vary across scientific disciplines and historical periods, and this diversity means that there may be different forms of scientific explanation appropriate for different questions in different scientific fields. However, once an explanandum and a contrast

class of alternatives are fully specified, many philosophers assert that the appropriate explanatory relation is also determined, leaving only one correct explanation that bears a unique, context-independent relation to the question. By contrast, contextualism holds that the explanatory relation may depend on further factors. On this view, what counts as a genuine explanation of a phenomenon may vary across theoretical and disciplinary contexts.

Philosophers of science have long associated contextualism with Bas van Fraassen's pragmatic account of scientific explanation (van Fraassen 1977; 1980; 1985). As van Fraassen frames it, questions consist of an explanandum, a contrast class of alternatives, and a relevance relation. For example, consider the question "Why is this aluminium conductor warped?" It consists of a fact to be explained (the conductor warped when heated); a class of alternatives (conductors made of other materials), and a relevance relation (causal history). A good scientific explanation provides information bearing the appropriate relevance relation to the fact to be explained; in this example, a causal story about how aluminium warps when heated by a given amount, while copper and steel do not. As van Fraassen famously points out, the conductor's causal history are empirical facts, but so are the positions of the sun, moon, and stars. Given a relevance relation of *astrological* significance, and a contrast class of conductors heated when the positions of heavenly bodies were in different locations, astrology also provides an explanation of the warping of the conductor. Of course, the astrological theory does not argue that the positions of heavenly bodies caused the warping, only that they have (astrological) explanatory relevance. On van Fraassen's account, there are no restrictions on the relevance relation.

This extreme contextualism about scientific explanation has unsavoury consequences. As the astrology example illustrates, without constraints on what constitutes a legitimate relevance relation, virtually any set of facts could be considered explanatory of any phenomenon. If almost any relation can count as explanatory in *some* context, the concept of explanation loses its normative force, and we can no longer distinguish good explanations from bad ones. We want an account of explanation in science that matches scientific practice, which privileges a restricted set of relevance relations, such as causal, nomological or mechanistic. Notably, Van Fraassen's account struggles with the problem of explanatory asymmetry. For instance, while the height of a flagpole explains the length of its shadow (via causal laws about light propagation), the length of the shadow does not explain the height of the flagpole. Yet on his account there is no way to debar the pathological case. The problem generalizes to non-causal cases in which explanatory asymmetry exists without causal asymmetry ([self-reference omitted]). Most reject this sort of radical contextualism about explanation. For one thing, it means that explanation is not actually an aim of science but rather a pragmatic concern

overlaid on what, for van Fraassen, is science's singular goal of empirical adequacy. For another, it is inconsistent with how scientists actually think about and use explanatory arguments.

On the traditional approach, by contrast, there is an extremely narrow role for context in explanation. As James Woodward puts it, our knowledge states and interests influence what we want to explain, determining legitimate explanatory questions within a particular scientific community at a specific time (2003, Ch. 2). However, once a question is fully specified, there exists only one valid explanation that bears a unique, context-independent relation to this question. On this approach, for example, we can ask a question about why this conductor warped rather than other similar ones by specifying a contrast class. Here, the contrast class may comprise different design choices that would have affected warping, such as using alternative materials with different thermal resistances. Importantly, once the question and contrast class are fully specified, only one correct explanation exists: a causal description of each materials' thermal conductivity and response to thermal stress. Contextual factors such as theoretical background, disciplinary perspective and practical considerations have relevance only with respect to the choice of question being asked.

The traditional view allows that there are different types of scientific explanation. For a given scientific question, the relevant explanatory factors may be causal, non-causal, nomological, mechanistic, mathematical or unificationist. These various types of explanation are fundamentally distinct with respect to the features that make them explanatory. For some, epistemic properties of the explanation are key, such as use of natural law or unification; for others, it is reference to mechanisms or causes in the world. However, each type of explanation answers its own special sort of question, and these special sorts of questions are mutually exclusive. If the warping of the aluminium conductor but not the copper one is explained by differing causal features of the metal, then no other type of explanation applies. In this case, the causal approach provides a genuine explanation of the phenomenon, while non-causal, nomological, mechanistic, mathematical and unificationist approaches do not. As Woodward emphasizes, explanations of a phenomenon can be deeper or shallower. However, these differences reflect different contrast classes, that is, different ways of specifying what is to be explained, rather than different answers to the same question. As we shall see, by ignoring the potential relevance of contextual factors to explanation, the traditional approach fails to capture some essential features of explanatory practice in science.

We have better options. We start with recent accounts from Lina Jansson, Marc Lange, and proponents of idealized model explanations, all representative of a current emphasis on the theory-relative aspects of scientific explanation. These accounts recognize that explanatory status depends not just on features of an argument itself but on its relation to a relevant high-level theory (Section 2). I then review two case studies presented by proponents of these

accounts and argue that far from supporting this approach, they undermine it. Properly understood, these cases show that what counts as a good explanation may depend on more than just their theoretical framework (Section 3). In Section 4 I develop an approach I call *constrained contextualism*, in which the explanatory value of an argument may depend on how well it serves other scientific goals within a particular disciplinary community, such as prediction, representation, intervention, realism and understanding. Different scientific communities sharing the same high-level theory may prioritize different goals, leading to different standards for what counts as explanatory. Section 5 responds to a challenge, from Angela Potochnik, that there is rarely if ever any such alignment between the aim of explanation and other scientific goals.

