

Scientific explanation and understanding: unificationism reconsidered

Sorin Bangu¹

Received: 18 May 2015 / Accepted: 29 April 2016 / Published online: 12 May 2016
© Springer Science+Business Media Dordrecht 2016

Abstract The articulation of an overarching account of scientific explanation has long been a central preoccupation for the philosophers of science. Although a while ago the literature was dominated by two approaches—a causal account and a unificationist account—today the consensus seems to be that the causal account (in one of its forms) has won. In this paper, I challenge this consensus and attempt to revive unificationism. More specifically, I aim to accomplish three goals. First, I add new criticisms (partly based on historical episodes) to the standard anti-unificationist arguments, in order to motivate the need for a revision of the doctrine. Second, and most importantly, I sketch such a revised version. Then I argue that, contrary to widespread belief (and in agreement with a small minority), the causal account and this revised unificationist account of explanation are compatible. Moreover, I also maintain that the unificationist account has priority, since a most satisfactory theory of explanation can be obtained by incorporating the causal account (properly spelled out), as a sub-component of the unificationist account. The driving force behind this reevaluation of the received view in the philosophy of explanation is a reconsideration of the role of scientific understanding.

Keywords Explanation · Understanding · Unification

1 Introduction

When reflecting on the idea that “unification is the essence of scientific explanation”¹ it is important to begin by remembering that this proposal is

¹Friedman (1974, 15). There have been attempts to endorse, refine and correct the view by several others, in particular by Kitcher (1981 and 1989); see also Jones (1995a, b, 2012), Weber (1999), Schurz and Lambert (1994), Schurz (1999), Bartelborth (2002). The thought that an explanation-understanding-unification link exists has been entertained in the past by several great philosophers such as Kant and Whewell. See Morrison (2000, ch. 2) for a review.

✉ Sorin Bangu
sorin.bangu@uib.no

¹ Department of Philosophy, University of Bergen, 12 Sydnespllass, 5007 Bergen, Norway

grounded in two assumptions: roughly put, that (a) an explanation provides understanding,² and (b) understanding is achieved through unification. Here I shall therefore begin by examining more closely these two premises, and this will naturally lead us to tackling several other, less-explored aspects of unificationism. This preliminary charting of the territory is undertaken with an eye to the main aim of this paper: to flesh out a proposal as to how we should (re)conceive the relation between unification, understanding and explanation (of nomic regularities³).

More specifically, I will make the case for the following two theses. The negative one is that in its current form unificationism is flawed. My main reason for claiming this has not yet been discussed in the literature despite the fact that the doctrine has received a great deal of attention. In a nutshell, the central problem is that although the unificationists' original motivation was to provide a solution to the difficulties raised by the notion of scientific understanding, they, too, end up by advancing a rather impoverished conception of understanding.⁴

The positive thesis is that unificationism doesn't have to construe understanding in this way, since there is a better way, which I will try to articulate here. But such a revision of the unificationist account of understanding can't be achieved unless certain modifications to the current form of the unificationist framework are implemented. The central one has to do with the very notion of 'unification'; as a general term, it is too imprecise a notion to carry on the major burden of connecting explanation and understanding. Thus, drawing on several analyses of unification available in the literature, I urge that this general notion has to be replaced with a more specific one—'ontologically-reductive' unification⁵—which, as I hope to show, does live up to these expectations.

2 Explanation, unification and understanding

Assumption (a) above is the thought that the very point of an explanation is to convey understanding.⁶ When this is done successfully, we should not confuse this accomplishment with the subjective, pleasant feeling which typically accompanies it

² Not everybody believes that the only way to increase understanding is to provide explanations; for such a dissenting view, see van Fraassen (1985, 642). Although I think this is wrong, here I can't argue against this position.

³ I will also refer to these as 'phenomena', although I don't intend to confine the discussion to the realm of observable (of the type 'metals expand when heated'); regularities taking place at the microscopic level (e.g., the photoelectric effect) are also meant to be covered by the account.

⁴ Woodward (2011 and 2003, ch. 8) contain inventories of some of the main criticisms; Morrison (2000) discusses the notion of understanding in connection to unification in more detail. Neither, however, concentrates on Friedman's distinction on which I focus here, between *local* and *global* understanding.

⁵ It will be spelled out in Diagram f* in sect. 7 and 8 below. As I will make it clear later on, I will be making use of the distinction between various kinds of unification (including the ontological kind), which has been noted before by several authors, including Redhead (1984), Maudlin (1996), Morrison (2000) and Mäki (2001).

⁶ Following others, I shall take it to be the case that explanation and understanding are the two sides of the same coin. Kim (1994, 54), for instance, writes: "The idea of explaining something is inseparable from the idea of making it intelligible; to seek an explanation of something is to seek to understand it, to render it intelligible. These are simple conceptual points, and I take them to be untendentious and uncontroversial." For a recent dissenting voice, see Gijsbers (2013).

(the ‘aha’ interjection, the ‘eureka’ moment, etc.) In this subjective sense, understanding is a mental state which, like other mental states, can be the result of a variety of psychological processes, and not necessarily a sign of reaching a genuine answer to a why-question: the aim of scientific explanation is *objective* understanding, and this kind of understanding can only be achieved when such a correct, publicly available answer is available. In saying this, I thus acknowledge from the outset that I’m taking for granted the distinction between the state of affairs in which one actually, genuinely, objectively understands why a natural phenomenon happens, as opposed to the state in which one only has the subjective feeling that one understood. While I agree that the distinction needs further philosophical discussion, it seems clear enough to serve at least as a starting point here: we are all familiar with self-deception situations, in which the feeling occurs in us, and yet genuine understanding is missing. I therefore set aside the complications generated by the fact that any explanation has to be *grasped* by an epistemic agent, and, since this grasping has a subjective nature, the suspicion arises that the psychological/subjective aspect of understanding can’t ultimately be eliminated.⁷

Premise (b) is the insight that this objective form of understanding is enabled by theoretical unification. The idea is that the possession of a unified scientific picture has, somehow, a positive effect on our understanding. This sounds appealing (and not only to philosophers, but to great scientists as well⁸), and yet the thought is not as clear as it appears to be at first sight. Several questions crop up right away: do unificationists claim that *only* unification can lead to understanding? What about then the discovery of patterns of causal connection in the world; doesn’t this lead to understanding, too? Furthermore, isn’t this causal way the *only* way to explain? Moreover, we have to ask: understanding *what*, exactly? Here we must be careful again, since the notion of understanding takes on two different meanings in the literature.⁹ Understanding refers either to understanding why an individual phenomenon occurs, or to understanding ‘the world’, as a whole—i.e., in a ‘global’ fashion. As is clear then, the unificationists’ view on the ‘all-explanation-is-causal’ refrain needs discussing; also, the distinction between the ‘local’ and ‘global’ senses of understanding will be of major importance in what follows, despite being virtually forgotten in the recent literature on understanding.¹⁰

3 Whence understanding?

Most generally, explanation can be conceived as an ordered triple [EXS, d, EXD], consisting in the phenomenon to explain (i.e., the *explanandum*, abbreviated as EXD), the assumptions made when explaining (the *explanans*, EXS), and what I’ll call a

⁷ See Trout (2002) and De Regt (2004) for a recent exchange on these issues. See also DeRegt and Dieks (2005).

⁸ Wolfgang Pauli also entertained the thought. See Heisenberg (1971, 29)

⁹ I should say *at least* two meanings. As a referee pointed out, there are recent discussions of understanding—that, –how, –with etc. I’ll leave them aside for reasons of space.

¹⁰ The distinction was introduced by Friedman (1974), mentioned briefly by Kitcher and Salmon in their early work on explanation, discussed in passing by Barnes (1992), and totally ignored ever since. None of the recent discussions of explanation and understanding in science even mentions it. See for example Woodward (2011), Gijsbers (2013), Grimm (2013, 2014), Hindricks (2013), Khalifa and Gadomski (2013), Khalifa (2013), Newman (2013), Strevens (2004, 2013).

dependence relation d , which makes explicit how EXS and EXD are connected.¹¹ In the diagrams below, relation d is illustrated as follows:

$$\text{EXS} \dashrightarrow \text{EXD}$$

Perhaps the most natural example of such dependence is logico-mathematical entailment. Hempel's now classical deductive-nomological model serves as an illustration here. To explain, in this model, is to derive the EXD (as the conclusion of a valid deductive inference) from one or more laws of nature, together with some initial conditions—and these items together constitute the EXS. But, as many clever counter-examples show (e.g., the flagpole/shadow, the barometer/storm, etc.), this kind of triple is not satisfactory. In particular, they show that it is required that the relation d encode more 'substantial'—causal/'productive' (i.e., asymmetric)—dependencies in the world.¹²

At this point one must admit that it is hard to see how unification can enter this picture. This sentiment often motivates the skeptical attitude toward this approach, and a reconstruction of the unificationist project is thus needed; I begin it here, and continue it in the next three sections. As will hopefully become clear, the critics' frustration is only fair, since the most intuitive way to approach these matters is not in unificationist terms. This idea needs to be invoked as a last resort, only after other approaches turn out less satisfactory.

