

# Why Unification Is Neither Necessary Nor Sufficient for Explanation\*

Victor Gijsbers<sup>†</sup>

---

In this paper, I argue that unification is neither necessary nor sufficient for explanation. Focusing on the versions of the unificationist theory of explanation of Kitcher and of Schurz and Lambert, I establish three theses. First, Kitcher's criterion of unification is vitiated by the fact that it entails that every proposition can be explained by itself, a flaw that it is unable to overcome. Second, because neither Kitcher's theory nor that of Schurz and Lambert can solve the problems of asymmetry and accidental generalizations, it follows that unification is not sufficient to ground explanation. Third, some good explanations are disunifying, which entails that unification is not necessary for explanation either.

---

**1. Introduction.** Much recent literature on scientific explanation (Kitcher 1985; Woodward 2003; Strevens 2004) states that there are two main philosophical theories of explanation. The first is the causal theory, associated with the work of Salmon (see especially Salmon 1984). The second is the unificationist theory, first proposed by Friedman (1974) and defended in radically revised form by Kitcher (1981, 1989) and by Schurz and Lambert (1994; Schurz 1999). In this paper I examine whether unification is indeed a concept which can ground explanation. This examination will have two parts: first, I will evaluate whether unification is sufficient for explanation; second, whether it is necessary. Both Kitcher's theory, which is by far the best-known theory of unification, and that of Schurz and Lambert will be considered. My conclusion is that unification is neither sufficient nor necessary for explanation.

In section 2, I review the two versions of unificationism. I argue that Kitcher's theory entails that every proposition explains itself, and that his proposed solution to this problem does not work. This problem, if not

\*Received October 2005; revised November 2006

<sup>†</sup>To contact the author, please write to: Faculty of Philosophy, University of Leiden, Postbus 9515, Leiden, 2300 RA, Netherlands; e-mail: V.Gijsbers@let.leidenuniv.nl.

Philosophy of Science, 74 (October 2007): 481–500. 0031-8248/2007/7404-0003\$10.00  
Copyright 2007 by the Philosophy of Science Association. All rights reserved.

solved, is fatal. The theory of Schurz and Lambert does not suffer from this flaw, and is therefore more promising.

I turn in section 3 to the sufficiency of unification. I argue that Kitcher's theory cannot generate the time asymmetry of causal explanation, and is thus unable to solve the so-called problem of asymmetry. Schurz and Lambert explicitly add causal principles to their theory, so for them the question of the derivability of causality from unification does not arise. In addition, I argue that neither of the two theories of unification is able to draw a distinction between a class of explanations and a class of non-explanations that are traditionally separated by means of the distinction between laws and accidental generalizations. I thus show that unification alone is insufficient to decide whether something is or is not a genuine explanation.

In section 4, the link between unification and explanation is analyzed in greater detail. Pace Schurz, I defend the thesis that one can explain 'surprising' events on the basis of equally or even more 'surprising' ones. This will show that unification is not a necessary condition for explanation. As I shall argue, unificatory power is relevant to explanation only as far as it serves as a reason for belief, but this is a much weaker connection than the one postulated by unificationism.

Finally, section 5 summarizes the conclusions.

**2. Two Types of Unificationism.** Both Kitcher's theory and that of Schurz and Lambert are complex. I will summarise them in a few pages, so inevitably some of the conceptual and formal machinery will not be touched upon.

*2.1. Kitcher's Theory.* The best-known unificationist theory of explanation is that of Kitcher (1981, 1989). He starts out from the set of all scientific knowledge,  $K$ , and develops criteria for its best systematization, which he calls the explanatory store over  $K$ ,  $E(K)$ . Kitcher defines explanatory patterns, sets of which are called generating sets. A generating set, when applied to  $K$ , generates a set of arguments: namely, all the instantiations of the explanatory patterns in the generating set that are acceptable in  $K$ . Kitcher then gives (incomplete) criteria for the unifying power of a generating set. Intuitively, a generating set is more unifying if it generates many conclusions from few patterns; and also if the patterns it uses are stringent, and not catch-all patterns that can be used to derive almost anything. Kitcher defines the *conclusion set*  $C(D)$  of a set of derivations  $D$  as the set of all statements that occur as a conclusion of at least one member of  $D$ . The *unifying power* of a complete generating set for  $D$  varies directly with the size of  $C(D)$ , directly with the stringency of the patterns in the set, and inversely with the number of patterns in the

set. The relative weight of these three criteria is intentionally left unspecified.

Kitcher claims that this theory suffices to characterize acceptable explanations: it is a total theory of explanation, giving both necessary and sufficient conditions. As part of this claim, Kitcher also says that he can generate the notions of causality and lawhood from the unificationist theory of explanation; and that he can thereby solve the asymmetry problem of causal explanation. I will examine these claims in sections 3 and 4.

However, Kitcher must first solve a problem which his theory faces. I address it in the next subsection.

*2.2. The Problem of Spurious Unification.* The problem of spurious unification is recognised in Kitcher 1981, 526–529, where Kitcher also attempts to solve it.<sup>1</sup> The problem is this: given Kitcher's theory, it appears to be the case that a unificationist is committed to the view that every fact  $F$  is explained by a derivation from  $F$  itself. The reasoning is as follows. Let us take only a single argument pattern:

$$\frac{\alpha}{\alpha}, \quad (1)$$

where the substitution instructions tell us to put an accepted scientific statement in place of  $\alpha$ . How unifying is this tiny generating set? The number of patterns it contains is minimal and the number of conclusions it generates is maximal, but the single pattern is not stringent at all. We may conclude that this pattern has little, if any, unifying power. This is good for unificationism, as we would be loath to accept that self-explanation is a universally valid type of explanation.