Constrained contextualism holds that what makes an argument genuinely explanatory depends not only on its internal structure and relation to high-level theory, but also, in at least some cases, on how well it serves other scientific goals within particular scientific communities. Constrained contextualism recognizes that communities working with the same scientific models and within the same theoretical framework may legitimately maintain different explanatory standards based on their disciplinary objectives. At the end of the day, constrained contextualism is able to defend reasonable normative standards for explanation while at the same time recognizing the considerable role contextual factors do and should play in the diversity of actual explanatory practices.

2. Theory-relative accounts of scientific explanation

In this section, we survey three recent accounts of scientific explanation from Lina Jansson, Marc Lange, and proponents of idealized model explanations. These accounts are broadly representative of current approaches to scientific explanation and are distinguished by their emphasis on what I will call the *theory-relative* aspect of explanation: the idea that whether an argument is genuinely explanatory of a target phenomenon depends not just on the argument itself—for example, whether it traces the correct causal structure—but also on its relation to the relevant high-level theory in a domain. On theory-relative approaches, explanatory standards are grounded in global theory.

Like the traditional approach, theory-relative accounts are non-contextualist in an important sense. For the traditional view, each type of explanation—causal, nomological, mechanistic, etc.—answers its own special sort of question, and these special sorts of questions are mutually exclusive. Similarly for theory-relative accounts: given a fully specified question and a relevant high-level theory, there is a uniquely best explanation. A precursor to these accounts is Kitcher's explanatory unification, where global theories are essential to

understanding how explanation functions in scientific practice (Kitcher 1981; 1985; 1989). For Kitcher, scientific explanation involves showing how a phenomenon fits into patterns of argument that unify our understanding of nature under a high-level theory such as Newtonian mechanics or Darwin's theory of evolution. High-level theory is at the centre of his account, as the set of argument patterns that underwrite explanation is defined relative to the current state of scientific knowledge. While the account of explanatory unification itself faces a number of serious challenges (Barnes 1992, [self-reference omitted]), taking an approach in which high-level theory is relevant has proven fruitful.

Jansson has developed a compelling account of explanation that highlights how higher-level laws of nature licence explanatory inferences (Jansson 2015; 2020). Jansson's starting point is the observation that laws of nature apply to particular cases only under certain conditions. For example, the warping of an aluminium conductor but not a copper one is explained by the much higher heat flux through copper compared with aluminium, given the same temperature gradient. What makes this a good explanation? Differences in thermal conductivity among metals is an empirical generalization that, on its own, is not sufficient to underwrite scientific explanation. Otherwise, we have something sounding suspiciously like Molière's explanation that opium induces sleep because it has dormitive virtue.

According to Jansson, the explanation is a good one because the empirical generalization in the explanation is licenced by a relevant higher-level law. In our example it is the Wiedemann–Franz law, which governs the thermal conductivity of metals but applies only within certain temperature ranges, pressures and metallic states. These conditions of application of the law are essential to the explanation. The explanation is successful because the Wiedemann–Franz law licences the inference from heat flux to warping. The law is silent about what explains warping outside its conditions of applicability, or indeed whether it can be explained at all; it speaks only to cases falling within those conditions. The Wiedemann–Franz law is not itself part of the explanation, but it tells us something about the high-level theory in which the explanation is embedded and about how this theory constrains the scope of the explanation.

A second recent account of explanation taking a theory-relative approach has been developed by Lange (Lange 2017, 2018, 2019, 2023). Lange's main focus is the explanations of laws, constraints and generalizations, and central to his account is the concept of *explanatory priority* (2017, Ch. 3; 2018, 34-37). Many non-causal explanations contain a symmetrical counterfactual dependence between the law or constraint being explained and laws or constraints doing explanatory work. Lange distinguishes laws and constraints in explanations that are *explanatorily fundamental* from those that are *explanatorily derivative*. This order of explanatory priority fixes the asymmetry of explanation. However, for Lange there is no general

or systematic way to distinguish, in a given explanation, which are explanatorily fundamental and which are derivative. Instead, the order of explanatory priority is determined by the relevant global theory—importantly, for Lange, the degrees of necessity of laws and constraints in the theory—and it is grounded differently in different cases (2019, 14).

When it comes to the explanation of particular phenomena, an analogous explanatory priority may be established by appealing to what Lange calls *natural properties*. Consider an explanation that contains a law L and a determinate value of a reducible (or higher-level) property P . He argues that for this to count as an explanation of a particular fact E , there must in turn be a distinct explanation of L in which P figures as a natural property, roughly a property that enters “as a unit” in the explanation of L (2017, 378, 386). For example, the center of mass of a system is defined in terms of the masses of its components and their positions. Though reducible to these component properties, valid scientific explanations can still appeal to the center of mass because it enters as a unit in some kinds of explanations. Consider the explanation of why a seesaw tips to the right. One explanation appeals to the generalization that a rigid body moves in a direction that lowers its center of mass. This law itself is explained through the principle of energy conservation, where the center of mass appears as a unified concept in the expression for gravitational potential energy. This establishes *center of mass* as a natural property with explanatory priority when accounting for this sort of behaviour of mechanical systems.

