

An Argument Against the Unification Account of Explanation

Michael Strevens

Draft of December 2009

ABSTRACT

Unification accounts of explanation equate unifying power with explanatory power; you might expect, then, that an increase in the known unifying power of a theory would be accompanied by an increase the perceived quality of its explanations. This paper presents examples suggesting that there is no such relationship: a theory's increased unifying power means that it explains many new things, but it does not explain the old things any better just because it now explains the new things. It seems that the relationship between unifying and explanatory power is rather complex—or non-existent. The negative implications for the unification account are investigated.

1. The Unification Approach to Explanation

According to a unification account of explanation, to explain a phenomenon—an event, a sequence of events, a law, an ongoing “effect” such as the aurora borealis—is to show that it belongs to a set of phenomena which can be unified by a relatively simple theory (Friedman 1974; Kitcher 1981, 1989). A generic form of the unification account might, then, require that an explanation of a phenomenon e do three things:

1. Present a sufficiently simple theory T ,
2. Present a sufficiently large and perhaps diverse set of phenomena E to which e belongs, and
3. Show that (the phenomena in) E can be derived in the right sort of way from T .

To flesh out the view, three things must be provided: an account of what counts as deriving E from T in the “right sort of way”, an account of what counts as a “sufficiently large” (or sufficiently diverse) set of phenomena, and an account of what counts as a “sufficiently simple” theory. Together, these criteria will determine the “degree of unification” achieved by the derivation of E from T . The strength of the explanation given by T is naturally taken to be proportional to the degree of unification achieved.

I will put pressure on this thesis of proportionality, arguing that an increase in the degree to which a theory unifies a set of phenomena does not seem to correspond to an increase in the quality of the explanations provided (although—in the case where the increase in a theory’s unifying power is due to an increase in the number of phenomena derived from the theory—there is of course an increase in the *quantity* of explanations provided). Indirectly, then, I put pressure on the unification account itself. Either there is something wrong with the account itself, or there is something wrong with the proportionality thesis. But the principal motivations for adopting a

unification account are equally good motivations for endorsing the proportionality thesis—in fact, the account is typically motivated by way of the thesis. Unificationism, then, looks to have a problem.

Let me begin by setting out briefly the elements of Friedman’s and Kitcher’s unification accounts.

Friedman (1974) holds that the derivation may be any deductive argument. He defines the degree of unification achieved by a derivation as a measure of how much smaller T is than E , where the size of these two sets is deemed proportional to the number of independently acceptable lawlike sentences required to characterize the content of each. (Friedman does not offer a precise characterization of “independent acceptability”. A necessary condition for independent acceptability, however, is logical independence. Also left open is the question of what counts as a “lawlike sentence”; possibly a certain class of universal generalizations, as in Hempel and Oppenheim (1948, §6), is intended.) The account of explanation is then roughly as follows: T explains e (together with everything in E) just in case the degree of unification achieved by deriving E from T is greater than zero (I omit some details here).

Kitcher agrees with Friedman that all derivation is by deductive argument (Kitcher 1989, §5). But not all deductive arguments are legitimate derivations. It is the unifying theory itself that specifies which kinds of derivations (or *argument patterns*, as Kitcher calls them) are legitimate. A theory, then, includes not only statements about the world (laws, facts and so on), but also rules specifying the kinds of arguments that may be used to derive phenomena from these statements (Kitcher 1989, §4.6).

Is there any limit on the argument patterns that might be endorsed by a theory? Can a theory recommend an apparently non-explanatory argument such as $A.B$, therefore B ? Kitcher builds into his definition of unifying power a desideratum that the argument patterns recommended by the unifying theories be *stringent*, a condition which he argues counts heavily against

intuitively non-explanatory derivations such as that of A from $A.B$. An argument pattern is stringent to the degree that it is “difficult to satisfy” (Kitcher 1989, 433), in particular, to the degree that it restricts the non-logical vocabulary that can appear in its instantiations and to the degree that it puts restrictions on the logical form of its instantiations. Kitcher does not say so, but as I understand it, the point of stringency is to restrict argument patterns to those that turn on the existence of what might be called real patterns in nature. (For example, I take it that deriving the fact that a particular F is G from a premise stating that all F s are G s turns on a real pattern of G -ness among the F s, whereas deriving B from $A.B$ does not turn on any non-trivial pattern at all.) Since unification is an attempt to identify patterns, a stringency requirement or something like it is a natural constraint for a unification theory to embrace. The question how properly to articulate the requirement is difficult and important, but I will not pursue it further here.