If the very point of explanations is to provide understanding, and explanations have the structure [EXS, d , EXD], then, as we saw, one immediately wonders how our understanding of EXD is being "produced" (Friedman's way of speaking, (1974, 6)), and how unification plays a role in this process. We can begin explicating this in terms of what I'll call here *schema E*:

A phenomenon (EXD) is explained iff:

- (i) its explanans (EXS) is identified, and
- (ii) relation d is specified.

Since objective understanding is not nominally mentioned here, the issue is where exactly, if anywhere, is its place in this schema? The problem of understanding is thus embodied in the following Question¹³:

How does a scientific explanation (of a natural phenomenon) produce objective understanding (of that phenomenon)?¹⁴

¹¹ The derivations are of course between *descriptions* of phenomena (i.e., sentences), not phenomena themselves. Although I'll speak loosely throughout the paper (as customary in this literature), the relevance of paying attention to the descriptions of phenomena will become crucial in section 7.

¹² Moreover, other types of derivation are possible, e.g., of statistical nature, or by taking limits (i.e., the 'asymptotic' type of explanation explored by Batterman 2002.)

¹³ I use a capital letter for it for further easier reference. Friedman (1974) takes it as a starting point.

¹⁴ Traditionally, the question asks about the 'epistemic gain' (Kim (1994, 54)) offered by an explanation.

Schema E suggests right away two general strategies to address this Question; they consist in supplementing conditions (i) and (ii), respectively. The supplementations will amount to additional clauses, which enable answering the Question. The first strategy primarily requires that what fills up the role of the EXS have certain properties; it can be illustrated in four ways, and in the next sections I shall examine them. None will be entirely unproblematic, but one—invoking the notion of unification—will arguably be better than the others. This notion (more precisely, Friedman’s (1974) version of it) will be involved in the fourth implementation. The second strategy consists in requiring that the instantiations of d satisfy certain constraints. There is one implementation of this strategy I’ll examine here, due to Philip Kitcher (1981; 1989), which also appeals to the notion of unification. I will discuss it in section 6 below.

4 The first strategy

This strategy consists in addressing the Question by constraining primarily the EXS. The thought underlying the first two implementations of this strategy below is the appeal to a kind of ‘closure’ principle: understanding EXD comes from understanding the EXS, in the sense that understanding is somehow inherited, or transmitted, from one to the other, via d . Thus, it is not surprising that one implementation of this strategy, natural although naïve, is to add to schema E the following clause:

(i+)₁ EXS is familiar,

(or, “less in need of an explanation” than the EXD¹⁵). If we show how EXS determines EXD, and the former is familiar to us, then, presumably, we can claim that we understand EXD.

As is perhaps obvious, this will not be satisfactory, for several reasons. The suggestion flies in the face of clear counterexamples—the liquidity of water is explained by the properties of water molecules, but we are surely more familiar with the former than with the latter. Moreover, familiarity is a hopelessly vague and subjective notion. Even if we identify a familiar EXS, all we can get from this is that we *feel* like we understood it (we are familiar with the familiar, and then, as a matter of psychological fact, aren’t tempted to question it), so we don’t need an explanation of it. This ‘need’, however, is of a psychological nature; pointing out that we don’t feel it does not *entitle* us to claim that we actually understood.

Another, more subtle, version of the same idea, is due to Toulmin (1961) and requires that EXS be familiar as well, but in a more different, and deeper way. It requires that it be an ‘ideal of intelligibility’, held by the scientific community, at a certain place and historical time. So, the clause to add is

(i+)₂ EXS is an ideal of intelligibility.

Does this answer the Question? It doesn’t, although it surely addresses it. Just as above, we may feel we understand EXS, because we find it familiar, and now this feeling

¹⁵ As Schurz (1999, 97) put it; see also Schurz and Lambert (1994, 105).

is reinforced by the fact that, by hypothesis, everybody around us finds it so. But, again, this is a matter of psychological (or sociological) fact; it doesn't mean that we are entitled to feel satisfied.¹⁶

The third implementation of the first strategy invokes the notion of causation. The primary requirement imposed on EXS would then be that

(i+)₃ EXS is the cause of EXD.¹⁷

According to this approach, to understand is to identify the cause. Given a phenomenon in need of explanation, to understand why it occurs is to identify what is 'responsible' for it—together with demonstrating a way (i.e., a dependency *d*), or a 'mechanism' to 'produce' it, or 'make it happen'.¹⁸ The idea is very popular, as E. Barnes summarizes: "The intuition that we understand an event in virtue of knowing its causal basis is no less sound than the intuition that we explain the fact by citing this causal basis. (...) To seek to understand almost any empirical *F* is just to seek the knowledge of its causal basis." (1992, 8)

This proposal would have surely answered the Question (assuming, as customary, that causal relations are objective), but stumbles upon a formidable, and well-known, obstacle. The appeal to causation faces the venerable battery of arguments against the very meaningfulness of this notion. After Hume, Russell was perhaps the most prominent figure suspicious of it, famously calling causation a "relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm." (Russell 1912, 1). Quine was equally critical ("(...) the notion of cause itself has no firm place in science. The disappearance of causal terminology from the jargon of one branch of science and another has seemed to mark the progress in the understanding of the branches concerned." (Quine 1976, 242)). Very recent authors are still uncomfortable with this notion; A. Ahmed, for instance, writes that "Causation is a pointless superstition" (2014, vii).

Yet, of course, nobody can deny the naturalness of condition (i+)₃, and lacking an account of causation should not be a reason to dismiss this clause; in fact, the model I sketch at the end will make room for it. For the moment, however, we'll just duly take note of the evidence, namely that (i+)₃ is (extremely) difficult to spell out; in saying this, I take it that at present there is no consensus on how to do this, despite the efforts of philosophers such as M. Scriven, J. Mackie, W. Salmon, D. Lewis, P. Humphreys, N. Cartwright, P. Dowe, D. Hausman, J. Woodward and, more recently, M. Strevens.

Given the appeal of (i+)₃, it's not surprising that the debate on explanation revolved around it. For quite a while (roughly, between 1950s and 1970s) it looked as if those who

¹⁶ However, this idea can't be pushed aside so quickly—as Friedman (1974) does—because of the following complication. It might be the case that everyone in a community claims that they have 'understood' (and thus have no further why-questions to ask) not just because they are intellectually-psychologically *comfortable* with EXS, but for a deeper reason: they may lack the necessary conceptual framework within which to conduct this further inquiry.

¹⁷ I ignore the difficulties raised by the uniqueness of causes.

¹⁸ As suggested, I include here (ignoring the obvious differences) the group of philosophers supporting the idea that mechanisms are what matters when it comes to explanation. The works by Machamer, Darden and Craver are exemplary in this respect.

disagreed with the logical-empiricist idea that understanding *is* ultimately an entirely subjective-psychological matter¹⁹ had no other option than to work out an acceptable theory of causation—and then, on the basis of such a theory, spell out (i+)₃; and, only after that, as a yet additional step, flesh out a causal theory of explanation. This was exactly what Salmon tried to do in the 1980s and 1990s (see Salmon 1984). However, Friedman, Kitcher, and several others afterwards, must be credited with the profound insight that another option was available: account for scientific explanation and objective understanding while bypassing the task of elaborating a theory of causation.²⁰ This dialectic is not always explicitly noted in the literature, but I submit that this is how the unificationist ideas must have originated—namely, as an attempt to decouple the articulation of a theory of explanation from the uncertain fortunes of a (metaphysical) theory of causation.

I will devote the rest of the paper to unpacking the triad explanation-unification-understanding, but a certain limitation of the subsequent discussion should be made clear at the outset. Although there are various ways to flesh out the unificationist insight, here I consider Friedman's and Kitcher's accounts to be the main philosophical expressions of it, so I'll engage only with their views.²¹ I will point out some differences and similarities between these two accounts, and I will argue (somewhat against the received view²²) that although Kitcher's proposal constitutes an admittedly necessary adjustment of Friedman's, it still fails to deliver; in the end, despite the convincing criticisms leveled against Friedman's key-idea, it is a (rather substantial) refinement of it (and, more generally, of the framework defined by it) that finally does the job.