By contrast, consider an electromagnetic system in which the dynamics of charged particles is governed by Maxwell’s laws. Here, the center of mass plays no fundamental role in the laws governing electromagnetic interactions. If the center of mass happens to appear in calculations about this system, it does so as a fortuitous algebraic combination rather than as a natural property with explanatory significance. At no point does the center of mass emerge as a unified concept in explaining the dynamics. The explanation depends on the individual masses and charges, not on their weighted average position. On Lange’s account, even though we could mathematically calculate the center of mass for this system, it would not qualify as a natural property because it does not appear as a unit in explaining the dynamics. It represents a mere mathematical combination without physical significance or explanatory priority.

A third group of accounts of explanation, focused on the pervasive role of modeling and idealization in science, also take the theory-relative approach. Many explanatory models are highly idealized (Strevens 2008; Batterman 2010, Bokulich 2011, Batterman and Rice 2014, [self-reference omitted]). For idealized models, explanations are not deducible from laws of the high-level theory, and in fact may be inconsistent with those laws. Some idealized models are explanatory, and these explanations describe the model but contain false statements about the

real-world explanatory target. Others idealized models are not explanatory. What is the difference?

In earlier work, I developed an account of idealized-model explanation that attempts to answer this question. For idealized models, the explanation is not deducible from or licenced by high-level theory in the way Kitcher and Jansson require. Rather, the pathways from high-level theory to local model involve procedures such as abstractions, singular perturbation methods, and contrary-to-fact assumptions. These pathways are not contained within the high-level theory, and the high-level theory does not provide an algorithm for constructing local models for specific target systems. Nonetheless, under certain conditions these pathways provide explanatory justification for the local model ([self-reference omitted]).

All these accounts of scientific explanation share two features that together define the theory-relative approach. First, they hold that the explanatory status of an argument depends not just on features of the argument itself, but on how the argument relates to relevant high-level theory. For Jansson, the laws or empirical generalizations contained in the argument must be licenced by a high-level theory. On Lange's account, the laws or empirical generalizations must themselves be explainable by high-level theory in such a way that properties appearing in them are natural properties. For idealized model explanation, entities referred to in an explanatory model must be physically possible within high-level theory. Second, and crucially, the dependence of explanatory status on extra-argumentative factors is strictly limited to features of the relevant high-level theory. Other factors—disciplinary goals, community standards, historical period—play no role. The remainder of this paper contends that the latter point is mistaken and that accounts of scientific explanation should adopt an approach in which other disciplinary factors may play a legitimate role in determining what counts as a genuine explanation.

3. Challenges to theory-relative accounts

Let us examine two case studies that challenge the theory-relative approach. I argue that in these cases high-level theory alone is insufficient to account for what makes the explanations successful. These are the case of Archimedes' principle, treated at length by Lange, and the case of idealized model explanations of gravitational waves. These cases demonstrate that additional, non-theoretical contextual factors, including disciplinary goals, play essential roles.

Lange uses the case of Archimedes' principle to articulate and defend his account of explanation (2018, Ch. 11). Archimedes' principle says that F , the buoyant force on an object in a fluid, is equal to W , the weight of the parcel of fluid displaced by the object. From this principle and a determinate value for W one can derive the value of the buoyant force on an

object (with many practical applications, such as determining whether a crown is made of pure gold). Whether this derivation is explanatory depends on a fact about the explanation of Archimedes' principle *itself*, namely whether W appears in the explanation of Archimedes' principle as a natural property.

Lange considers two contemporary explanations of Archimedes' principle. In the field of hydrodynamics, Archimedes' principle is well explained in terms of the pressure difference on the upper and lower surfaces of the object. In this explanation of the principle, W is not a natural property, in the sense that it does not function in hydrodynamics as a unit. W is simply the product of the density of the fluid, the acceleration due to gravity, and the volume of the object. That is to say, W is not explanatory because it is simply the product of three base-level properties, just like center of mass in the electromagnetics case. The key point for Lange is that hydrodynamics does not explain *why* this combination of three disparate base-level properties should appear in Archimedes' principle (2018, 388). In the context of hydrodynamics alone, therefore, W does not function as a natural property with explanatory power.

The second explanation of Archimedes' principle is in the context of the law of energy conservation, which is a universal constraint on all fundamental laws in physics. The explanation begins by noting that as an object rises in a fluid, its gravitational potential energy increases by the weight of the body times the distance travelled. At the same time, the potential energy of the displaced fluid decreases by W times the distance travelled. From there, the net force on the submerged object—Archimedes' principle—can be derived and explained. Conservation of energy explains Archimedes' principle in a different way than hydrodynamics, and the weight of the displaced fluid W comes in at the start, as a unit. Here, W functions in the explanation as a natural property. On Lange's criterion, this establishes W as a natural property wherever Archimedes' principle is applied: W is natural enough to explain if and only if there is any explanation of the principle in which W enters as a unit, and the energy conservation explanation provides exactly this. Lange concludes, "The explanation from energy conservation makes W natural enough to join Archimedes' principle in explaining the magnitude of the buoyant force" (2017, 392). On his account, this explanatory status holds not only in the context of energy conservation but wherever Archimedes' principle is deployed, including in hydrodynamics.