Kitcher (1989, 435) characterizes unifying power as follows. A theory T from which phenomena E can be derived offers a degree of unifying power that

1. Increases with the size of E ,
2. Decreases with the size of T , and
3. Increases with the stringency of the arguments recommended by T .

The resulting notion of unifying power is very close to Friedman’s, but with the notion of stringency doing roughly the work of the notions of independent acceptability and lawlikeness.

What Kitcher does with unifying power, however, is rather different. For Friedman, any unifying theory qualifies as explanatory. For Kitcher, only the most unifying theory counts as explanatory. Take all the world’s phenomena, then. Whatever theory or set of complementary theories (physical, chemical,

biological, and so on) best unifies the phenomena gives their correct explanation; no other theory, however unifying in its own right, explains anything at all (Kitcher 1989, §8).

Kitcher himself prefers to present his theory in epistemicized terms, an expository tactic that will suit me equally well. Accordingly, define the *acceptable* explanatory theories, relative to a given body of knowledge, as those that together best unify the known phenomena. An acceptable explanation of a known phenomenon is therefore a derivation of the phenomenon from such a theory. Kitcher accordingly calls the set of acceptable explanatory theories at any given time the *explanatory store*.

The relationship between explanatory correctness and explanatory acceptability seems to be straightforward: explanatory correctness is acceptability in the epistemic limit—when all the information is in—and acceptability is apparent correctness.

2. The Proportionality Thesis

To say that a theory has great explanatory power might be to say one of two different things. On the one hand, it might be to say that the theory explains many phenomena. On the other hand, it might be to say that the theory explains the phenomena that it does very well—that it explains them in a profound or deep way. Call the first kind of power *extensive* explanatory power and the second kind of power *intensive* explanatory power.

The arguments in this paper are directed at a thesis about intensive explanatory power that I attribute to the unificationist approach:

Proportionality Thesis: A theory's intensive explanatory power is proportional to its unifying power.

In other words, the more unifying a theory, the better it explains each of the phenomena that can be derived from it. Or: a more unifying theory not

only explains more phenomena, it provides a stronger explanation of each individual phenomenon.

Why subscribe to the proportionality thesis?

Imagine a universe, *Discordia*, whose phenomena exhibit no large-scale patterns—a collective idiosyncrasy, in which the regularities, where they exist, are local and transient. Gases in *Discordia* might obey Boyle’s law in some spatiotemporal regions, but they would conform to other generalizations in other regions. A Gods-eye attempt to construct a unified description of the goings-on in *Discordia* would be trying and perhaps tedious, but it would not end in failure: it would capture the local and transient regularities, and so would constitute a specification of the *Discordia*’s history that would be considerably simpler than an enumeration of every event in isolation.

Such a description—the ideal end product of *Discordian* science—might be taken to explain what goes on in *Discordia*. But the unfortunate *Discordians* would surely be, by comparison with ourselves, explanatorily impoverished. Their best explanation of the Boylean behavior of gases would amount to not much more than “Gases tend to behave that way around here, for now”. By contrast, our best explanation of Boylean behavior proceeds by examining the implications that a handful of wide-ranging fundamental laws of nature have for any system possessing the physical structure of a gas.¹ We have the luxury (or so most of us think) of living in a place where everything that happens, happens for reasons that make sense in the light of a few fundamental laws of nature. Our explanations are simply better than the explanations of *Discordian* science. That is not to find fault with *Discordian* science, but rather with *Discordia* itself: it is a place where even the best explanations for what goes on are not very satisfying. They are explanations, but they are not strong explanations.

Unificationism provides an appealing account of this intensive notion of explanatory strength or depth: the Boylean behavior of a particular gas is

1. Also important are some posits about probability, ignored here.

explained better in our universe than in Discordia because the best theory of our universe is far more unifying than the best Discordian theory. Explanatory satisfaction comes from seeing the phenomenon to be explained subsumed under as wide-ranging a pattern, or under as few and as general principles, as can be found. The wider the scope of the pattern or principles, the more satisfying—the intensively better—the explanation. In making sense of the Discordian situation, then, the unificationist has every reason to take advantage of the proportionality thesis that is in any case at their fingertips.