5 The first strategy, once again: Friedman's unificationism

It was Michael Friedman's 1974 paper 'Explanation and Scientific Understanding' that first articulated explanatory unificationism fully and convincingly. Against the background I provided so far, this paper can be read as advancing a certain interpretation of what the problem of understanding is, and then, under this interpretation, as formulating an answer to it. I will discuss these two points in turn.

On my reading, Friedman's first idea is as follows: if the Question embodies the problem of scientific understanding, then this problem just can't be solved. As we have seen, none of the proposals above works. Yet, Friedman continues, this may be so because the problem was formulated under a certain assumption—that we explain (and understand) phenomena *individually*. I will call this (following him) the 'locality' assumption.

But, as Friedman also remarks, if we give the problem of understanding a different interpretation, we can tackle it more fruitfully. That is, instead on taking 'understanding'

¹⁹ Recall that Hempel's own suggestion was that understanding is to be identified with the subjective notion of nomic expectability (1965, 327).

²⁰ And even turn things upside-down, by accounting for the 'because' of causation on the basis of the 'because' of explanation, as Kitcher urges (1989, 477).

²¹ Schurz and Lambert (1994) detailed account deserves a special mention here (together with Schurz (1999)), as it remains the most carefully elaborated alternative to Friedman and Kitcher's ideas.

²² I take it to be expressed in Woodward's 2011 *Stanford Encyclopedia of Philosophy* article. Friedman's position is mentioned, but not discussed.

to mean ‘understanding an individual EXD’, Friedman suggests we should introduce a different notion: take ‘understanding’ to mean ‘global understanding’, or understanding *the world*. What is *global understanding*, then? In essence, achieving this kind of understanding amounts to reducing the number of why-questions on the scientific agenda, since a world raising less why-questions, i.e., containing less ‘mysteries’, is more comprehensible.

This global view of understanding is plausible, but it is at odds with the causal account: when implementing that account, we don’t decrease, but keep the same, the number of such questions—since the answer to the question ‘why does EXD happen?’ (namely, ‘it was caused by EXS’) simultaneously (and proverbially) raises *another* question ‘but, what caused EXS then?’ No reduction in the number of why-questions is achieved, hence no understanding is obtained; and this is so, to repeat, since ‘understanding’ is now interpreted in these global terms, as an exercise in why-questions elimination.

Under Friedman’s reformulation of the problem of understanding, the relevant question reads as follows—this will be, for convenience, Question*:

How does a scientific explanation generate objective understanding *of the world*?

This Question* is, I submit, standard unificationism’s main concern; and, as I noted above, a quite convincing answer to it can be proposed, to which I now turn.

The answer proceeds by imposing a different requirement on the EXS, namely that it must be “*more comprehensive*” (Friedman 1974, 19; italics in original). This means that the EXS should serve as a basis not only for the derivation of the phenomenon we wanted to explain initially, but for a multitude of *other* phenomena as well.²³ Next, when an EXS having this ‘comprehensiveness’ property can be identified (not an easy task!), we can call it a *unifier* of these various EXDs. Accordingly, the fourth proposal to complete schema E is to add

(i+)₄ EXS is a unifier.

We are now in the privileged position to: answer Question*, show how global understanding is getting produced, and that it is objective.

First, let’s make the additional and trivial assumption that doing science involves dealing with more than one puzzling phenomenon, i.e., with more than one EXD.²⁴ Let’s then index these elements accordingly: see Diagram e below, representing the traditional view on explanation. Now, *if* we happen to be in the situation that the explanans of a phenomenon EXD₁ is identified and it’s a comprehensive unifier (call it ‘UEXS’), then other explanations ([UEXS, d, EXD₂], [UEXS, d, EXD₃], etc.) can be formulated as well; see Diagram f. What Friedman argues, in essence, is that the

²³ A subsequent requirement is that *d* be a logical derivation; Friedman—and, as we’ll see, Kitcher too—follow Hempel and the logical positivists in this respect. I should note, however, that although the view I will eventually espouse here is unificationist, in the end I will not rush to follow Friedman and Kitcher in regard to the nature of *d*; as we’ll see (sect. 9), in my account I will ultimately leave the nature of *d* open.

²⁴ For instance, we want to explain the trajectory of projectiles in the air, as well as why gases expand when heated.

situation described in Diagram f is epistemically superior to the situation captured by Diagram e:

EXS₁ \mapsto EXD₁

EXS₂ \mapsto EXD₂

Diagram e

UEXS $\begin{cases} \mapsto \text{EXD}_1 \\ \mapsto \text{EXD}_2 \end{cases}$

Diagram f

What Diagram f describes is *unificatory explanation*, or explanation-via-a-unifier. Several remarks on these diagrams are in order.

Since (in Diagram f) UEXS serves as the explanatory basis for a series of EXDs, we gained something important: when explaining EXD₁ and EXD₂, instead of having to appeal to several, different EXSs (i.e., one corresponding to each EXD), we now have to invoke only one (UEXS). That is, instead of having to accept several brute, unexplained facts ('mysteries'), we have now to accept only one such fact.²⁵ Thus, the reduction of the number of phenomena we must accept as brute/unexplained is the main achievement of imposing the requirement (i+)₄. If Diagram f obtains, the world contains fewer 'mysteries'; hence, notes Friedman, the world is more comprehensible, and thus we gained understanding—of the global kind.

Furthermore, this *quantitative* reduction in the number of brute phenomena satisfies the condition to be an *objective* measure of increase in understanding. The required independence from one's preferences and historical context is now achieved: no matter what one happens to be familiar with, or what kind of metaphysical or epistemological 'ideal of intelligibility' (teleological, mechanical, etc.) happens to prevail during a historical period, anyone who basically knows how to count²⁶ should accept that it is an objective fact that our global understanding increases. This is so since the underlying key-idea—a world with few(er) mysteries is more comprehensible than one with many—is surely acceptable regardless of personal metaphysical/epistemological preferences and historical context.²⁷

Taking stock, we are at the point where we can see that, for (standard) unificationists:

²⁵ Despite what Diagram f may suggest, scientific research doesn't stop when a unifier UEXS is found; scientists will presumably treat it as an object of further investigation (a 'mystery') as well, and it will set the agenda for future scientific research. So, in a sense (as a referee has pointed out), unificationism is threatened by some form of infinite regress, or—just like causalism—by the incapacity to reduce the number of why-questions: if we take this number to be initially infinite, then eliminating a finite number of them will leave us with an infinite number still, and this can't count as a reduction. Here I shall set this difficulty aside; the assumption that the number of why-questions is infinite seems to me not strictly speaking false, but rather exaggerated.

²⁶ This is a point where we'll return later, when discussing the conjunction problem.

²⁷ This point aims to address Morrison's (2000, 26–29) criticism, to the effect that Friedman's account of understanding is not entirely objective, as it depends, after all, on what happens to be the explanatory ideal during a certain historical period.

- (a) explanation is only ‘wholesale’ (global), not ‘retail’ (local), i.e., explanation-via-a-unifier doesn’t yield understanding of individual phenomena, but
- (b) produces understanding of the world (i.e., of global kind), and
- (c) this global understanding is objective.

And yet this can’t be the end of the matter: claim (a) sounds worrisome. A unificationist must confront the strong intuition that, contrary to (a), explanations should yield gains in our understanding of individual phenomena too. This was the main insight to be retained from the causal doctrine: it identifies and articulates interdependencies between *parts* of the world. One may even hint that perhaps (Friedman’s) comprehension is not an unproblematic synonym for understanding. While unificationists aim to make our view of the world broader and more comprehensive, the causalists aim to fill in relevant details about the relations between parts of the world, and so to increase our understanding—yet in a different sense. It is the filling-in of these details that increases understanding in a sense that is more ‘constructive’, and contrasts with the more ‘reductive’ unificationist approach.²⁸

The standard unificationist doctrine dismisses locality (what underlies this ‘constructive’ sense of understanding), and although this dismissal is justified by the failures of proposals (i+)₁ to (i+)₃, on a more considered view the force of locality can’t be denied. Thus, I take it to be a desideratum for *any* theory of explanation to do justice to locality—and thus this poses a serious problem for standard unificationism.²⁹ This is to say that I am not content with how Friedman (and Kitcher) have dealt with locality—in essence, by setting it aside (as a result of interpreting ‘understanding’ to mean exclusively ‘global understanding’.³⁰) Yet, while this is a feature of standard unificationism, I submit that a revision of the doctrine *can* do justice to locality—and I shall try to show how in sect. 7.