On Lange's account, however, the explanatory status of W in hydrodynamics is grounded not in anything internal to hydrodynamics nor in what Archimedes principle enables in terms of prediction, intervention, or understanding within a disciplinary context. Rather, it is grounded in a fact about an entirely separate theoretical domain, namely whether W enters as a unit in the energy conservation explanation of Archimedes' principle. This constitutes a significant cost for any normative philosophical account of scientific explanation that aims to represent actual

scientific practice. Constrained contextualism offers a more principled account on which Archimedes' principle is explanatory in hydrodynamics because it directly enables the counterfactual reasoning required for causal intervention and control within a disciplinary disciplinary that has this as a goal.

One might worry that scientists in such a community are simply mistaken about the explanatory power of Archimedes' principle, and, for example, that the principle was indeed merely predictive until energy conservation was developed in the mid-19th century. Yet Archimedes' principle is explanatory in hydrodynamics on philosophical accounts of explanation as well, for reasons that have nothing to do with energy conservation. On Woodward's approach, for instance, a causal explanation contains a causal generalization that enables us to answer a range of "what-if-things-had-been-different" questions. The generalization does this by supporting the correct counterfactuals about what would happen under relevant interventions on the system. Here, Archimedes' principle is the generalization that does this, allowing us to determine the truth of relevant counterfactuals such as, "if the weight of displaced water were doubled, the buoyant force on the object would double." As we shall see in the next section, Woodward draws a tight connection between the scientific goal of intervention and the goal of explanation.

The case of Archimedes principle is meant to motivate the idea that limiting philosophical accounts of explanation to consider only theoretical context does not fully capture the explanatory practices of science. Gravitational waves provide a similar challenge to the theory-relative approach. Here, rather than a connection between goals of explanation and manipulation, the disciplinary goal of understanding is the most salient. Einstein predicted the existence of gravitational waves shortly after developing his theory of general relativity. The most common source of gravitational waves are systems of two massive objects, such as black holes or neutron stars, orbiting each other and gradually spiralling closer together, eventually merging in a collision. During this binary inspiral process, they release energy in the form of gravitational waves. Unlike light and other forms of electromagnetic radiation, gravitational waves pass through almost everything, making them extremely difficult to detect. In 2015, gravitational waves were detected for the first time, confirming Einstein's century-old prediction and opening a new way to observe the universe.

Gravitational wave phenomena are also extremely difficult to predict and explain. Two distinct theoretical approaches are used to model gravitational waves from binary inspiral systems. The traditional Post-Newtonian (PN) methodology, which traces its origins to Einstein's work, divides spacetime into near and far zones, and applies approximations of the Einstein Field Equations, fundamental equations of general relativity. PN predicts gravitational waves by employing different approximation schemes of the fundamental equations in each

zone and matching them at their boundaries. Developed over many decades, this approach has achieved remarkable success, predicting gravitational waveforms closely matching experimental results. However, computational complexity increases exponentially as the approximation methods are made more precise, introducing technical challenges that severely limit further progress using this framework.

The effective field theory (EFT) approach, developed in 2006, represents a methodological paradigm shift by adapting techniques from quantum field theory to classical gravitational systems. This framework conceptualizes gravitational effects as effective fields operating at different length scales, utilizing familiar mathematical tools from particle physics such as Feynman diagrams and renormalization group techniques. The EFT approach makes use of quantum-like models that are entirely fictional and in fact physically impossible in the classical theory of general relativity. In a matter of months, the EFT approach is able to reproduce results that required years to derive through traditional PN methods, and it now has predictive accuracy and scope well beyond the capabilities of the PN approach ([self-reference omitted]).

In earlier work, I contrasted PN and EFT approaches and argued that despite their similar predictive success, they differ significantly in explanatory power. My claim was that PN models can genuinely explain gravitational waves while EFT models cannot. PN models are highly idealized, yet the various pathways to construct a PN model of an inspiral system from the fundamental equations provide explanatory justification for the model. By contrast, EFT models make ineliminable use of physically impossible assumptions, including point particles as fundamental building blocks as well as impossible quantum-like properties in a classical theory. Thus, PN models can explain gravitational waves while EFT models cannot ([self-reference omitted]). The argument assumes a theory-relative approach, in which explanatory power depends on the right sorts of connections between the explanation and the high-level theory in which it is embedded and no other contextual factors are relevant.

Here, as in the case of Archimedes' principle, there are good reasons to suspect that consideration of high-level theory in gravitational physics alone is not sufficient to capture the range of contextual factors relevant to the explaining gravitational phenomena. Gravitational physicists often treat gravitational EFT models as explanatory, not merely predictive. When Goldberger and Rothstein introduced gravitational EFTs, their work was celebrated precisely because it provided novel and distinct explanatory insights into inspiral systems (Goldberger and Rothstein 2006). EFT models were held to explain gravitational wave phenomena by identifying the relevant degrees of freedom at each scale and showing how these interact according to well-defined principles. According to one of the founders of EFT gravitational models, this approach provides not just computational advantages but also deeper physical

insight due to its unified treatment of a diverse range of gravitational phenomena (Foffa and Sturani 2014). Indeed, this is the central theme of an influential review of gravitational EFTs by R.A. Porto. As he puts it, “We will attempt to convince the reader that adopting an EFT framework, when possible, greatly simplifies the computations and provides the required intuition for ‘physical understanding’” (Porto 2016, 5). EFTs in gravitational physics, he argues, “allow us to get a grip on the analytic side, often providing a deeper and more systematic understanding of the dynamics” (2016, 4).