However, as Kitcher himself points out (1981, 527), there is a procedure that creates, for any generating set  $G$ , a generating set  $G'$  that contains only self-derivations but is just as unifying as  $G$ . Take a single argument pattern,  $A$ , from  $G$ .  $A$  generates a set of arguments, which has a set of conclusions  $C(A)$ . We construct an argument pattern  $A'$  that is at least as stringent as  $A$  and has the same set of conclusions. Argument pattern  $A'$  has the form (1), where the substitution instructions tell us to put a sentence  $p$  in place of  $\alpha$  that conforms to the rule  $p \in C(A)$ . Evidently,  $C(A') = C(A)$ . And because each member of  $C(A)$  is the result of a different substitution of terms for the dummy letters in  $A$ , the substitution instructions of  $A$  allow at least as many substitutions as those of  $A'$ . So

1. It is not discussed in Kitcher 1989, even though his theory as expounded there is just as vulnerable to it.

$A'$  is at least as stringent as  $A$ . We repeat this procedure for each argument pattern in  $G$  and together these patterns form  $G'$ , a generating set that is just as unifying as  $G$ , but generates only self-explanations. Hence everything is explained by itself. This is an unacceptable consequence of a theory of explanation. If the problem of spurious unification cannot be solved, it is fatal for Kitcher's unificationism.

In order to solve it, Kitcher introduces a requirement that I will call  $R$ :

If the substitution instructions associated with a pattern  $P$  could be replaced by different substitution instructions, allowing for the substitution of a class of expressions of the same syntactic category, to yield a pattern  $P'$  and if  $P'$  would allow the derivation of *any* sentence, then the unification achieved by  $P$  is spurious. (1981, 527–528)

What motivates this requirement? Kitcher writes:

Why should patterns whose substitution instructions can be modified to accommodate any sentence be suspect? The answer is that, in such patterns, the nonlogical vocabulary that remains is idling. The presence of that nonlogical vocabulary imposes no constraints on the expressions we can substitute for the dummy symbols, so that, beyond the specification that a place be filled by expressions of a particular syntactic category, the structure we impose by means of substitution instructions is quite incidental. Thus the patterns in question do not genuinely reflect our beliefs. (1981, 528)

The patterns  $A'$  do not conform to requirement  $R$ , as changing the substitution instruction to 'put any sentence whatsoever in place of  $\alpha$ ' allows us to derive any sentence whatsoever. Requirement  $R$  thus gets rid of this example of spurious unification. But is Kitcher's reply successful in general? I will argue that it is not.  $R$  is both too strong and not strong enough: it banishes some patterns that we need to keep, but does not bar all forms of spurious unification.

What patterns are excluded by requirement  $R$ ? Those that can yield any sentence whatsoever if the dummy letters can be replaced by anything. This is just the class of arguments that have a dummy letter as their final conclusion. For suppose that the conclusion also contains elements that are not dummy letters. Then these will be present in all possible instantiated conclusions, which means that sentences that do not contain these elements cannot be derived.

This raises two questions: Are all derivations with a dummy letter as their final conclusion spurious explanations? And are all spuriously unifying argument patterns of this form? I will argue for a negative answer to both questions. A negative answer to the first question means that  $R$

is too strong, whereas a negative answer to the second means that  $R$  is not strong enough.

We take the first question first. Some logical derivations are barred by criterion  $R$ . For example:

$$\frac{\alpha}{\frac{\alpha \rightarrow \beta}{\beta}}, \quad (2)$$

where the substitution instructions tell us to put an accepted sentence in place of  $\alpha$  and any sentence in place of  $\beta$  such that  $\alpha \rightarrow \beta$  is an accepted sentence. According to Kitcher's criterion  $R$ , this derivation cannot be explanatory. Relaxing the substitution instructions completely—as any test of criterion  $R$  demands us to do—will also remove the need to ensure that  $\alpha$  and  $\alpha \rightarrow \beta$  are *accepted* sentences, since that need was encoded in the substitution instructions; and with that need removed, we can put any sentence we like in place of  $\beta$ . But logical derivations *can* be explanatory. “Why is this rose red?” “Well, you know that it was planted by John?” “Yes, I figured that out.” “And you know that John plants only red roses, right?” “Ah yes, I see—I really should have been able to make that inference myself.” (This explanation works even though all non-logical vocabulary is ‘idling’.)

Let us look more closely at the role of logic in Kitcher's theory.  $K$  is a deductively closed set of statements, so if  $p$  and  $q$  are members of  $K$ , then  $p \wedge q$  is also a member of  $K$ . Now surely, if we can explain  $p$  and we can explain  $q$ , we can also explain  $p \wedge q$ . This holds for all (deductive) logical derivations: if we can explain the premises, we can explain the conclusion. So Kitcher's theory should imply that the set of explainable sentences,  $C(D)$ , is closed under logical deduction. There are three ways of getting this result from the theory, but they are all problematic.

1. We can add every valid deductive inference to the generating set as a new argument pattern. This strategy will leave us with an infinity of argument patterns, and hence every generating set will be completely non-unified. In addition, when we apply requirement  $R$ , some of these patterns will be rejected. Deductive closure of  $C(D)$  cannot be guaranteed.
2. Alternately, we can add a single argument pattern,  $LD$  (for ‘Logical Derivation’), that has the form ‘ $\alpha$ , therefore  $\beta$ ’. The substitution instructions tell us to replace  $\alpha$  with any set of accepted conclusions from  $E(K)$ , and  $\beta$  with some proposition that deductively follows from this set. In this way,  $C(D)$  is deductively closed and there are still only finitely many argument patterns. Unfortunately,  $LD$  falls

prey to requirement  $R$ , because relaxing the substitution instructions completely allows us to derive any sentence whatsoever.

3. Finally, we can choose to add every valid deductive inference to the generating set as a new argument pattern; but change the criteria of unification so that deductive inferences no longer count towards the number of patterns. They are ‘free’, so to speak. However, this choice has the consequence that we will always achieve the greatest unifying power by using only deductive inferences as argument patterns, for instance, only self-explanations.