Could the unificationist nevertheless renounce the proportionality thesis? Not without great ideological contention. Consider two bodies of phenomena, the *A*-effects and the *B*-effects. Suppose that you have two theories of equal intrinsic complexity. The *A*-effects, but not the *B*-effects, can be derived (in an explanatorily legitimate way) from theory 1. Both the *A*-effects and the *B*-effects can be derived from theory 2. All forms of unificationism hold that we should prefer to explain the *A*-effects using theory 2. I will argue that it is very difficult to make sense of this recommendation while denying the proportionality thesis.

Why should we use theory 2 to explain the *A*-effects, if the *A*-effects can equally well be derived from theory 1? It seems that the unificationist must answer: because theory 2 provides a better explanation of the *A*-effects than does theory 1. (If not, why not use theory 1?)

But then what is the basis of theory 2's explanatory superiority? There seems to be only one possible answer: theory 2 explains the *A*-effects better than does theory 1 because theory 2 also explains the *B*-effects. That is to endorse the proportionality thesis.²

2. Another possible answer is that, because theory 2 predicts more actual phenomena than theory 1, we have more reason to think that theory 2 is correct than that theory 1 is correct. On this view, unificationism is a kind of combined theory of explanation and confirmation, telling you both which theories are likely to be true given the evidence and which theories explain the evidence. It seems to me to be a bad idea to pursue such a

There is more to say: I will return to the question whether the unificationist is committed to the proportionality thesis in section 5. But now is the time to give my reasons for thinking that the thesis is false.

3. Explanatory Power Is Not Proportional to Unifying Power

Consider again Boyle's law, the law that in a gas held at a constant temperature, pressure is inversely proportional to volume. In virtue of the law, every known gas (provided that its density and pressure are not too high) exhibits Boylean behavior: when a quantity of oxygen is kept at constant temperature, its pressure varies inversely with volume; when a quantity of radon is kept at constant temperature, its pressure varies inversely with volume; and so on. The unifying power of Boyle's law is proportional to the number of such generalizations, or perhaps to the number of instances of such generalizations, depending on how the unificationist counts phenomena. Thus if explanatory power is proportional to unifying power, the explanatory power of Boyle's law is proportional to the number of gases. As a consequence, if a new gas—say, vapor of element number 118—is created, any explanation that makes use of Boyle's law becomes more powerful. Upon the synthesis of element 118, our explanation of the Boylean behavior of oxygen becomes better—we understand better why oxygen conforms to Boyle's law. Is that not a reductio of the proportionality thesis?

A unificationist might reasonably reply that the unifying power of a theory is to be judged not relative to currently existing phenomena, but relative to all the phenomena that have ever and will ever exist. In a world where element 118 is fated to be created, the unifying power of Boyle's law

project. Following Hempel, we should distinguish the question of what makes a theory a good potential explanation from the question of whether the theory is true (in Hempel's terms, the theory's "empirical adequacy"). Unificationism should be seen as a theory of potential explanatoriness; the question of truth should be bracketed, then, when asking why theory 2 provides a better explanation of the *A*-effects than theory 1.

will reflect its ability to cover the behavior of the element even before it exists. The unifying power of a theory— thus, on the proportionality thesis, its explanatory power— can never change.

In response, let me follow Kitcher's lead, focusing on the *known* unifying power and apparent explanatory force of Boyle's law. When element 118 is discovered, we see that Boyle's law unifies one more phenomenon than it did before, and so— if we are unificationists— we ought to think that our explanation of the Boylean behavior of every gas has turned out to be better than we previously realized. Our explanation does not actually become more powerful, but its perceived explanatory power increases.

Again a reductio, or so I claim: we do not think that the explanation of the behavior of oxygen is any better after the creation of element 118. Rather, the quality of the explanation of oxygen's behavior depends on the structure of oxygen and what our fundamental laws of nature have to say about things with that structure. That the structure is shared with other gases— that the Boylean explanation has the power unify a wide range of phenomena— is explanatorily irrelevant. The proportionality thesis is false.

Another unificationist riposte: perhaps the case of element 118 fails to take into account the importance of the *variety* of phenomena unified. The phenomena that are unified by Boyle's law are just instantiations of exactly the same type of behavior over and over again. After the Boylean behavior of ten thousand gases has been unified, it does not seem a very significant unifying achievement that the behavior of one more gas has been added to the list. As it happens, neither Friedman nor Kitcher take the variety of phenomena unified into account when calculating unifying power— and there may be considerable technical difficulties in doing so— but perhaps it is something that needs to be done all the same. To take this strategy to an extreme, you might count every kind of phenomenon only once in calculating unifying power. Unifying power would be proportional the the number of *types* of phenomena unified.