As in the case of Archimedes principle, one cannot simply take the scientists’ intuitions at face value. I suggest that one good reason proponents of EFT gravitational models take them to be explanatory has to do with the unified structure of the models. EFT models are used in gravitational physics in the same way they are used across a wide range of scientific fields in which systems exhibit different behaviours at different energy or length scales: quantum field theory (from which gravitational EFT models are adapted), condensed matter physics, fluid dynamics, even biology and chemistry where hierarchical organization makes them useful for modeling complex phenomena on multiple levels. These disparate EFTs use similar techniques, such as renormalization, to make tractable what were previously very difficult or impossible calculations. In this way, EFTs enable a kind of understanding that many philosophers and scientists link with scientific explanation (a point we return to in the next section).

In these case studies, what makes an argument a good scientific explanation is not just their connection to high-level theory but also their role in furthering distinct scientific goals within disciplinary communities. Archimedes’ principle gains explanatory power in hydrodynamics not through some theoretical connection to energy conservation but because it aligns with the discipline’s goals of intervention and control. Similarly, EFT models are explanatory not because they can be reconciled with the fundamental field equations of general relativity but because they provide understanding through cross-disciplinary unification. Disciplinary goals constitute a fundamentally different category of context from theoretical structure. Recognizing this distinction requires moving beyond theory-relative approach to an account that explicitly incorporates how explanatory standards emerge from the intersection of theoretical constraints and broader scientific objectives.

4. Constrained contextualism

I propose an alternative approach to explanation I call *constrained contextualism*, in which contextual factors beyond high-level theory may be relevant to the explanatory value of an argument. Our starting point is the idea that explanation is not disconnected from other scientific aims. Indeed, explanatory goals are sometimes closely aligned with other disciplinary

goals. In other words, whether an argument is a valid explanation may depend in part on how well it serves other objectives, such as prediction, intervention, representation, realism or understanding. How explanation aligns with these sorts of goals, and the mix of goals that is important, depend upon the state of a scientific discipline at a given time, a contextual dependence that goes beyond the relevant high-level theory. Therefore, what renders an argument explanatory may depend on this additional context. Let's unpack this argument.

Often, explanatory success in a discipline is important not just for its own sake but also because it furthers other disciplinary goals. Causal information about a target system, for example, is often relevant to interventions on the system. This causal information is gleaned from models that accurately represent the values of observable variables and dynamics. Proponents of causal explanation have long maintained a connection between our interest in causal knowledge and causal explanation, on the one hand, and our practical interests in prediction, manipulation and control on the other. Having causal knowledge underwrites scientific explanation as well as prediction and intervention. Woodward draws this connection very tightly; for him, knowledge about invariant relationships in a system is necessary for knowledge of the effect of interventions on the system (both practical and in principle), which in turn is necessary for causal explanation. Woodward describes a range of scientific fields in the social and biological sciences where scientists explicitly link the progression of their field from a descriptive to an explanatory science with gaining knowledge of causal relations and mechanisms that enable manipulation and control (Woodward 2003, Ch. 2).

The goal of scientific realism, that is, knowledge of unobservable entities and processes, is also closely connected with explanation. Typically, inference to the best explanation (IBE) arguments provide this link. The idea is that given a range of candidate explanations for a given phenomenon, we can infer that the best explanation of the phenomenon is the true hypothesis, provided that the best explanation meets some standard of goodness as an explanation. Variants of this argument conclude that the hypothesis is closer to the truth, or more likely to be true, than the others. IBE is often supported by claiming that what makes an explanation successful—picking out causal structure, for example, or unifying disparate phenomena—also make the explanation likely to track truth. Alternatively, as Peter Lipton argues, the methods scientists use to develop explanations, such as controlled experimentation and elimination of alternatives, are also effective to track truth. Hence IBE, since the very reliability of these methods is best explained by realism (Lipton 2004). Recently, Chris Pincock made the case for a much stronger link between realism and explanation, arguing that the nature of evidential relations themselves give rise to explanatory connections, and in many cases these explanatory connections involve unobservables (Pincock 2024).

The connection between realism and explanation illustrates how explanatory standards do vary across communities with different scientific goals. Disciplinary communities committed to causal explanation typically hold that genuine explanation requires identifying unobservable mechanisms, processes, or entities responsible for a phenomenon. For such communities, an account that confines itself to observable regularities—however accurate and predictively powerful—falls short of genuine explanation. Other communities reject this standard, accepting theories as fully explanatory without positing unobservable causal relations or entities. Of course, one might maintain that one of these communities simply has the correct view of what valid scientific explanation requires and the other does not, independent of whatever other scientific aims the community holds. The constrained contextualist framework offers an alternative approach, emphasizing that divergent yet correct explanatory standards in this case should be tied to substantive differences in scientific goals.

Understanding is another goal closely aligned with scientific explanation. The connection is drawn by Kitcher, for example, who argues that explanations provide understanding by showing how seemingly different phenomena can be unified under common patterns of reasoning. Understanding increases as we reduce the number of independent assumptions needed to account for diverse phenomena. There are a variety of other approaches here, but common among them is the idea that scientific explanation is a vehicle to produce understanding. Understanding is a *sui generis* epistemic state that is distinct from knowledge, even explanatory knowledge. Good explanations generate understanding, so that if one knows the explanation of a phenomenon, then one understands it. On one common approach, when a theory is understood it becomes intelligible to scientists, which enables them to recognize qualitative consequences of the theory without performing precise calculations. Explanations are valuable, then, because they lead directly to improving scientists' practical reasoning skills to work with a theory (de Regt and Dieks 2005; de Regt 2017).