Requirement  $R$  is not as harmless as it seemed: when it is combined with the claim that if we can explain a set of sentences, we can also explain every logical consequence of that set, it follows that every generating set  $G$  contains an infinite number of patterns. Requirement  $R$  cannot be accepted by unificationists, as it would make unification impossible.

I will now show that requirement  $R$  does not eliminate all spurious unification. The demonstration is easy. Let  $A$  be a pattern in  $G$  that does not fall prey to requirement  $R$ . This means that its conclusion is not a dummy letter but has additional structure, like ‘ $\alpha \rightarrow \beta$ ’, or ‘ $\alpha$  is bigger than the moon’. The set  $C(A)$  contains all the conclusions that are generated by  $A$  when  $G$  is applied to  $K$ . We can now construct a new pattern  $A'$  that is at least as stringent as  $A$ , which generates the same conclusions, which is not rejected by requirement  $R$ , but which is nevertheless spurious. For example:

$$\frac{\alpha \text{ is bigger than the moon}}{\alpha \text{ is bigger than the moon}}, \quad (3)$$

with the substitution instruction ‘choose an object for  $\alpha$  such that “ $\alpha$  is bigger than the moon”  $\in C(A)$ ’. Evidently, this pattern cannot generate every sentence, no matter how far the substitution instructions are relaxed; it passes the test of requirement  $R$ . But it gives only spurious unification. If we repeat this procedure for every argument pattern in  $G$ , we will get a  $G'$  that is at least as unified as  $G$ , and yet contains only self-derivations. Requirement  $R$  is not powerful enough to solve the problem of spurious unification. This completes my demonstration that Kitcher has not solved the problem of spurious unification.

I wish to look briefly at one way in which the problem of spurious unification can be avoided by unificationists who do not accept Kitcher’s theory. Let  $G$  be the generating set such that it contains as few patterns as possible, that are as stringent as possible, yet that generate as many conclusions from  $K$  as possible with as small a deductive basis of facts from  $K$  as possible. The idea is to derive a lot of conclusions from a relatively small number of premises. Self-derivations are not unifying pat-

terns in this theory, since they do not generate any conclusions that have not been taken as premises. With self-derivations, you cannot derive many conclusions from few premises. So by adopting a theory along these lines, one can avoid the problem of spurious unification. This possibility is explored by Schurz and Lambert.

2.3. *Schurz and Lambert's Theory.* Intuitively, unification is reduction of the number of underived facts. In the approach of Schurz and Lambert (1994; Schurz 1999), a corpus of knowledge is unified by connecting its individual elements through 'arguments in the broad sense', keeping as few basic facts as possible. Their notion of unification is defined in the context of a theory of understanding (and explanation). I will first briefly survey their account of understanding and then go on to sketch their analysis of unification. I will also indicate how their theory avoids spurious unification.

We start from the corpus of knowledge of the epistemic subject (an individual or a community). This cognitive corpus  $C$  is an ordered pair,  $\langle K, I \rangle$ , where  $K$  is a relevant representation of the set of sentences that the subject believes (*KNOW*) and  $I$  is the set of 'arguments in the broad sense' (these include deductive, inductive, and probabilistic arguments) that he or she has mastered. That  $K$  is a relevant representation of *KNOW* means that it contains only *KNOW*'s *relevant elements*, which correspond to *basic phenomena*. These elements can be extracted from *KNOW* using the notion of a 'relevant conclusion' explicated in Schurz 1991. The effect is that  $K$  may contain  $P$  and  $Q$ , but not  $P \wedge Q$ ; that if  $K$  contains  $\forall x : F(x) \rightarrow G(x)$ , then it will not contain  $\forall x : F(x) \wedge H(x) \rightarrow Gx$ ; and so forth. *KNOW* is, as it were, represented by its logical atoms.

An answer  $A$  to a question 'Why  $P$ ?' can contribute understanding of  $P$  to  $C$  only if it shows how  $P$  fits into  $C$ . It must include the claim that there is an argument in the broad sense  $I_p$  that connects  $P$  to other elements of  $C$ . An argument can do this either by having elements of  $C$  among its premises and  $P$  as the conclusion, or by having  $P$  among the premises and some element of  $C$  as the conclusion. In addition,  $A$  must make  $C$  more unified. That is,  $\langle K + P, I + I_p \rangle$  must be more unified than  $\langle K, I \rangle$ .

Unification is 'coherence minus circularity'. Connecting statements in  $K$  by arguments in  $I$  increases coherence; but circular connections do not increase unification, since circular 'explanations' do not yield understanding. Formally, unification is defined as follows.  $K$  consists of two parts: the set of *basic phenomena*  $K_b$ , and the set of *assimilated phenomena*  $K_a$ . A *basis* of  $K$  is any subset  $K'$  of  $K$  such that every element of  $K$  not in  $K'$  can be inferred from elements of  $K'$  using arguments in  $I$ . The *unification basis* of  $K$  is that basis of  $K$  that yields the greatest unification of  $K$ ,

according to criteria explained &forall;.  $K_b$  is the unification basis of  $K$ ;  $K_a$  is  $K - K_b$ .

Every element of  $K$  is assigned a value, which is negative or positive depending on whether it is a datum or a hypothesis, and on whether it is in  $K_a$  or in  $K_b$ . An experimental datum in  $K_b$  has value zero: new data neither increase nor decrease unification. An experimental datum in  $K_a$  has a positive value: assimilating data by inferring them from the unification basis is exactly what scientific unification amounts to. A hypothesis in  $K_b$  has a negative value: adding new theories to  $K$  decreases unification, unless a significant amount of data from  $K_b$  is moved to  $K_a$  as a result. A hypothesis in  $K_a$  has zero value: as a consequence of more fundamental hypotheses it has already been paid for. The exact values are not defined by Schurz and Lambert, who view unification as a comparative concept (1994, 78). But the following two conditions do obtain. First, adding a theoretical statement to  $K_b$  costs more than transferring a datum from  $K_b$  to  $K_a$  yields: it is disunifying to think up a theory that explains only one datum. Second, complex theoretical statements cost more than simple ones.