Before we move on, two caveats are in order. First, the revised unificationism I’ll advocate here, might not convince (or not even interest) a unificationist who follows the orthodoxy rigidly, and thus ignores the local aspects of explanation. Yet I think it is fair to say that such an attitude would be unwise. While the global aspects are paramount, I also find it imperative to recognize the strong intuitive appeal of locality; in essence, the underlying thought here is that understanding has something of a dual nature. On one hand, it is global—it seems eminently correct to say that a world with fewer unanswered why-questions is more comprehensible. On the other, it is also undeniably local, in the sense detailed above—and note that this point holds *even if* we grant the unificationists the view that this doesn’t lower the total number of unanswered why-questions. This shows that the two kinds of understanding *can* coexist; and, as I believe, they *do* in fact coexist.

Thus, a more satisfying conception of scientific understanding seems to be one in which, schematically

$$\text{Understanding} = \text{Local} + \text{Global}$$

²⁸ This is part of Morrison’s point that unification comes apart from explanation (Morrison 2000, 4).

²⁹ The point has been made by Barnes (1992, 8).

³⁰ As Friedman emphasizes, science has as its ultimate target ‘understanding the world’ (1974, 19).

Then, the best we can hope from an account of scientific explanation is to maximize this ‘sum’. However, as we saw, standard unificationism takes care only of the global component; hence I consider a welcome improvement of the unificationist doctrine to be able to show that this account can make room for local understanding too. The ability to address the locality worry introduced above—how is our *understanding of an individual phenomenon increased* when we derive it from a unifier?—would be a desirable feature of unificationism.

The second caveat is the recognition that the dualism above does not assume to be obligatory for one and the same account of explanation to cover both the local and the global components. Instead—and this idea can be traced back to Salmon—one may say that the causal account of explanation gives us local understanding, while the unificationist account gives us global understanding. *This complementarity*³¹ seems tenable; however, the ambition here is to probe further, and show that unificationism can offer both. As suggested above, this would constitute the fulfillment of some kind of ideal for a theory of explanation; although not mandatory, I take this unified account of explanation to be philosophically preferable and methodologically in sync with unificationism as an account of explanation.

Now I shall put Friedman’s own brand of unificationism aside for a moment, and turn to Kitcher’s influential version of the doctrine.³² After I highlight how it is different from Friedman’s, I will also stress their (often unacknowledged) similarities, emphasizing the reasons why it should be eventually discarded.

6 The second strategy: Kitcher’s unificationism

Given my charting of the territory so far, Kitcher’s development of unificationism comes out naturally as another attempt to tackle Question*. Going back to schema E, Kitcher’s project illustrates the second strategy; he imposes constraints on the second component (ii), the derivation relation d .

The well-known idea is, in a nutshell, as follows. Given the multiplicity of EXSs and EXDs, and the variety of ways d to get from the EXSs to the EXDs,³³ the requirement is, in essence, to reduce the number and variety of the types of derivations d , or “argument-patterns” (1981, 515). The proposal amounts to claiming that the fewer and more “stringent” types of instantiations of the d relation, the more unified our scientific picture is, and thus the more our understanding increases.³⁴ Let’s denote this clause as (ii+).

Schematically, Kitcher’s point is that the situation described in Diagram k below is preferable to the one described in Diagram e. The term ‘ d_u ’ designates the d -unifier

³¹ See De Regt (2006) who discusses Salmon’s complementarity thesis.

³² As we’ll see in some detail in section 7, from here Friedman goes in the direction of worrying how we count the explanans. We’ll return to this later, as the issue is extremely important for the purposes of this paper.

³³ The set of such triples [EXS, d , EXD] corresponds to what Kitcher calls the “explanatory store” (1981, 512)

³⁴ Kitcher articulates this idea carefully, and some of the specifics are worth reflecting on, but here I’ll leave them aside, as they don’t really matter for what’s next (detailed expositions of them are available; see for instance Roland 2008)

argument pattern (just as with Diagram f above, I depict the simplified case when only one argument pattern is needed).

EXS₁ \mapsto EXD₁

EXS₂ \mapsto EXD₂

Diagram e

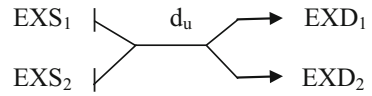


Diagram k

As it should be clear by now, once we hear of an increase in understanding, we have to ask for clarifications—understanding, in what sense: subjective or objective? local or global? Since we can count argument patterns (Kitcher’s never challenged assumption, as far as I know), we deal with objective understanding; as for local v. global, although Kitcher doesn’t use these terms, it is clear that he accepts Friedman’s global sense:

On both the Hempelian and the causal approaches to explanation, the explanatory worth of candidates—whether derivations, narratives, or whatever—can be assessed individually. By contrast, the heart of the view that I shall develop in this section (and which I shall ultimately try to defend) is that successful explanations earn that title because they belong to a set of explanations, the *explanatory store*, and that the fundamental task of a theory of explanation is to specify the conditions on the explanatory store. Intuitively, the explanatory store associated with science at a particular time contains those derivations which collectively provide the best systematization of our beliefs. Science supplies us with explanations whose worth cannot be appreciated by considering them one-by-one but only by seeing how they form part of a systematic picture of the order of nature. (Kitcher 1989, 430)

Moreover, when it comes to the connection between explanation and understanding, it is important to observe that Kitcher also accepts Friedman’s views *on understanding* entirely.³⁵ In terms of textual evidence, here are two rather long, but I believe revealing, quotes:

Friedman argues that a theory of explanation should show how explanation yields understanding, and he suggests that we achieve understanding of the world by reducing the number of facts we have to take as brute. Friedman’s motivational argument suggests a way of working out the notion of unification: characterize [the explanatory store] as the set of arguments that achieves the best tradeoff between minimizing the number of premises used and maximizing the number of conclusions obtained.

³⁵ One should not find this surprising. Just by reading carefully Kitcher’s two relevant works (1981 and 1989), one notices that each time he mentions the notion, he agrees with Friedman. In fact, he doesn’t seem to have much to say about understanding per se; for example, in his 1981 paper, he discusses understanding briefly at the outset on, p. 508–9, and then gets back to it only at the very end, on p. 529, where he agrees with Friedman.

Something like this is, I think, correct. Friedman's own approach did not set up the problem in quite this way, and it proved vulnerable to technical difficulties (see Kitcher 1976 and Salmon in (Kitcher and Salmon (1989))). I propose to amend the account of unification by starting from a slight modification of the motivational idea that Friedman shares with T. H. Huxley (see note 17³⁶). Understanding the phenomena is not simply a matter of reducing the “fundamental incomprehensibilities” but of seeing connections, common patterns, in what initially appeared to be different situations. Here the switch in conception from premise-conclusion pairs to derivations proves vital. *Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again, and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute)*. So the criterion of unification I shall try to articulate will be based on the idea that [the explanatory store] is a set of derivations that makes the best tradeoff between minimizing the number of patterns of derivation employed and maximizing the number of conclusions generated. (Kitcher 1989, 431–2; italics in original)

In conclusion, let me indicate very briefly how my view of explanation as unification suggests how scientific explanation yields understanding. By using a few patterns of argument in the derivation of many beliefs we minimize the number of *types* of premises we must take as underived. That is, we reduce, in so far as possible, the number of types of facts we must accept as brute. Hence we can endorse something close to Friedman's view of the merits of explanatory unification (Friedman 1974, pp. 18–19). (Kitcher 1981, 529–30)

Thus, initially Kitcher takes a different route than Friedman (as he proposes to add clause (ii+), instead of (i+)₄, to schema E). Yet, when he deals specifically with understanding, he simply defers to Friedman's proposal. A closer look at the last part of the italicized lines in the first quote (“...and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute)”), and at the middle part of the second quote (“By using a few patterns... number of types of facts we must accept as brute.”), bears this out eloquently. In the end, the consequence of his theory is precisely what Friedman's theory required: a reduction in the number of types of explanans. And this is not even surprising since, naturally, the constraints on the variety of patterns of derivation will turn into constraints on the variety and number of accepted types of premises.