The goals of explanation and understanding can come apart in important ways. On de Regt's account, a scientist who grasps an explanation thereby understands the phenomenon explained. A scientific community that prioritizes understanding as a primary goal will hold explanations to standards that differ markedly from those of a community focused principally on prediction or intervention. De Regt's intelligibility criterion, for instance, reflects an interest in achieving the kind of practical, qualitative grasp of a theory that enables productive scientific work, even at some cost to formal precision or mechanistic detail. Constrained contextualism accommodates this: the explanatory standards appropriate to a given community reflect, in part, how that community weights understanding relative to other scientific goals.

The common theme here is that what counts as a genuine explanation sometimes depends in part on how well it aligns with other scientific goals. The foregoing is meant to illustrate a few

common ways in which this can happen. Whether explanatory success is linked with predictive power, accurate representation, intervention, realism or understanding, a combination of them, or other goals entirely, will vary across disciplines. Different scientific communities may have different objectives even in the context of the same higher-level theory.

A classic example of contextual dependence that goes beyond theoretical structure involves the very different approaches to quantum-mechanical explanation between the Copenhagen interpretation and the pilot wave theory developed by Louis de Broglie and later expanded by David Bohm. To oversimplify somewhat, both groups worked with the same theoretical formalism and experimental evidence, yet they had differing scientific goals. In the Copenhagen interpretation, the goal of quantum mechanics was not to give a realistic account of the microscopic world and certainly not to describe its causal structure. Rather, the goal was to provide as much predictive power and, where possible, understanding of a realm radically different from that of everyday experience. For the pilot wave group, by contrast, the most important goal was realism, that is, the fullest possible knowledge, using classical concepts, of unobservable quantum entities and processes.

Different approaches to explanation are aligned with these disparate goals. For the Copenhagen group, the explanatory power of quantum mechanics is a consequence of its ability to unify a wide range of phenomena under a small set of theoretical principles, first and foremost the Schrödinger equation. This explanatory unification produces understanding. By contrast, proponents of the pilot wave theory take a mechanistic approach to explanation, not in terms of causal mechanisms (due to non-locality) but in terms of classical deterministic physical processes. These explanations function as premises in IBE arguments in support of realist commitments to subatomic particles with classical properties. What are we to make of the disparate standards of explanation? I submit that just like the cases of Archimedes' principle and gravitational waves, in quantum mechanics extra-theoretical factors such as community objectives of realism, predictive power or understanding play a role.

As in the discussion of realism and explanation above, one might simply claim that the Copenhagen interpretation does not provide genuine explanations at all. While its calculations and operational rules are predictively powerful, they are explanatorily inert. Only the pilot wave theory, with its commitment to a realistic underlying ontology, can underwrite explanations of quantum phenomena. Or perhaps it's the reverse. In any case, this sort of position requires one to have a correct, non-contextual account of explanation that discounts both scientists' self-understanding about explanation and their wider disciplinary goals. Constrained contextualism provides another, and I claim a better, alternative. Constrained contextualism acknowledges that Copenhagen and pilot wave proponents are operating with different standards of

explanatory success, and these reflect a genuine difference in how the relevant community goals shape explanatory norms.

Have we arrived at an approach to scientific explanation unmoored from theoretical or other objective constraints on explanation? Does this mean that almost any set of facts could be explanatory of any phenomenon, in the right scientific context? The worry here is that constrained contextualism makes explanation too subjective, variable or discipline-dependent. Let me emphasize, however, that constrained contextualism does not claim that any argument can count as a good explanation when a community decides it does. Rather, it maintains that explanatory standards emerge from the intersection of (i) objective features of the world as captured by relevant theories and models, (ii) normative standards of explanation as articulated in philosophical accounts, and (iii) properly scientific aims of a scientific community, significant among which are prediction, representation, intervention, realism, and understanding. The latter are often referred to as epistemic aims of science, in contrast with non-epistemic aims such as guiding policymaking, obtaining funding, and gaining public support. In this way, constrained contextualism can account for the relative stability of explanatory standards across scientific communities and time periods, because explanatory standards are tied to other goals that persist across contexts. There is a shared nature of many scientific goals and a cumulative character of scientific knowledge. Similar epistemic goals are common across much of science, albeit with different emphases. It may even be that as scientific knowledge accumulates, certain explanatory approaches prove more successful at advancing these goals than others, leading to a convergence of explanatory standards.

I contend that constrained contextualism avoids “anything goes” relativism through three principled constraints. First, only contextual factors tied to established epistemic aims of science can influence explanatory standards. These aims are not arbitrary preferences but constitute what makes an enterprise recognizably scientific. Second, any proposed explanation must maintain consistency with the mathematical formalism and empirical content of the relevant high-level theory, per theory-relative accounts. Third, the connection between explanatory success and other scientific goals must be demonstrable through the actual content of the explanation—not merely asserted by a community. For instance, Archimedes' principle is explanatory in hydrodynamics not simply because scientists say so, but because the principle directly enables the counterfactual reasoning required for intervention. These constraints ensure that while explanatory standards may vary across contexts sharing the same theoretical framework, such variation is bounded by both the objective features of the world (captured in our theories) and the constitutive aims of scientific inquiry.