An argument  $A$  can add elements to  $K_b$  or  $K_a$ , take them away or move elements from  $K_b$  to  $K_a$  or *vice versa*. If the sum total value of all these changes is positive,  $A$  is unifying; if it is negative,  $A$  is disunifying. It may not always be possible to find out whether  $A$  has a positive or a negative effect, as the criteria of Schurz and Lambert define only a partial ordering.

Schurz and Lambert's theory is immune to the problem of self-explanations that haunted Kitcher's proposal. Since these argument patterns do not decrease the number of phenomena in  $K_b$ , they are not unifying on their criteria. Only relevant inferences that decrease the set of basic phenomena or increase the set of assimilated phenomena count as unificatory. Thus, their theory is more promising than Kitcher's—as a theory of unification. Whether either of the two is successful as a theory of explanation will be the question I address in the rest of this paper.

**3. Causality and Lawhood.** In this section, we will consider whether the concept of unification is *sufficient* for grounding the concept of explanation, leaving the question of its *necessity* to section 4. My arguments that unification is not sufficient for explanation will have to do with the concepts of causality and lawhood. These have been introduced into the theory of explanation to make distinctions between certain classes of explanations and of non-explanations. If unification is to be sufficient for grounding explanation, it must be able to make these same distinctions, either by grounding the concepts of lawhood and causality themselves, or in some other way. I will show that this is not the case.

Casuality and lawhood are natural starting places for investigating the



sufficiency of unification as a ground for explanation. It is often claimed that causes explain their effects. Some theories of explanation, such as Salmon's (1984), even postulate that causality is *the* essential ingredient of explanation. It is also often claimed that laws of nature explain their instances. The theory of Hempel and Oppenheim (1948) assumes that all explanations must use a law of nature; from a very different perspective, Armstrong (1991) and Dretske (1977) argue that laws explain their instances in ways that mere regularities do not.

What I have to show is that the concepts of causality and lawhood allow us to distinguish between explanations and non-explanation, which unificationists cannot keep apart. In subsections 3.1 and 3.2, I will analyse Kitcher's attempt to generate causality and lawhood from his unificationist theory of explanation. I will argue that he fails. In subsection 3.3, we take a brief look at the possibility of getting these notions from the theory of Schurz and Lambert, and conclude that they do not succeed either. The conclusion is that unification is not sufficient for explanation.

*3.1. Kitcher and Causal Asymmetry.* One of the most pressing problems that beset traditional accounts of explanation was the problem of explanatory asymmetry. The paradigmatic example is that of a flagpole and its shadow: we can use the position of the sun, the length of a flagpole and the laws of optics to explain the length of the flagpole's shadow, but we cannot use the position of the sun, the length of the shadow and the laws of optics to explain the length of the flagpole, even though there is a valid deduction in both directions. The causal approach, pioneered by Salmon (1984), is in large part inspired by such problems of asymmetry. The length of the flagpole is the *cause* of the length of the shadow, whereas the latter is the *effect* of the former. Causal theories can solve the asymmetry problem.

In order to prove its sufficiency, Kitcher's theory should be able to reproduce the explanatory asymmetry of the flagpole case. The notion of unification must somehow generate these asymmetries. Kitcher accepts this challenge, and argues (1989, 484–488) that the best systematization  $S(K)$  of  $K$  that contains the pattern deriving the length of a pole from the length of its shadow is less unified than the best systematization *tout court*,  $E(K)$ . We will follow Kitcher's argument in order to assess it.

According to Kitcher,  $E(K)$  contains a very general argument pattern that he calls the *origin-and-development* pattern. This pattern allows the derivation of the size of material objects from the conditions in which they originated and the changes they have since undergone. Using the origin-and-development pattern, the length of a flagpole can be explained by describing its genesis and the substantial changes it has since undergone. Since this pattern can be used to explain the sizes of all objects,

adding a new pattern that explains these sizes from the lengths of shadows does not allow us to derive more conclusions, and is therefore disunifying.

We may object that  $K$  may not contain the premises needed to derive the size of every object using the origin-and-development pattern. In particular, it is possible that  $K$  contains no statements about the origin and development of the pole, but *does* contain statements about the length of its shadow and the position of the sun. If this were the case—and this situation is not particularly far-fetched—the shadow pattern *would* allow us to derive new conclusions, and Kitcher's argument would grind to a halt. As far as I can see, the only way to avoid this counterargument is to restrict ourselves to the ideal situation in which all information is available. This is a heavy concession, as Kitcher explicitly wishes to avoid such idealising assumptions.

Returning from our critical excursion, we find Kitcher looking at the possibility of entirely replacing the origin-and-development pattern with the shadow pattern. If the shadow pattern can be used to derive the sizes of all objects, then it might entail the same consequences as the origin-and-development pattern and  $E(K)$  and its rival  $S(K)$  would be equally unifying. However, not every object casts a shadow, as some are unilluminated, transparent, or strong sources of light. That means we cannot instantiate the shadow pattern to explain the sizes of all objects. The consequence set of  $S(K)$  is smaller than that of  $E(K)$ , and  $E(K)$  is to be preferred over its rival  $S(K)$ . If this analysis is correct, it would solve (at least part of) the problem of explanatory asymmetry.