This means that a better description of Kitcher's proposal is captured in Diagram k' below. Kitcher's unificationist strategy turns out to be a variation on the first unificationist strategy we have already examined (Friedman's), since it ultimately

³⁶ Note 17 in Kitcher's 1989 paper cites Friedman (1974); it runs as follows: “...our total picture of nature is simplified via a reduction in the number of independent phenomena that we have to accept as ultimate” (Friedman 1974, 18). There is an interesting recapitulation here of T. H. Huxley's summary of Darwin's achievement. “In ultimate analysis everything is incomprehensible, and the whole object of science is simply to reduce the fundamental incomprehensibilities to the smallest possible number.” (Huxley 1896, 165)”

amounts to another way to impose the same kind of constraints, namely on the explanans.

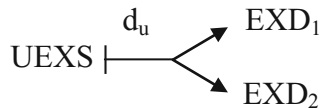


Diagram k'

Summing up, *when it comes to tackling the problem of understanding*, Kitcher doesn't have, at the end of the day (and by his own admission), a novel answer³⁷; he adopts Friedman's answer. Thus—and this is an important result—once one grasps this answer, one is in the possession of the gist of the standard unificationist account (Friedman-Kitcher) of the relation between explanation and understanding.

As I announced at the end of the previous section, I will argue that this standard unificationism is deficient in an important respect. Although bringing in the notion of global understanding is a major advance, unificationists shouldn't dismiss the local aspects of understanding. Local understanding is an intuitive, central component of the whole notion of understanding, and it would be utterly disappointing if the unificationists had nothing to say about it; if so, they would work with an impoverished notion. But there is good news for the advocates of unificationism. The doctrine doesn't have to be committed to the dismissal of locality, as it has the resources to say illuminating things about it. In the next section, I attempt to show precisely this.

7 How unificationism allows for gains in local understanding

We can now return to the issue I flagged up at the end of sect. 5, the standard unificationists' mishandling of the local aspects of understanding. As we recall, the difficulty cropped up naturally. Suppose we look at Diagram f, and wonder: is our understanding of *an individual phenomenon* (say, EXD₁) increased when we derive it from a unifier (UEXS)? Standard unificationists answer that 'it is not'—while also adding that 'it *need not be*', since the unificationists' concern is different, namely global understanding.

I now begin to argue that unificationism can do better than this. But, as I pointed out at the outset, to present my arguments I need to operate a revision of the unificationist framework as we have inherited it. This is needed because unification, when examined in historical perspective and in relation to scientific practice, functions quite differently than pictured by the sympathizers of the unificationism (recall especially Diagram f above). The examination of a couple of scientific examples (otherwise widely used in the literature) can help clarify what has been overlooked, and what modification is needed.

Consider the unifying Newtonian mechanical framework and its capacity to provide derivations of results obtained previously by Kepler and Galileo in celestial and terrestrial mechanics, respectively. We can take this framework to play the role of the

³⁷ The italicized clause is important, since I'm *not* claiming that Kitcher's theory is reducible to Friedman's. For reasons of space I can't discuss, as I warned above, Schurz and Lambert's (1994) version of unificationism, which, unlike Kitcher's, does take an account of understanding to be of paramount importance.

UEXS unifier. Also, consider Kepler's Third Law. It states that the square of the orbital period (p) of a planet is directly proportional to the cube of the semi-major axis (A) of its orbit. That is,

$$p^2/A^3 = \text{constant}$$

The phenomenon we deal with here is the constancy of this ratio and, as is well known, this constancy can be (easily) derived using Newtonian concepts. Yet, the crucial point to note is that what we derive within the Newtonian framework is a *conceptually enhanced* variant of this constancy claim, namely one in which the constant of proportionality is specified. Thus, the Newtonian variant of Kepler's Third Law is

$$p^2/A^3 = (4\pi^2)/(gM),$$

where g is the gravitational constant, and M is the mass of the Sun. (To be precise, M approximates the sum of the masses of the Sun and of the Earth).

This example is by no means unique. Here is another one, the predecessor of the 'ideal gas law', the so-called 'combined gas law' (combining Boyle's, Charles' and Gay-Lussac's laws.) It reads:

$$PV/T = \text{constant},$$

where P is the pressure of the gas, T is the absolute temperature, and V is the volume. Just as with Kepler's Third Law, we can show that this constancy holds within the Newtonian framework (to be exact: the kinetic theory), where we can derive the *conceptually enhanced* relation

$$PV/T = Nk_B,$$

where N is the number of particles of gas (interacting via Newtonian forces), and k_B is Boltzman's constant.³⁸

More examples like these can surely be found, and they help convey the following important, and general, point: within a unificationist framework, we are in the position to make *new, true claims* about the *individual phenomena* (regularities) under study. Moreover, the kind of gain just highlighted—an increase in the conceptual sophistication of the *description of the phenomenon* (EXD) under scrutiny—is often complemented by a related one, namely the ability to derive, within the unifying framework, *corrected* versions of the regularities discovered before. (A well-known example, among many others, is Galileo's laws of free fall, which are recovered in corrected form by Newton.) Thus, a unificationist can claim an important epistemic gain, namely an improvement in our capacity to provide *better descriptions of individual explananda*, both in terms of making corrections to their descriptions and by increasing the conceptual sophistication of their description. This is a gain of 'local'

³⁸ From ($PV/T = \text{constant}$), by taking Avogadro's law into account, one derives the ideal gas law ($PV/T = nR$), and this quantity is constant for a given amount n of moles of gas, since R is the universal gas constant. As a result of the derivation of the ideal gas law within the kinetic theory, we have $nR = Nk_B$.

nature—it regards one individual phenomenon/regularity, as disconnected from any other phenomenon—and thus it is a contribution to local understanding.

This leads us to the central modification to standard unificationism proposed here, which I will represent in Diagram f* below. EXD₁ and EXD₁* refer to the descriptions of the same phenomenon, before and after unification, respectively. The *-version is the conceptually enhanced/corrected one.

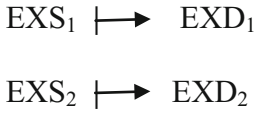


Diagram e

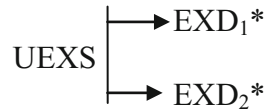


Diagram f*

The key-difference between standard unificationism (Diagram f) and the kind of unificationism I endorse here (Diagram f*) is that the *-version takes into account a certain aspect of unification overlooked by standard unificationism. This aspect (introduced above) is the existence of a kind of ‘conceptual enhancing’ effect, which occurs when (genuine) unification is accomplished.³⁹ We may begin by puzzling over a series of natural phenomena, apparently unrelated regularities described as EXD₁, EXD₂, etc. (see Diagram e). Then, *if* we are in the felicitous position to find a unifier UEXS (Diagram f), we realize that something interesting happens: this unifier doesn’t allow the derivation of EXD₁ per se, but of EXD₁*; not of EXD₂ exactly, but of EXD₂*, etc. That is, the unifier UEXS, once identified, does *not* leave the descriptions of the phenomena to be unified unchanged, but *unifies while re-describing (and/or correcting)*. Thus, the key-question to which the standard unificationist didn’t have an answer—how is our understanding of an individual phenomenon increased when we derive it from a unifier?—now receives one: (in terms of Diagram f*.) we can claim a better understanding of an individual explanandum because of our capacity to provide a conceptually enhanced description of it. To exemplify, we can say we understand the phenomenon of constancy of the ratio of the two characteristics of a planet’s trajectory *better* in the unified framework of Newtonian astronomy, since we now see how and why this constancy holds: it is due to the constancy of other elements of the astronomical system (the mass of the earth and the gravitational constant).

That this enhancement is possible is not surprising.⁴⁰ The unifying framework typically brings in additional conceptual resources (e.g., the gravitational constant), and, as we’ll see in more detail in the next section, this framework affords these resources because it postulates a different *ontology* than that of the EXDs. To anticipate: unlike Kepler’s, Newton’s ontology contains *forces*, gravity in particular. This is also

³⁹ This aspect is by no means new. It was signaled in several careful and detail-oriented philosophical analyses of unification and reduction in science. In addition to the recent reference to Morrison’s book (2000, 2–3), earlier work by Sklar (1993) and Schaffner (1967) (as well as by others, including Feyerabend and Kuhn), should be mentioned. I’m not aware of any place where either Friedman or Kitcher discuss this ‘enhancing’ aspect in their work.