There will, of course, be boundary cases. These often arise as explanatory standards change and evolve within a discipline. For example, when EFT models were first developed gravitational

physics, they were regarded by most physicists as effective calculational tools rather than models providing any explanation or understanding. Over time, EFT models emerged as a competing explanatory strategy in the field. As we have seen, by identifying relevant degrees of freedom across scales, they are able to predict gravitational wave phenomena using a model successfully applied in a wide range of other scientific fields. Most importantly, EFT models are conceptual tools that enhance scientists' ability to work with gravitational systems without engaging in first-principle calculations from higher-level theory, thereby rendering the theory intelligible, and intelligibility is a key element of understanding. What this example illustrates is that these sorts of changes are neither arbitrary, nor merely pragmatic, but reflect the community's evolving standards for what constitutes legitimate explanatory practice. This prevents constrained contextualism from collapsing into extreme relativism while still acknowledging the legitimate contextual variability of explanatory standards.

5. The diverse aims of science

My argument for constrained contextualism begins with the assumption that explanatory goals are closely tied to other scientific aims, at least in some cases. It is precisely this additional context—alignment with other scientific aims—that distinguishes constrained contextualism from theory-relative accounts. As we have seen, alignment between scientific goals occurs when the features that make an argument explanatory are the same features that advance other disciplinary objectives. For instance, goals are aligned when the causal information that underwrites a good explanation also enables successful intervention, or when a good explanation supports a realistic inference about unobservables. The connection must be substantive rather than accidental. It is not enough that the same argument happens both to explain and to predict; rather, its explanatory virtues must derive from or contribute to its predictive power.

According to Angela Potochnik, however, there is rarely if ever any such alignment between the aim of explanation and other scientific goals. On her view, the opposite is true: the diverse aims of science are often in tension (Potochnik 2015; cf. 2017; 2018). If Potochnik is right, constrained contextualism is not a viable approach to explanation and not relevant to the practice of science. Potochnik is certainly right about the diversity of scientific aims, and she has done important work to highlight the relationship between these aims and the widespread use of highly idealized models across scientific disciplines. Taking a closer look at her argument will enable us to better understand how and why explanatory and other scientific goals may align and thus clarify the scope of constrained contextualism about explanation.

Potochnik argues that various scientific aims, such as explanation, prediction or accurate representation, require different types of scientific activities and produce distinct kinds of scientific products (2017). Key to her argument are the pervasive and multifaceted uses of idealized models in science. A model of a target system that provides an accurate representation of the system's components and properties may be poorly suited to predict its future dynamics. For example, in fluid dynamics or statistical mechanics, where systems have many components, models that aggregate or coarse-grain deliver the most predictive and explanatory power. By contrast, very different models, ones that accurately represent the system's individual components and their properties, are good for goals of accurate representation and realism. In other cases, the goals of prediction and explanation may pull in opposite directions. As Potochnik notes, an idealized model that provides a good explanation of a system's behaviour may leave out many causal details to focus on one causal difference-maker essential to the explanation. Such a model abstracts away from causal details that may be essential to making accurate predictions (2018).

Potochnik also has a stronger claim to make about the relationship among scientific aims. It is not just that some idealizations well suited to achieve one aim may be poorly suited for others, but more generally that the aims themselves are in opposition. There is an intuitive sense in which various aims of science are aligned; it seems normal for a model that accurately represents a target system also to be useful for prediction and manipulations. Potochnik is not just claiming that this may not always be the case, that this intuition about the alignment among scientific aims is occasionally wrong, or that successful representation sometimes pulls in a different direction from prediction or intervention. Rather, as she puts it, "...something stronger is true: success with one aim often inhibits success with other aims" (2015, 75; cf. 2017, 204); "Science's diverse aims, both epistemic and non-epistemic, often conflict..." (2015, 79). Specifically, she claims that "Accurate prediction is achieved by tools poorly suited to explain" (2015, 75; cf. 2017, Ch. 7).

How often do aims really conflict in this way? Can we say categorically that prediction and explanation pull in different directions? These are empirical questions about the practice of science, and we have some evidence that the conflict is not as widespread as Potochnik's stronger claim asserts. For Woodward, the aims of explanation, prediction and intervention are closely aligned. In numerous cases across a wide range of scientific disciplines, he shows that exactly the same knowledge of invariant causal relationships in a system underwrites predictions about the system's behaviour, the effect of interventions on the systems and causal explanation (2003). I argued above that explanations using Archimedes' principle are a case in point: the explanatory power of the principle is a product of the principle's effectiveness for the prediction and control of hydrodynamic systems. Similarly, we have seen that there is an

alignment between the goals of explanation and realism in IBE arguments, with arguments from Lipton and Pincock about why this is so.

I recognize that IBE arguments are applicable only under very limited conditions at best. Likewise, many explanations in science are non-causal, with no reference to causal notions of intervention or invariant relationships. Some explanatory models in science involve coarse-graining techniques that forego any accurate representation of the target system. I am not making an empirical claim about the extent to which aims align in scientific practice, except to say we have reason to question Potochnik's stronger, categorical claim that they do not. For our purposes, the important point is that the scope of constrained contextualism—where extra-theoretical factors influence what counts as a *bona fide* explanation—is limited to cases in which there is such an alignment.