But Kitcher recognises that the asymmetry problem 'cuts deeper':

Suppose that a tower is actually unilluminated. Nonetheless, it is possible that it should have been illuminated, and if a light source of a specified kind had been present and if there had been a certain type of surface, then the tower would have cast a shadow of certain definite dimensions. So the tower has a complex dispositional property. . . . From the attribution of this dispositional property and the laws of propagation of light we can derive a description of the tower. (1989, 485–486)

However, Kitcher argues, there has to be one pattern for unilluminated objects, another pattern for transparent objects (involving a dispositional property of casting shadows when coated with an opaque substance), yet another pattern for light sources (perhaps involving a dispositional property of casting shadows when illuminated by a much stronger light source), and so on. A large number of *shadow* patterns is needed to do the work that the origin-and-development pattern did all by itself. That means that  $E(K)$  is better unified than  $S(K)$ ; consequently, the theory of unification excludes explanations of the size of objects by the size of their shadows.

This argument is a complex tangle of thorns, and we will have to move carefully in appraising it. First, notice that Kitcher allows dispositional properties. Dispositional properties support counterfactuals, and hence they have a close connection with both laws of nature and causality. This is not the place to speak about the nature of this connection, but building up a theory of causality by appealing to dispositional properties does not appear to be an unproblematic strategy. So much the better for Kitcher, perhaps: he can simply abandon dispositional properties and without them the shadow pattern will be even less successful. However, it may be the case that some of our scientific knowledge is dispositional, and thus part of  $K$ . 'Electrons have mass  $m$ ' might be thought to imply 'if a force  $\vec{F}$  is applied to an electron, it will undergo an acceleration of  $\vec{F}/m$ '. If this is the case, and causal claims are implicit in the set of scientific knowledge  $K$ , then causality cannot be generated by unificatory constraints on the systematization of that knowledge.

We will not pursue this issue here. There is an easier way to show that causal asymmetry cannot be grounded in unificatory constraints. As a rival to the origin-and-development pattern, I propose to define the end-and-regression pattern. (A similar idea is pursued by Barnes [1992].) This pattern uses the *final* state of an object and the transformations it previously went through as premises in a deduction of facts about its *earlier* states. Given the fundamental time symmetry of the known laws of nature and the ideal cognitive situation that we earlier had to suppose, this new pattern generates explanations of all the phenomena that the old pattern generated explanations of.<sup>2</sup> The old pattern has been replaced with a new pattern that has the same consequence set. It seems, then, that unificatory constraints cannot discriminate between argument patterns that explain causes by their effects, and patterns that explain effects by their causes. But if this is the case, neither the flagpole and shadow example, nor any other causal asymmetry, can be generated by a unificationist theory. Kitcher's theory does not give sufficient constraints on explanatory power.

*3.2. Lawhood in Kitcher.* We will now strengthen the conclusions of the previous subsection by demonstrating that Kitcher's theory is not sufficient for distinguishing between a class of explanations and a class of non-explanations that can be prized apart by using the opposition between laws and accidental generalizations. Laws of nature featured prominently in Hempel and Oppenheim's (1948) influential attempt to analyse expla-

2. A possible exception is the final states of objects. This is exactly counterbalanced by the end-and-regression pattern's ability to explain initial states.

nation using the ‘deductive-nomological model’.<sup>3</sup> In this model, an event can be explained *only* by invoking a law of nature of which the event is an instance. The distinction between generalizations that are simply true, and generalizations that are laws of nature was of the essence for Hempel and Oppenheim because not every generalization is explanatory: that all members of a certain club are bald cannot be used to explain John’s baldness, even if we know he is a member of the club—assuming, of course, that there is no shaving ritual involved in becoming a member.

The observation is this: “All men with hair of this-and-this type are bald before the age of 50” might feature in an explanation of John’s baldness, but “All members of the local Rotary are bald” might not. The opposition between laws and accidental generalizations allows us to make this distinction. The question is this: can unification also be used to make this distinction?

Kitcher deals with laws in a short section of Kitcher 1989, writing:

So we can suggest that the statements accepted as laws at a given stage in the development of science . . . are the universal premises that occur in explanatory derivations. (1989, 447)

According to Kitcher, then, lawhood is conferred upon statements by their role in explanatory derivations. Laws simply are the universal premises in genuine explanations. Lawhood is thus conferred on generalizations by virtue of their appearance in explanations. (This is the exact reverse of the claim of Hempel and Oppenheim, who based explanatory power on lawhood.) In order to establish that Kitcher’s criterion of lawhood is unacceptable, it suffices to show that there are explanations which contain generalizations that are not laws. I will do that in the rest of this subsection.

Why is not a single member of the local Rotary a member of the Luxuriant Flowing Hair Club? Because all members of the local Rotary are bald, and bald people cannot become members of the Luxuriant Flowing Hair Club. This, surely, is a perfectly good explanation. One of its premises is “All members of the local Rotary are bald,” and hence Kitcher’s theory indicates that this is not an accidental generalization, but a law. But if it is a law, there is no reason why we reject the proposed explanation of John’s baldness by his membership of the local Rotary.

The unificationist can reply in two different ways. First, he or she can attempt to show that my explanation is not, after all, a good explanation; and thus try to rescue the idea that lawhood is something that is grounded in unification. Second, he or she can attempt to show that the unificationist

3. The question of lawhood and explanation has remained topical; see, for instance, Psillos 2002.

theory can reject the explanation of John's baldness by the generalization about the Rotary in some way that does not involve lawhood. Our response to the first strategy will lead to a response to the second.

In order to reject the explanation, the unificationist would have to say that it will not be part of the most unifying set of argument patterns of our knowledge. The real scientific explanation of the non-overlap between members of the Rotary and those of the Luxuriant Flowing Hair Club will be in terms of real laws: perhaps sociological or psychological laws; perhaps even the laws of physics.

Two responses are open to us. First, if unificationists are bound to reject the explanation we gave—an explanation all of us should accept—this is in itself a counter-argument against unificationism. There are presumably many explanations of the phenomenon we question, and rejecting all but one (or a few) of them in the interest of having a 'minimal amount of argument patterns' does not seem justified. This might be developed into a general line of argument against unificationism: by seeking to retain as few potential explanations as possible, it is blind to the abundance of explanations. But we will not attempt to do so here.