⁴⁰ This is the more cautious version of the stronger claim I have in mind, that unification *always* leads to this enhancement—more precisely, as we’ll see, the *right kind* of unification, of the ontologically-reductive type. What follows are plausibility considerations in favor of the stronger claim, to be developed later on; the examples here are only illustrative.

evident in the second example above: our initial EXD ($PV/T = \text{constant}$) was captured by an expression ‘blind’ to the composite structure of matter, while the expression for EXD^* was $PV/T = Nk_B$, a conceptually enhanced one, since it takes into account (on the right-hand side) the corpuscular composition of the physical systems.

It should now be clear that despite agreeing with the overall standard unificationist picture, the revised kind of unificationism I’m articulating here rejects the dismissal of locality. On the contrary, this revised version does find a place for it in the general scheme of the doctrine. This inclusiveness, however, should not overshadow the crucial role of global understanding, the defining feature of unificationism (both standard and revised).

8 The conjunction problem

It is now time to tackle perhaps the central problem of unificationism, the so called ‘conjunction problem’—a well-known difficulty plaguing *any* attempt to flesh out a unificationist project ever since Hempel and Oppenheim’s (in)famous footnote 33 in their (1965, 273).⁴¹ In a nutshell, the problem is this: if unificationism rests in the end on claiming the reduction of the number of explanans, the question is how we count them, since “it is not at all clear what counts as *one* [EXS] and what counts as *two*.” (Friedman 1974, 16; italics in original) In our notation (Diagrams e and f), the difficulty for the unificationist is to say what would prevent us from considering the conjunction of two explananda (EXD_1 & EXD_2) as one explanans, re-label it as a unifier UEXS, and then derive each of the explananda from it—thus trivializing the unificationist idea. If we take EXD_1 to be Kepler’s Third Law ‘K’ ($p^2/A^3 = \text{constant}$), and EXD_2 to be Boyle’s Law ‘B’ ($PV = \text{constant}$, for constant T), we can recover the conundrum in the form Hempel and Oppenheim originally formulated it.

Friedman’s first attempt to spell out precisely his unificationism, and deal with the problem, was definition “ D_1 ” (1974, 17)⁴²—but, as the readers familiar with the intricacies of this literature will recall, that didn’t work well. Failing to come to terms with this problem, Friedman gave up the first definition and offered another one (D_1')—which, alas, turned out, once again, to have the absurd consequence that *no* sentence of the type Friedman generally took to be explanatory can explain, as Kitcher promptly pointed out (see Kitcher (1976, 211–2) for the details of D_1' and the formal

⁴¹ Salmon (in Kitcher and Salmon (1989, 95)) discusses it in relation to Friedman’s (1974) views, and it is pretty clear that Richard Feynman thought about it too, as Lange (2002) observes.

⁴² D_1 states that ‘ S_1 explains S_2 iff $S_2 \in \text{con}_\kappa(S_1)$ and S_1 reduces $\text{con}_\kappa(S_1)$ ’, and an interesting consequence of it, proved by Kitcher (1976, 209), is that only so-called ‘ κ -atomic’ sentences can explain. I shall explicate some of this formalism in the benefit of completeness, but I stress that these details are not needed to understand the subsequent discussion here. A sentence S is ‘ κ -atomic’ if it has no (κ)-partition, i.e., if there is no pair of sentences $\{A, B\}$ such that A and B are acceptable independently of S , and the conjunction $A \& B$ is logically equivalent to S . κ is the set of accepted law-like sentences by the scientific community at a certain point in time, and is deductively closed. The set ‘ $\text{con}_\kappa(S)$ ’ is the set of independently acceptable consequences of S . A given sentence S ‘reduces’ a set Δ iff the following relation holds: $\kappa\text{-card}(\Delta \cup \{S\}) < \kappa\text{-card}(\Delta)$. For a set of sentences M , the term ‘ $\kappa\text{-card}(M)$ ’ denotes the κ -cardinality of M . Friedman stipulates (naturally) that $\kappa\text{-card}(M) = \inf \text{card}(\Gamma)$, where Γ is a κ -partition of M . As above, a κ -partition of a set of sentences M is a set of κ -atomic sentences which is logically equivalent to M (it is assumed that such a κ -partition exists for every set M).

proof.) Thus, Friedman seemed to be jumping from the frying pan right into the fire, and it looked like this was the end of the road for the project. This impasse provided Kitcher with the motivation to articulate a new version of unificationism—which, however (as we saw in sect. 6), was affected by the lack of its own account of understanding, which led him to defer to Friedman’s flawed account.

At this point we need to see if a revised version of unificationism can do better with regard to explicating understanding—and this includes providing a solution to the conjunction problem.

We now know what the problem is, and how it impacts the issue of understanding: we shouldn’t accept the obviously odd claim that EXD_1 (for instance) is explained, and thus we *understand* why it occurs, as a result of the trivial derivation from EXD_1 & EXD_2 . Unlike Friedman’s, the revised version of unificationism I described here in Diagram f^* can deal with the problem, since this form of unificationism has built in, by design, the solution. In essence, the revised account is immune against trivialization because the key-manoeuver (forming the conjunction) fails to ensure that the unifier is a genuine one, or the kind of unifier complying with the requirements of Diagram f^* .

To see why this is so, we should begin by noting that the trivialization move is meant to guarantee that what is derived from the conjunctive unifier is *exactly* each of the conjuncts: we form $EXD_1 \& EXD_2$ to ensure that we derive EXD_1 (for instance). But we recall (from section 6) that genuine unification in science just doesn’t work like that: what we derive from a genuine unifier is not exactly the description of the phenomenon we wanted to explain initially, but an enhanced version of this description (even a corrected one), here called EXD_1^* . This happens in cases of genuine unification, but it *never* happens in cases of spurious-conjunctive unification. Because the standard version of unificationism overlooked this aspect—namely, the systematic enhancements prompted by unification—this version had to undertake the burden of dealing with the Hempel-Oppenheim conjunction problem. While the problem is indeed fatal for the standard unificationism captured by Diagram f , it simply doesn’t appear for the revised version of unificationism captured in Diagram f^* . To exemplify: if we conjoin K&B (i.e., ‘ $p^2/A^3 = \text{constant}$ ’ & ‘ $PV = \text{constant}$, for T constant’), this is intuitively not a unifier. And now we can say why: K&B entails K, not K^* (i.e., $p^2/A^3 = (4\pi^2)/(gM)$), the more precise version of Kepler’s Third Law, which the genuine unifier (the Newtonian framework, and its ontology of point-masses and forces) derives. Moreover, K&B can’t *in principle* entail K^* , as K^* involves conceptual resources—i.e., a (Newtonian) ontology (including gravitational forces and the gravitational constant g)—‘unknown’ to mere K&B. Let’s elaborate this point.

Keeping Diagram f^* in mind, one can now ask: can’t we now concoct the conjunction $EXD_1^* \& EXD_2^*$, and say that it is a unifier? To deal with this challenge we have to return to an important observation about the ontology of the unifying framework introduced briefly in sect. 7. Now we must expand on this observation, and express it more precisely: the trivializing, conjunctive unifier is bound to always be *ontologically conservative*—while a genuine unifier is not. ‘Conservativeness’ here is the idea that the ontology presupposed by the conjunction will *always* be the union of the ontologies assumed by the two conjuncts, and never less than that. No ontological reduction, i.e., reduction in the number of the *types* of entities, can take place when the mere logical

conjunction is employed to concoct a unifier. So, once we make a necessary condition for a genuine unifier to assume a sparser ontology than the ontology of the phenomena derived from it, we see that the simple conjunction EXD_1^* & EXD_2^* can't be a genuine unifier if the ontology assumed by it is precisely the ontology of EXD_1^* together with that of EXD_2^* —and not *less* than that, as required.

This major difference between a trivial-conjunctive unifier and a genuine one can be illustrated with Hempel's own example. Kepler's Third Law, in its exact form K^* , is about the relation between a planet and the Sun. The objects mentioned in Boyle's Law are gases; if we call the exact version of it B^* (and take it to be the Ideal Gas Law), then the objects mentioned in it are gas molecules. The conjunction of these two, K^* & B^* , is then bound to assume an ontology containing *all* these entities—i.e., the planets, the sun, and the gas molecules. Thus, K^* & B^* , if we look at it as a unifier, is ontologically conservative; by contrast, a genuine unifier is *not* ontologically conservative, but always works by decreasing the number of types of objects assumed in formulating an explanation - hence K^* & B^* does not qualify as a genuine unifier. To clarify further this crucial property of a genuine unifier, we should now consider such a genuine unifier, the Newtonian framework. The crucial remark is that within it, all phenomena, both celestial (motions of planets, comets, etc.) and terrestrial (apples falling, gases expanding, etc.) are on a par—ontologically speaking. No ontological difference between these phenomena is recognized, since *all there is*, according to the ontology assumed by this framework, is two categories, or types, of entities: *point-masses* and *forces* (where the latter accelerate the former).⁴³ We can of course label these point-masses in various ways—as 'apples', 'molecules', 'electrons', 'planets', 'the Sun', 'the Earth', etc.—and the forces acting on point-masses as 'gravitational', 'electric', and so on. Yet the ontological reduction at work here reveals that, strictly speaking, there is no Earth, no Sun, no apples and no molecules in the Newtonian ontology.