A revised version of the Boyle's law case will circumvent this strategy. The kinetic theory of gases entails and so explains the truth of a number of gas laws, including Boyle's law, Gay-Lussac's (or Charles's) law, Avogadro's law (these three make up the ideal gas law), Graham's law of effusion (concerning the escape of a gas from a small hole), and many laws concerning diffusion (which can be derived to order from the kinetic theory).

According to the proportionality thesis, the explanatory power of the kinetic theory increases with the number of gas laws explained. For example, the degree to which kinetic theory explains Boyle's law is enhanced by the fact that it also explains Graham's law. Again, a *reductio*. The quality of the kinetic theory's explanation of Boyle's law depends on what the kinetic theory says about the pressure and volume of gases, but not on what it says about the rate with which a gas escapes from a small hole. When we come to see that the kinetic theory can explain effusion behavior, we learn something new about kinetic theory and something explanatory about effusion—but our understanding of Boyle's law is enhanced not at all.

The point can be underlined by considering the historical development of the kinetic theory. It is generally agreed that the first more or less correct explanation of Boyle's law was that proposed by Daniel Bernoulli in his *Hydrodynamics* of 1738. Bernoulli assumed that gases are made up of fast, independently moving particles, and that pressure is proportional to the number of these particles impacting a container wall in a given time. He then showed that when the volume of gas is decreased, the number of impacts, and hence the pressure, increases by the same factor. Bernoulli made some false steps, but his explanation (in the cleaned-up form in which I have presented it) is still presented as the explanation of Boyle's law in modern textbooks (Strevens 2009, chap. 8).

Avogadro posited his eponymous law in 1811. Thomas Graham discovered his effusion law in 1846. In the following decades various scientists including John Herapath, Rudolf Clausius and James Clerk Maxwell pre-

sented explanations of these laws, in ever-improved forms (Brush 1983). If scientists were unificationists, then each time one of these laws was explained by kinetic theory, the kinetic explanation of *Boyle's law* would be revealed to be more powerful than had previously been supposed. By the twentieth century, the kinetic theory was known to have several times as much unifying power as it was known to have in 1738. If the proportionality thesis is correct, it follows that the apparent explanatory power of the kinetic theory had by then increased several times over as well, not only in the extensive sense—that the theory was known to explain more phenomena than before—but also in the intensive sense, that the theory was known to provide several times more understanding of each such phenomenon than it was previously known to provide. Whatever satisfaction Bernoulli's contemporaries derived from his explanation of Boyle's law, then, we should derive several much times as much satisfaction now—though the structure of the explanation has barely changed. That cannot be right.

(Let me emphasize again the epistemic dimension of this argument: the actual unifying power of the kinetic theory has not changed in the last several hundred years. It is the apparent unifying power that has changed. Thus it is not the actual explanatory power of the theory, but its apparent explanatory power, that will according to the proportionality thesis also have changed. But I assume that we are at least as well able to detect apparent explanatory power as actual explanatory power, and so that a theory that makes false predictions about apparent explanatory power is in serious difficulty.)

Finally, a variation on the previous argument. When Bernoulli published his explanation, Boyle's law and a few other, less well-formulated facts were the only behaviors of gases that the kinetic theory could reasonably be hoped to explain. These facts, if written down, would be at least as compact as the kinetic theory of gases, and would contain about as many assertions. Thus the apparent unifying power of the kinetic theory after Bernoulli's explanation was—zero. To put it another way, a mere compendium of gaseous laws had

just as much power as the kinetic theory to unify the behaviors of gases that were known at the time, hence just as much apparent ability to explain those behaviors. If unificationism is correct, Bernoulli's theoretical achievement ought not to have been considered an explanatory advance at all.

Suppose that the kinetic theory first achieved positive unifying power relative to the epistemic background in 1847, when Herapath derived Graham's law of effusion from the kinetic theory. (At this point, I am supposing, it became apparent that kinetic theory provided a simpler and more integrated summary of the known behaviors of gases than did a list of gas laws covering those behaviors.) Then according to the unification account, kinetic theory first acquired the apparent ability to explain Boyle's law not in 1738, when Bernoulli derived the law from the theory, but in 1847, when Herapath used the kinetic theory to derive . . . Graham's law!

4. Unificationist Replies

A unificationist might contemplate two kinds of reply to the arguments assembled in the previous section. First, they might embrace those consequences of the proportionality thesis that I have done my best to argue are absurd, or at least, contrary to actual scientific explanatory practice. Second, they might find some way reject the proportionality thesis.