The second response is more straightforward. It is simply this: we construct a scenario in which the only explanation of the non-overlap between members of the Rotary and those of the Luxuriant Flowing Hair Club is the one given above, while no explanation in terms of real laws warrants acceptance.

Suppose that, in an old shoe box in the basement, we find the following items: a membership list of the Rotary and a membership list of the Luxuriant Flowing Hair Club, both in the same town and in the same year; and a black-and-white group photograph of the Rotary, all members of which are bald. This is a historic discovery, because this town was completely destroyed by a tornado, and all the information about its inhabitants was thought lost. In fact, all of it *is* lost, except for these items.

If we attempt to explain why no Rotary member became a member of the LFHC on the basis of social or physical laws, we face the problem of a radical underdetermination of the theory by the evidence. There are many potential explanations—perhaps the town employed a rigid caste system, with each caste having their own clubs—all of which have their unique presuppositions about the social or physical structures in place. For the sake of the example we will suppose that none of these presuppositions is confirmed by the data to a degree that warrants its inclusion in the store of scientific knowledge, *K*.

The scientific situation of which this example is a colorful illustration is quite common. It often happens that the data underdetermine the choice of a general theory to such an extent that we do not accept any theory,

but confess that we are ignorant. At the same time, we see patterns in the data, and try to explain them. Since no general theories are accepted, and since an explanatory argument pattern must use only premises that are in  $K$  (Kitcher 1981, 519), we cannot use general theories to explain the patterns in the data. But sometimes we can explain it using a local story featuring no general laws whatsoever.

In our example, we can explain why no Rotary member became a member of the LFHC by showing people the photograph and saying: “Well look, they were all bald!” It is a good explanation. It is also the only explanation we have, because all explanations based on social or physical laws are unacceptable as their presuppositions are not in  $K$ . The best explanation, therefore, is the one that does not contain laws. The first unificationist strategy therefore fails.

By modifying the scenario, we can also use it to defeat the second unificationist strategy: showing that unificationists can reject an explanation of John’s baldness by his membership with the Rotary in a way that does not use the notion of law. We will do this by showing that there are cases in which the generalization that all members of the Rotary are bald is genuinely unifying.

Assume that we find a list of names of everyone who lived in the town. Behind every name is written what clubs the person is a member of, and whether he is bald or not. This is the entirety of our knowledge about the town.

There is one strong correlation between the entries of the list: everyone who is a member of the Rotary is also bald. In the unificationist theory of Kitcher, adding the argument pattern “ $X$  is a member of the Rotary; all members of the Rotary are bald; therefore,  $X$  is bald” will increase  $C(D)$ . By making the number of members of the Rotary large enough, we can always make sure this will more than balance the addition of a new argument pattern, thus increasing unification. Hence, Kitcher must accept the non-explanation as a real explanation.

We conclude that unification by itself is not enough to solve the problem of asymmetry and the problem of accidental generalizations. For both of these reasons, unification is not sufficient to ground explanation.

*3.3. Lawhood in Schurz and Lambert.* Schurz and Lambert explicitly add a causal theory to the body of knowledge  $KNOW$ , which is meant to reflect the best knowledge about causality that is available to a given cognitive agent or community. Arguments that proceed from causes to effects get a unification bonus, whereas arguments that proceed the other way incur a unification penalty. This strategy ensures that causal explanations are preferred to non-causal or counter-causal ones; but it also

means relinquishing the ambition of Kitcher to generate causality from unification.

Nor do Schurz and Lambert fare better where lawhood is concerned. Let us recall the final scenario given in the previous section, where we had found a list of names, club membership, and degree of baldness. In the theory of Schurz and Lambert, adding the theoretical statement “all members of the Rotary are bald” moves several pieces of data from the ‘basic’ to the ‘assimilated’ category. If the Rotary has enough members, this increases unification and “John is bald because he is a member of the Rotary and all members of the Rotary are bald” must be a genuine explanation—but it is not.

In general, a generalization is allowed in  $K_b$  whenever enough particular facts that used to be in  $K_b$  can be derived from it by arguments in the broad sense. These facts will then be moved to  $K_a$ , generating a unification bonus. This bonus will outweigh the cost of adding the generalization to  $K_b$  if and only if some (unspecified) number of particular facts is involved. Thus, whether a generalization is unificatory and hence allowed in  $K_a$  depends only on the number of its previously unassimilated instances. But the number of previously unassimilated instances cannot be a criterion of lawhood: some accidental generalizations have huge numbers of instances, while some genuine laws may have none, like Newton’s first law.

This means that the theory of Schurz and Lambert must also condone non-explanation as explanation, or invoke the criterion of lawhood (or derive lawhood from causality, if such a thing is possible). Either way, unification is not sufficient for explanation.

I conclude that neither of the two unificationist theories I have discussed gives sufficient conditions for explanatory power.

**4. Is Unification Necessary for Explanation?** In the previous sections I argued, first, that Kitcher’s theory of unification is beset by a profound internal difficulty, and second, that neither Kitcher’s nor Schurz and Lambert’s theory is strong enough to explain the roles of causality and lawhood in explanations. I have thus argued that unification does not yield sufficient conditions for explanatory power: additional conditions involving causality and lawhood have to be added. In the present section I will claim that unificationism does not provide necessary conditions either: explanations do not have to be unificatory. I will defend the positive counter-claim: some explanations disunify our knowledge.

Gerhard Schurz (1999, 97) presents a necessary condition of explanation, (U):

The explanatory premises Prem must be *less in need of explanation* (in  $C + A$ ) than the explanandum  $P$  (in  $C$ ).