So, the ontology assumed at the level of the explanans (the Newtonian framework) contains only point-masses and forces; it is in terms of these objects (this ontology) that we formulate (the exact version of) Kepler's Third Law K^* , i.e., without any reference to the planets or the Sun, but by referring only to two point-masses interacting via a (specific) force (gravity). We then derive the relation K^* , and after the derivation is accomplished, we label the point-masses involved as 'planet' and 'the Sun'. Thus, before this labeling (i.e., when we work out the derivation of K^*), both the explanans (the 'premises'—as it were) and the explanandum (the 'conclusion') are formulated in terms of the same ontology: we begin with two point-masses and a force (obeying a certain law), and derive a relation holding between these two point-masses. The appropriate labeling of the point-masses in the explanandum (as 'planet' and 'the Sun') allows the explanandum to be read in the usual way later on, as specifying features of the trajectory of a planet around the Sun. The very same procedure applies when we derive the laws of kinetic theory (such as B^*) within the Newtonian framework. We label the point-masses as 'molecules', and then the collection of molecules (i.e., the collection of point-masses) as a 'gas'. In both cases, the ontology of the explanans is massively reduced (contains only point-masses and forces), while

⁴³ There is of course more, e.g., absolute space and absolute time, but these are irrelevant here.

the ontology of the explananda is more diverse (contains planets, stars, molecules, gases, gravitation, etc.)

It is then this ontological-reductive requirement that is crucial to genuine explanatory unification and which has been overlooked by the traditional unificationism; in addition, this feature helps solving the conjunction troubles. But, one may now ask, isn't the price we pay for this solution too high? For, is this kind of explanatory unification ever illustrated in science? The answer is affirmative—see the Newtonian framework just examined—and yet one should also acknowledge that such illustrations, while well-known, are not common. Thus, this account of explanation may seem to have the disadvantage that it makes formulating scientific explanations *that provide (both local and global) objective understanding* an extremely difficult task. But, why should one assume that such a task—especially given the italicized requirements—should be easy in the first place?! On the conception of explanation advocated here, to (genuinely) explain is an epochal intellectual achievement, and thus it's just not surprising that it occurs rather rarely—although, granted, there seems to be plenty of explanations in science. Yet they are 'local', of the metaphysically dubious 'causal' type, and they, as we recall, can be said to only marginally enhance our understanding, as they just shuffle 'mysteries' around. Although truly understanding-producing explanations are not that common indeed, another important historical case—Einstein's unification of electricity and magnetism in the Special Theory of Relativity—can be invoked here. This unification consists in essence in demonstrating that the occurrence of electric and magnetic phenomena is merely an observer-relative effect of switching between reference frames; these phenomena turn out to be manifestations of one underlying entity, the electromagnetic field. Here, again, the ontology of the explanans contains one entity (the electromagnetic field), while the ontology of the explananda recognizes, and thus labels, more (two), upon switching frames. As a contrast, when this ontological deflation doesn't take place, unification is not genuine—and thus the danger of conjunctive trivialization is always present. The version of unificationism endorsed here preempts this danger: it urges that unification should be conceived not as a move from Diagram e to Diagram f, but to Diagram f* while performing ontological reduction.

Finally, note that this last example illustrates what some authors call 'perfect' unifications (Maudlin 1996), or 'reductive' unifications (Morrison 2000). It should be distinguished from other, less perfect cases, which lack the defining feature of reductive unification—ontological reduction/simplification. Consider, again, Maxwell's well-known achievement, oftentimes also described as 'the unification of electricity and magnetism'. It is an underappreciated fact that, as it turns out, this unification falls short of the reductive perfection displayed above by Newton's and Einstein's. All that Maxwell's equations give us is a nomological connection among physical phenomena of apparently different nature: electrical currents create magnetization, and vice versa. This shows that there is some truth to the unification claim, since we can't disentangle these phenomena. Yet, as Maudlin puts it "[T]he electric and magnetic fields retain a completely distinct ontological status in Maxwell's [theory]. They may be nomically correlated, they may give rise to one another, but at base they are still entirely different entities." (1996, 131) What we have here is unification, but a weaker form of it: not ontological reduction (simplification), but rather unification-cum-synthesis—hence Morrison's (2000) introduction of the notion of 'synthetic unity'. (This point is also

in agreement with Redhead's (1984) illuminating account of the senses of unification in physics.)⁴⁴

9 A revised version of unificationism

All elements are now in place to re-cast schema E, and condition (i+)₄, into a revised version of unificationism, as follows (see Diagram f* again):

A phenomenon is explained iff:

- (i) an *ontologically-reductive unifier* explanans is identified, and
- (ii) the way in which this explanans ensures the phenomenon arises is presented.⁴⁵

This ontological-reductive version of explanatory unificationism has been designed to deal with the problem(s) that plagued its ancestors, in particular the conjunction problem. It was also designed to warrant an increase in objective understanding of the global kind, while also allowing for gains in objective understanding at the local level. We can now finally return to the relation between unificationism (in this revised form) and causalism.

As I said at the beginning, I take it to be universally agreed that presenting a theory of causation is a formidable task, hence the suspicion that it will never be accomplished. Thus, under this pessimistic assumption, a causal approach to explanation can't even get off the ground; without this theory, we don't quite know what (i+)₃ actually means, so there's no way to add it to schema E. Thus, we can see one sense in which the unificationist approach is to be preferred to the causal approach: the former is a safer bet (so to speak), as it can be spelled out independently of such an achievement.

However, as I said at the end of section 4, I don't discount the optimistic scenario that a theory of causation *can* be worked out. So, let's assume we know how to spell out (i+)₃, i.e., that the notion of cause is successfully captured by a metaphysical theory of causation: 'cause' will then stand for 'cause as explicated by ... [insert the true theory

⁴⁴ A consequence of all this is that Maxwell's unificatory achievement mentioned above should not, on the view I propose here, count as genuinely explanatory. I agree with the referee who raised this issue that this sounds a bit counterintuitive, but I'm prepared to bite the bullet and point out that there is a sense in which the discovery of the nomological connections entangling electricity and magnetism only partially increased our understanding of the world—rather, it raised another question, as to *why* are they so entangled—to be answered later, by Einstein's STR. However, the other, related, major Maxwellian idea, that light *is* (a form of) electromagnetic radiation, does count as explanatory, as it should, since it involves ontological reduction. The same referee has also raised the issue of other cases traditionally described as unifications (Mendelian genetics, Darwinian evolutionary theory, the theory of the chemical bond, etc.), and the verdict to be given is similar: the central question to ask when analyzing them (task which unfortunately can't be carried out here) is whether ontological reduction is present. I suspect it is, in most of these cases; but if it is not, then, although they are still recognized as unifications, the view I support here will not deem them explanatory (despite the impression to the contrary).

⁴⁵ As is immediate, the phenomenon in question here is in fact a member of a larger class of phenomena, each of them derivable from the same unifier.

of causation here]’. Under this assumption, schema E and (i+)₃ amount to the following causal account:

A phenomenon is explained iff:

- (i) its explanans—its *cause*—is identified, and
- (ii) the causal dependence of the phenomenon (‘the effect’) from its cause is demonstrated.

It is now important to realize that the unificationist account is still more appealing. This is so because of the dual nature of understanding, as local + global. If we conceive of understanding as this composite, then its maximization becomes a desideratum for any account of scientific explanation: the best account would be the one that delivers a maximum of understanding (so to speak). Thus, the trouble with the causal account above comes from the fact that it can give us *only* local understanding. Explanations, on this account, are cause-effect pairs, and this makes it *essentially* a local account. Showing that an individual phenomenon holds because its cause ‘produced’ it, increases local understanding of that phenomenon indeed, but does nothing to address the need to increase global understanding. The total number of mysteries to confront overall remains the same, as we recall.⁴⁶ Yet, within the optimistic scenario, it’s not all bad news for the causalist, and this thanks to the existence of the unificationist framework sketched above. A hybrid account suggests itself, taking the following form:

A phenomenon is explained iff:

- (i) its explanans, i.e. its *cause*, is identified, and it is an *ontologically-reductive unifier*, and
- (ii) the causal dependence of the phenomenon from the explanans is demonstrated.