6. Conclusion

This paper argues that constrained contextualism provides a useful framework for understanding scientific explanation. It is a framework that, in some cases at least, better represents actual scientific practice and better supports normative philosophical accounts of explanation than do the traditional approach, theory-relative approaches, or a radical “anything goes” approach. It acknowledges the role of both theoretical constraints and disciplinary objectives in determining what counts as a genuine explanation. By foregrounding how explanation connects with other scientific goals within a disciplinary context, it aims to capture both the diversity and the normativity of explanatory standards in science. It maintains that explanations track objective features of the world while acknowledging that what renders an argument explanatory depends partly on the goals and methods of the local scientific community.

As always, the proof is in the pudding. Our focus has been on approaches to explanation rather than particular accounts, an investigation into the importance of context, and how explanation connects with other scientific goals—a meta-analysis of explanation, so to speak. I have not developed an account of scientific explanation that takes a constrained contextualist approach. What's more, scientific aims are sometimes in tension, limiting the extent to which other goals function as constraints on explanation. Nonetheless, I hope to have convinced you that constrained contextualism about explanation is worth serious consideration.

Acknowledgements

[Omitted]

Disclosure of interest

The author reports there are no competing interests to declare.

References

- Barnes, E. (1992). "Explanatory Unification and the Problem of Asymmetry." *Philosophy of Science* **57**: 558–571.
- Batchelor, G. K. (2000). *An introduction to fluid dynamics*. Cambridge, Cambridge University Press.
- Batterman, R. and C. Rice (2014). "Minimal Model Explanations." *Philosophy of Science* **81**(3): 349–376.
- Batterman, R. W. (2010). "On the Explanatory Role of Mathematics in Empirical Science." *British Journal for the Philosophy of Science* **6**(1): 1–25.
- Bokulich, A. (2011). "How scientific models can explain." *Synthese* **180**(1): 33–45.
- de Regt, H. W. (2017). *Understanding Scientific Understanding*. New York, Oxford University Press.
- de Regt, H. W. and D. Dieks (2005). "A Contextual Approach to Scientific Understanding." *Synthese* **144**(1): 137–170.
- Foffa, S. and R. Sturani (2014) "Effective field theory methods to model compact binaries." *Classical and Quantum Gravity*, **31**(4), 43001. DOI: 10.1088/0264-9381/31/4/043001.
- Goldberger, W. D. and I. Z. Rothstein (2006) "Towers of Gravitational Theories." *General Relativity and Gravitation*, **38**(11), 1537–1546. DOI: 10.1007/s10714-006-0345-7.
- Jansson, L. (2015). "Explanatory Asymmetries: Laws of Nature Rehabilitated." *Journal of Philosophy* **112**(11): 577–599.
- (2020). "Network explanations and explanatory directionality." *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**(1796): 20190318.
- Kitcher, P. (1981). "Explanatory Unification." *Philosophy of Science* **48**: 507–531.
- (1985). "Two approaches to explanation." *Journal of Philosophy* **82**: 632–639.
- (1989). "Explanatory unification and the causal structure of the world" in P. Kitcher and W. C. Salmon, eds., *Minnesota Studies in the Philosophy of Science, volume XIII*. Minneapolis, University of Minnesota Press: 410–506.

- Lange, M. (2017). *Because without cause: non-causal explanations in science and mathematics*. New York, NY, Oxford University Press.
- (2018). "Because Without Cause: Scientific Explanations by Constraint." In Reutlinger, A., & Saatsi, J., eds., *Explanation beyond causation: philosophical perspectives on non-causal explanations*. Oxford University Press. DOI: 10.1093/oso/9780198777946.001.0001.
- (2019). "Asymmetry as a challenge to counterfactual accounts of non-causal explanation." *Synthese*, 198(4), 3893–3918. DOI: 10.1007/s11229-019-02317-3.
- (2023). "Explanations by Constraint: Not Just in Physics." *International Studies in the Philosophy of Science*, 36(4), 265–277. DOI: 10.1080/02698595.2023.2298085
- Lipton, P. (2004). *Inference to the best explanation*. London; New York, Routledge/Taylor and Francis Group.
- Pincock, C. (2024). "Defending a Realist Stance." *International Studies in the Philosophy of Science* 37(1-2): 1–15.
- Porto, R. A. (2016). "The effective field theorist's approach to gravitational dynamics." *Physics reports* 633(C): 1–104.
- Potochnik, A. (2015). "The diverse aims of science." *Studies in history and philosophy of science. Part A* 53: 71–80.
- (2017). *Idealization and the aims of science*. Chicago, The University of Chicago Press.
- (2018). "Eight Other Questions about Explanation." In Reutlinger, A., & Saatsi, J., eds., *Explanation beyond causation: philosophical perspectives on non-causal explanations*. Oxford University Press. DOI: 10.1093/oso/9780198777946.001.0001
- Strevens, M. (2008). *Depth: an account of scientific explanation*. Cambridge, Mass., Harvard University Press.
- van Fraassen, B. (1985). "Salmon on Explanation." *The Journal of Philosophy*: 630–651.
- (1977). "The Pragmatics of Explanation." *American Philosophical Quarterly* 14(2): 143–150.
- (1980). *The scientific image*. Oxford, Clarendon press.
- Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford, Oxford University Press.