One page later, he claims that this condition leads to a unificationist theory of explanation:

In condition (U), *being-in-need-of-explanation* is the crucial concept that leads to a unification- or coherence-based approach of explanation. The being-in-need-of-explanation of a phenomenon *P* in cognitive state *C* comes in degrees, and it depends of how well *P* fits into *C* or coheres with *C*. . . . [I]f condition (U) is satisfied, then the loss of coherence due to the addition of *Prem* to *C* must be smaller than the gain of coherence due to the assimilation of *P* to *Prem* in *C + A*. . . . Hence condition (U) implies that the answer can be explanatory only if the *total* coherence of the cognitive corpus has been increased because of this addition.

Being-in-need-of-explanation is equated to fitting badly into the cognitive corpus. Condition (U) thus demands that the premises from which the explanandum *P* is derived fit better into the cognitive corpus than *P* itself does. The ‘total amount’ of being-in-need-of-explanation must decrease, which is another way of saying that the unification of the cognitive corpus must increase. If condition (U) holds, it is necessary that explanations increase unification; and if it is necessary that explanations increase unification, condition (U) holds. Whether condition (U) holds or not and whether unificationism does or does not furnish necessary conditions for explanation will be decided together.

Does condition (U) hold? Let us examine the example used by Schurz. While sitting in your third-floor office, you see your colleague Peter falling past the window. ‘Why did Peter fall past the window?’, is the question that naturally comes to mind. After all, it is surprising that Peter falls past the window; the proposition *P*, ‘Peter just fell past the window’, does not fit well into your cognitive corpus. It was not to be expected. According to condition (U), an explanation of *P* must derive *P* from premises that fit better into the cognitive corpus *C* than *P* does.

Schurz illustrates this with two proposed explanations. Explanation *A1* is: ‘Because one second ago, Peter was falling past the window of the fifth floor’. According to Schurz, although my background knowledge allows me to derive *P* from *A1*, *A1* is nevertheless not explanatory because it is just as much in need of explanation as *P*. It does not fit well into *C* either. The second explanation is *A2*: ‘Because the fire brigade is testing a new jumping sheet at our building’.

There is nothing puzzling about firebrigades testing jumping sheets: though the event is not very likely, it has plausible ‘how possible’-explanations and thus is heuristically assimilated. Hence, the answer *A2* is completely satisfying. (Schurz 1999, 108)



Is it possible that the fire brigade testing jumping sheets at my office building at this moment of the day (a phenomenon which I will call  $Q$ ) is less in need of an explanation than Peter falling past the window? Let us assume that we are not overly puzzled by its being *Peter* who fell, by his falling past *my* window, by the fall's happening at this exact moment, or by any other detail that is not explained by answer  $A2$ —let us assume, in other words, that  $Q$  is connected to  $P$  by a deductive or strong probabilistic argument. Is it possible that  $Q$  does not stand in need of an explanation while  $P$  does? It certainly cannot be the case that  $Q$  has a high probability and  $P$  a low one. An argument in the broad sense guarantees that a high probability of the premises implies a high probability of the conclusion; there is an argument in the broad sense connecting  $Q$  to  $P$ ; and therefore,  $Q$  cannot be very likely and  $P$  very unlikely at the same time. So if being-in-need-of-explanation is a matter of probability, the premises from which a conclusion is reached can never be less in need of explanation than the conclusion itself.

According to Schurz, being-in-need-of-explanation is not to be construed in terms of probabilities. Something is in need of explanation if it has no plausible 'how possible'-explanations. A 'how possible'-explanation is an explanation that either shows that the phenomenon is truly random (such as, perhaps, quantum wave collapse), or shows that the phenomenon can be inferred from a theory  $T$  in  $K$  using boundary conditions  $Cd$  which do not have to be in  $K$ , but must be compatible with  $K$ . Presumably,  $Q$  is not a truly random phenomenon; but it is plausible that some theories in  $K$  (about the practices of fire brigades, for instance) can generate  $Q$  when combined with appropriate boundary conditions. So  $Q$  has a 'how possible'-explanation, and is not in need of explanation.

However, there is *ex hypothesi* an argument in the broad sense that connects  $Q$  to  $P$ . At the very least this must mean that if  $Q$  is possible,  $P$  is possible. Surely, then,  $P$  also has a valid 'how possible'-explanation using  $T$  and  $Cd$ . In addition,  $P$  has many independent alternative 'how possible'-explanations including Peter's being suicidal, Peter's having been thrown out of the fifth-floor window by an angry customer, Peter's testing a new bungee jumping cord for the local bungee club, and so forth.

It is impossible that the conclusion of a valid argument in the broad sense does not have 'how possible'-explanations if the premises *do* have them. Furthermore, having a 'how possible'-explanation does *not* ensure that a phenomenon no longer stands in need of explanation. The possibility of Peter's fall has not been contested or doubted by anyone. Anyone with a little imagination can come up with ten possible explanations of Peter's fall in the space of two minutes. What is asked for when we want Peter's fall explained is not an explanation of  $P$ 's possibility, but an explanation of  $P$ . (It should be noted that a fact that has *no* known plausible

‘how possible’-explanations will be very disconnected with the rest of *C*, and will thus generally be very much in need of explanation. The reverse, however, is not true: having a plausible ‘how possible’-explanation is a very weak condition, and does not imply being well-connected with the rest of *C*.)

Giving an explanation of *P* can take two forms. *P* can be explained using only propositions in *K*, by pointing out an argument in the broad sense that leads from these propositions to *P*. In such a case, the explanation merely adds arguments in the broad sense to *I*, the set of inferences in *C*. The other possibility is that new propositions have to be added to *K* in order to explain *P*. These new propositions *must* be surprising given the rest of *K*, that is, not certain or highly likely given the rest of *K*, as otherwise it would not have been necessary to add them. (One could simply have derived them.) Thus, as far as ‘standing-in-need-of-explanation’ is an objective term, newly introduced premises must always stand in need of explanation. Condition (U) does not hold.<sup>4</sup>

Phenomena can be explained by other phenomena that are just as unlikely and unexplained. Indeed, I venture the claim that the majority of explanations we encounter in practice are like that. Are Newton’s laws less in need of an explanation than the phenomena they help to explain? Is the length of the flagpole less in need of an explanation than the length of the shadow it helps to explain? Rather, what happens in each of these cases is that we satisfy our curiosity about one ‘unlikely’ phenomenon by deriving it from another ‘unlikely’ phenomenon about which we are less curious. But the explanations would be just as good if the phenomena that feature in the explanans were much *more* in need of an explanation than those in the explanandum—provided, of course, that both these sets of phenomena are admitted into the cognitive corpus *K*. This completes my demonstration that unificationism does not furnish us with necessary conditions for explanatory power.