Most of what follows explores the second option, but a surprising comment of Friedman's, discussing of all things the kinetic theory, suggests that he would prefer the absurdist response:

Consider . . . the kinetic theory of gases. The theory explains phenomena involving the behavior of gases, such as the fact that gases approximately obey the Boyle-Charles law [$PV = kT$, the law explained by Bernoulli], by reference to the behavior of the molecules of which gases are composed. For example, we can deduce that any collection of molecules of the sort that gases are,

which obeys the laws of mechanics will also approximately obey the Boyle-Charles law. How does this make us understand the behavior of gases? I submit that if this were all the kinetic theory did we would have added nothing to our understanding. We would have simply replaced one brute fact with another. But that is not all the kinetic theory does—it also permits us to derive other phenomena involving the behavior of gases, such as the fact that they obey Graham’s law . . . (Friedman 1974, 14)

In short, Bernoulli’s contemporaries—who had no reason to think that the kinetic theory unified anything over and above Boylean behavior—ought not to have regarded Bernoulli’s derivation of Boyle’s law as an explanation of the law. Presumably Friedman would be happy to say in addition that the perceived explanatory power of kinetic theory increases as the its known unifying power increases. There is not much I can say in reply, except that this prescription diverges dramatically from the historical reception of the kinetic theory as explainer.

Can a unificationist coherently reject the proportionality thesis? I will consider several different reasons for thinking that the answer might be *yes*.

Consider first the “winner-takes-all” aspect of Kitcher’s unificationism, according to which only the most unifying theory (or set of complementary theories) has the power to explain. According to Kitcher, if theory 2 unifies the phenomena 20% better than theory 1, then deriving a phenomenon from theory 2 constitutes an explanation but deriving it from theory 1 does not. Even though theory 1 has almost as much unifying power as theory 2, then, it has none of its explanatory power. It would appear that Kitcher denies the proportionality thesis.

A closer look at Kitcher’s reasons for endorsing a winner-takes-all theory, however, suggests otherwise. Kitcher (1986) proposes that the most unifying theory is the one that is correct, not because the truth tends to unify best, but because truth for a theoretical posit *consists in* belonging to a maximally

unifying set (compare Lewis's (1994)'s best-system account of laws of nature). On the standard assumption that only truth explains, the winner-takes-all element of Kitcher's account of explanation follows immediately. By contrast, as explained earlier, the proportionality thesis concerns potential explanatory power, that is, explanatory power bracketing the question of truth. Winner-takes-all provides no reason, then, to think that Kitcher is not committed to proportionality about potential explanatory power.

Might Kitcher, or some other unificationist, nevertheless deny the proportionality thesis? One way to do so would be to define the strength of a theory's explanation of a phenomenon solely in terms of the intrinsic properties of the theory, the phenomenon, and the relation between them. The theory's ability to explain other phenomena would then be unable to contribute to explanatory strength.

Considered in itself this is a plausible suggestion, but it works badly within the context of unificationism, for the following reason. Suppose that theory 1 and theory 2 both explain a phenomenon. Suppose further that theory 1 provides the better explanation of the phenomenon (perhaps theory 1 itself, or the derivation of the phenomenon from theory 1, is simpler). It might nevertheless turn out that theory 2 is more unifying, if it explains many more phenomena than theory 1. The unificationist would then be committed to the following perverse claim: although theory 1 explains the phenomenon better—it provides more or deeper understanding of the phenomenon—we should nevertheless prefer to explain it using theory 2. (I take it as axiomatic for a unificationist that the more unifying theory is always explanatorily preferable.)

A better strategy for the unificationist would be to renounce the notion of explanatory strength altogether, that is, to hold that all derivations from sufficiently unifying theories are equally explanatory. Let me give you three reasons why this is a bad idea.

First, without a notion of explanatory strength, the unificationist loses

all the benefits outlined in section 2. They cannot make sense of the claim that some worlds are explanatorily richer than others—that in a world in which all behavior conforms to a few wide-ranging laws, we have better and more satisfying explanations than in a world where regularities are many, varied, and disconnected. Further, though they remain committed to the proposition that section 2's theory 2, which you will recall accounts for both the *A*-effects and the *B*-effects, ought to be preferred over theory 1, which accounts only for the *A*-effects, as an explanation of the *A*-effects, they must at the same time deny that theory 2 explains the *A*-effects better than does theory 1—an awkward package of commitments.