As is immediate, this account addresses both types of understanding, global and local (its representation would still be Diagram f*, in which the arrows, and the relations d, have a causal interpretation.) Thus, the unificationist framework can naturally subsume the causal idea—and rescue it. Moreover, a virtue of this unificationism is that it does this regardless of which specific theory of causation happens to be adopted.

Finally, we see that by rescuing (through incorporation) the causal account, this revised unificationism can now deal with another family of problems unificationism traditionally faced—the asymmetry problems of the flagpole/shadow type mentioned above⁴⁷—since the causal component of the account will take care of them. Note, on the other hand, that if a satisfactory theory of causation can not developed—i.e., we can’t tell what ‘A causes B’ means—then this revised unificationism can’t incorporate a causal account (since it doesn’t exist!), and thus seems incapable to deal with these

⁴⁶ And this is of course Friedman’s (1974, 19) worry: “How our total understanding of the world is increased by replacing one puzzling phenomenon with another”?

⁴⁷ As Gijsbers (2007, 489–491) cogently argues, such problems continue to plague standard unificationism.

problems. But this is not an issue anymore, as the problems themselves dissolve: if ‘A causes B’ is problematic, so is an instance of it, the asymmetry ‘the flagpole causes the shadow, and not the other way around’ that generated the challenge in the first place.

To close. This importation of the causal element into the unificationist account suggests a win-win denouement of the debate between causalism and unificationism in the philosophy of scientific explanation. Thus, this revised unificationism is consistent with the ecumenism proposed, albeit rarely, in the literature (see Salmon (1990, 1998)). While I’m not denying that the view I advocate here is ecumenical, I must stress that it is also *hierarchical*: the unification approach subsumes the causal approach, and not the other way around (again, under the important proviso that such an approach can be presented). Hence, in the end, it is unificationism, not causalism, that has priority, since (this revised form of) unificationism is capable to offer the general framework within which the best possible account of scientific explanation— one maximizing understanding—is to be articulated.⁴⁸

Acknowledgments I thank the two anonymous referees of this paper for really insightful suggestions and criticism. I am of course responsible for the final version of the paper.

References

- Ahmed, A. (2014). *Evidence, decision and causality*. Cambridge Univ. Press.
- Barnes, E. (1992). Explanatory unification and scientific understanding. In D. Hull, M. Forbes, & K. Okruhlik (Eds.), *PSA 1992* (Vol. 1, pp. 3–12). East Lansing: Philosophy of Science Association.
- Bartelborth, T. (2002). Explanatory unification. *Synthese*, *130*, 91–107.
- Batterman, R. W. (2002) *The devil in the details*. Oxford Univ. Press
- De Regt, H. W. (2004). Discussion note: Making Sense of Understanding. *Philosophy of Science*, *71*: 98–109.
- DeRegt, H. W. 2006. Wesley Salmon’s complementarity thesis: causalism and unificationism reconciled? *International Studies in he Philosophy of Science* *20*, 129–147.
- De Regt, H. W., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, *144*, 137–170.
- Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, *71*, 5–19.
- Gijsbers, V. (2007). Why unification is neither necessary nor sufficient for explanation. *Philosophy of Science*, *74*, 481–500.
- Gijsbers, V. (2013). Understanding, explanation, and unification. *Studies in History and Philosophy of Science*, *44*, 516–522.
- Grimm, S. (2013). Understanding. In D. Pritchard & S. Berneker (Eds.), *The Routledge companion to epistemology*. Routledge.
- Grimm, S. (2014). Understanding as knowledge of causes. In A. Fairweather (Ed.), *Virtue scientia: Essays in philosophy of science and virtue epistemology*. Special Issue of *Synthese*.
- Heisenberg, W. (1971). *Physics and beyond. Encounters and conversations* (translation from German by A. J. Pomerans). New York: Harper and Row.
- Hempel, C. (1965). *Aspects of scientific explanation*. New York: Free Press.
- Hindricks, F. (2013). Explanation, understanding, and unrealistic models. *Studies in History and Philosophy of Science*, *44*(3), 523–531.
- Huxley, T. H. (1896). *Darwiniana*. New York: Appleton.

⁴⁸ I should emphasize that Salmon’s ecumenism is based on the belief that the causal approach and unificationism are on a par, while I believe that they are not. Moreover, although I suppose that Strevens’ (2004) *kairctic* account can be called ecumenical, it explicitly prioritizes causation. Furthermore, what I’m saying here is also different from Kitcher’s previously mentioned ecumenistically-inclined idea that “the ‘because’ of causation is always derivative from the ‘because’ of explanation.” (1989, 477)

- Jones, T. (1995a). How the unification theory of explanation escapes asymmetry problems. *Erkenntnis*, *43*, 229–40.
- Jones, T. (1995b). Reductionism and the unification theory of explanation. *Philosophy of Science*, *61*, 21–30.
- Jones, T. (2012). Unification. In S. Psillos & M. Curd (Eds.), *The Routledge companion to the philosophy of science* (2nd ed., pp. 97–106). New York: Routledge.
- Khalifa, K. (2013). The role of explanation in understanding. *British Journal for the Philosophy of Science*, *64*(1), 161–187.
- Khalifa, K., & Gadowski, M. (2013). Understanding as explanatory knowledge: the case of bjorken scaling. *Studies in History and Philosophy of Science*, *44*, 384–392.
- Kim, J. (1994). Explanatory knowledge and metaphysical dependence. *Philosophical Issues, Truth and Rationality*, *5*, 51–69.
- Kitcher, P. (1976). Explanation, conjunction, and unification. *Journal of Philosophy*, *73*, 207–12.
- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, *48*, 507–31.
- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.
- Kitcher, P., & Salmon, W. (Eds.). (1989). *Scientific explanation*, Minnesota studies in the philosophy of science, vol. 13, Minneapolis: Univ. of Minnesota Press.
- Lange, M. (2002). *An introduction to the philosophy of physics: Locality, fields, energy, and mass*. Blackwell.
- Mäki, U. (2001). Explanatory unification: double and doubtful. *Philosophy of the Social Sciences*, *31*, 488–506.
- Maudlin, T. (1996). On the unification of physics. *Journal of Philosophy*, *93*, 129–44.
- Morrison, M. (2000). *Unifying scientific theories. Physical concepts and mathematical structures*. Cambridge: Cambridge University Press.
- Newman, M. (2013). Refining the inferential model of scientific Understanding. *International Studies in the Philosophy of Science*, *27*, 173–197.
- Quine, W. V. (1976). *The ways of paradox and other essays* (Revised and enlarged ed.) Cambridge: Harvard University Press.
- Redhead, M. (1984). Unification in science. Review of *The Unity of Nature* by C. F. von Weizsäcker. *British Journal for the Philosophy of Science*, *35*(3), 274–9.
- Roland, J. (2008). Kitcher and the obsessive unifier. *Philosophy and Phenomenological Research*, *2*, 493–506.
- Russell, B. (1912). On the notion of cause. *Proceedings of the Aristotelian Society*, *13*, 1–26. New Series.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- Salmon, W. (1990). Scientific explanation: Causation and unification. *Critica*, *22*(66): 3–21.
- Salmon, W. (1998). *Causality and explanation*. Oxford: Oxford University Press.
- Schaffner, K. (1967). Approaches to reduction. *Philosophy of Science*, *34*(2), 137–147.
- Schurz, G. (1999). Explanation as unification. *Synthese*, *120*, 95–114.
- Schurz, G., & Lambert, K. (1994). Outline of a theory of scientific understanding. *Synthese* *101*, 65–120.
- Sklar, L. (1993). *Physics and chance*. Oxford Univ. Press.
- Strevens, M. (2004). The causal and unification approaches to explanation unified—causally. *Noûs*, *38*(1), 154–76.
- Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A*.
- Toulmin, S. (1961). *Foresight and understanding*. London: Hutchinson.
- Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, *69*, 212–33.
- Van Fraassen, B. (1985). Salmon on explanation. *Journal of Philosophy*, *82*, 639–651.
- Weber, E. (1999). Unification: what is it, how do we reach and why do we want it? *Synthese*, *118*, 479–99.
- Woodward, J. (2003). *Making things happen. A theory of causal explanation*. New York: Oxford Univ. Press.
- Woodward, J. (2011). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2011 Edition). URL = <http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/>.