I wish to make two more, related, points concerning this topic. First, I wish to point out the difference between local ‘connectedness’ and global ‘connectedness’. Second, I wish to offer a brief explanation of the popularity of the idea that unification and explanation are closely linked.

4. My analysis is not inconsistent with the well-known fact that explanations are not in general infinitely regressive. We stop asking explanatory questions and feel satisfied not because some objective state of not-being-in-need-of-explanation has been reached, but because at some point we are no longer interested in following the chain of explanations further down. We are satisfied on being told that the fire brigade is testing a new jump sheet today at our office using Peter as test subject, simply because we are not interested in further explanations of this fact. It is lack of interest, rather than achievement of unification, that stops the potentially infinite chain of explanatory questions.

Schurz and Lambert represent our knowledge as a web of statements connected by arguments. We may speak about the ‘connectedness’ of statements as a measure of the number and strength of the arguments connecting them to other statements in *K*. Schurz and Lambert’s unificationist theory of explanation then states that something is an explanation of *P* only if it has two effects: it increases the connectedness of *P*, and it increases the total connectedness of *K*. *P* must be linked to other statements in order to be explained; *P*’s local connectedness must be increased. But the total set of knowledge must also become more unified; the global connectedness of *K* must increase. And global connectedness is equal to unification.

My analysis suggests that increasing the global connectedness of *P* is not a necessary part of explaining *P*.<sup>5</sup> *Q* can explain *P* even if adding *Q* to *K* unravels large parts of the web. Planck’s postulation of light quanta explained the black body radiation curve, even though this postulate unravelled many connections based on the wave theory of light. Of course, almost nobody was willing to accept Planck’s postulate, including Planck himself. This brings me to my second point.

*Q* can explain *P* only if we are willing to believe that *Q* is true. If *Q* is a disunifying postulate, it is incompatible with statements we formerly believed to be true. This will often decrease our willingness to believe *Q*. Therefore, we are often unwilling to accept disunifying explanations; not because explanations cannot be disunifying, but because the statements we are asked to believe are incompatible with established parts of our knowledge. The premises of unifying explanations, in contrast, can be compatible with all our previous beliefs. The reason unifying explanations are often deemed superior to disunifying ones is simply that we are more inclined to believe the premises of the former. But if, for whatever reason, we are willing to accept the premises of a disunifying explanation, it can function perfectly well as an explanation. If Planck had been willing to accept the particle nature of light, he would have regarded his theory of black body radiation as a perfectly good explanation. And this would have been justified.

Unification is used as a criterion of belief. It is in this capacity that it is linked to explanation, because an explanation is acceptable only when its premises are believed to be true (or probable). But this paper has shown that there is no stronger link than this between unification and explanation. Whether an argument that contains premises we believe to be true actually explains its conclusion is a question that will have to be answered separately from any considerations of unificatory power.

5. It may be necessary to increase the local connectedness of *P*, but we will not pursue this question.

**5. Conclusion.** The notion of unification is important and worthy of analysis. I have tried to show that Kitcher's proposal faces serious difficulties, but the theory of Schurz and Lambert is more successful. As a theory of unification, I have no quarrel with it.

However, as unificationist theories of explanation, both Kitcher's and Schurz and Lambert's theory face serious difficulties. They are not sufficient for the task, as they cannot generate the notions of causality and lawhood which many believe to be important to characterize explanatory power. Moreover, unification is not necessary for explanation. Explanations can have a disunifying instead of a unifying effect. The only reason unifying explanations are deemed preferable is that we are often more inclined to believe their premises.

Therefore, whatever the merits of these theories as theories of unification, as unificationist theories of explanation they are not successful.

#### REFERENCES

- Armstrong, David (1991), "What Makes Induction Rational?", *Dialogue* 30: 503–511.
- Barnes, Eric (1992), "Explanatory Unification and the Problem of Asymmetry", *Philosophy of Science* 59: 558–571.
- Dretske, Fred (1977), "Laws of Nature", *Philosophy of Science* 44: 248–268.
- Friedman, Michael (1974), "Explanation and Scientific Understanding", *Journal of Philosophy* 71: 5–19.
- Hempel, Carl, and Paul Oppenheim (1948), "Studies in the Logic of Explanation", *Philosophy of Science* 15: 135–175.
- Kitcher, Philip (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 Philip Kitcher and Wesley C. Salmon (eds.), *Scientific Explanation, Minnesota Studies in the Philosophy of Science*, vol. 13. Minneapolis: University of Minnesota Press, 410–505.
- Psillos, Stathis (2002), *Causation and Explanation*. Montreal: McGill-Queen's University Press.
- Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- Schurz, Gerhard (1991), "Relevant Deduction: From Solving Paradoxes towards a General Theory", *Erkenntnis* 35: 391–437.
- (1999), "Explanation as Unification", *Synthese* 120: 95–114.
- Schurz, Gerhard, and Karel Lambert (1994), "Outline of a Theory of Scientific Understanding", *Synthese* 101: 65–120.
- Strevens, Michael (2004), "The Casual and Unification Approaches to Explanation Unified—Causally", *Noûs* 38: 154–176.
- Woodward, James (2003), "Scientific Explanation", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2003 ed. <http://plato.stanford.edu/archives/sum2003/entries/scientific-explanation/>.