Second, discarding explanatory strength is in any case not enough to solve the final problem laid out in section 3, that on the unificationist view, Bernoulli's account of Boyle's law ought not to have been regarded by his contemporaries as explanatory. Although the Bernoulli problem is of a piece with my other objections, it does not turn on the proportionality thesis, but on a corollary of the thesis that is even more difficult for a unificationist to deny: a new theory that does not bring any net unifying power—that is at least as complex as a mere summary of the phenomena that it explains—cannot explain anything.

Third and finally, to ditch the idea of explanatory strength and along with it the proportionality thesis, seems at odds with the empiricist ideology underlying the unification account.

That ideology's principal tenet is that the basic subject matter of science is limited to, as David Lewis puts it, "local matters of particular fact, just one little thing and then another" (Lewis 1986, ix). Old-fashioned empiricists hold that the particulars are facts about sense impressions; newfangled empiricists such as Lewis that they are physical facts about local arrangements of physical qualities. Both old-fashioned and newfangled empiricists hold that further facts—about causes and laws, in particular—must, if they exist at all, be higher order properties of the basic, local facts, or to put it another

way, that they must be facts about patterns of basic facts (in a broad sense of “pattern”).

All higher order scientific activity, on the view I am describing, is concerned with the discovery and deployment of patterns. There is a strand in this tradition according to which all pattern discovery is directed towards the end of finding the most unifying set of patterns, in the same sense of “unifying” employed by Friedman and Kitcher. Lewis holds, for example, that the laws of nature are the deductive closure of the set of axioms that best unifies all the basic facts (Lewis 1994); the search for the theory of everything, then, is a search for the most unifying description of the basic facts.³ In its purest incarnation, this strand of empiricism holds that only two activities are proper to science: discovering the basic facts and discovering the most unified description of these facts, an attitude well exemplified by Mach’s view that the goal of science is “intellectual economy”. Call this view *unification empiricism*.

The unification empiricists once held that explanation, being a metaphysical pursuit, was not a proper goal of science (Duhem 1954—though Duhem was not quite an empiricist). If, as later empiricists decided, explanation is after all a legitimate scientific activity, then for a unification empiricist it can only be a kind of unification. Hence the motivation for a unification account of explanation. Explanation is essentially a matter of showing how well the explanandum fits into a unifying system of laws. Understanding comes in appreciating the niceness of fit.

For a unification empiricist, the scientific component of explanation is not distinct from the process of unification (though the pragmatic component may be). Explanation is just a kind of unification, a kind where the basic fact to be incorporated into the unified corpus is called the explanandum, and

3. Lewis’s view is that the laws are the closure of the axioms that best combine *simplicity*, *strength*, and *fit*; I take it combined simplicity, strength, and fit in Lewis’s sense is a close cousin, at the very least, of unifying power.

the unifying theory is called the explainer. Thus the power of an explanation is nothing over and above the unifying power of the explainer. Unification empiricism leads naturally, then, to the proportionality thesis.

* * *

Much of what Friedman and Kitcher have to say fits the unification empiricist paradigm. Yet both give, as a motivation for the unification account, a rationale that might be interpreted as distinct from the “intellectual economy” rationale, namely, that the role of scientific explanation is to “reduce the incomprehensibility of the world”. A theory of great unifying power is explanatory, then, because it is great reducer of incomprehensibility.

In this paper’s concluding section, I will show that there is a way of understanding the account of explanation as incomprehensibility reduction that does not commit its proponents to the proportionality thesis. Does this mean that there is a variety of unificationism that is immune to the objections raised above? No; on closer examination, the account of explanation as incomprehensibility reduction undermines, rather than supporting, the unification approach.

5. Reducing Incomprehensibility

Understanding, according to Friedman and Kitcher, comes from reducing the number of fundamental, hence incomprehensible, facts that the scientist must accept:

This is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more. (Friedman 1974, 15)

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). (Kitcher 1989, 432).

The goal of explanation is, then, not so much to increase comprehension as to reduce incomprehension, a rather pessimistic view: we begin science with a massive deficit of understanding, and our goal is to raise our understanding to as near zero as possible. (Compare T. H. Huxley (1896, 165), cited by Kitcher: “In ultimate analysis everything is incomprehensible, and the whole object of science is simply to reduce the fundamental incomprehensibilities to the smallest possible number.”)

On the incomprehension reduction view, the understanding induced by a unifying theory is directly proportional to the number of formerly brute facts made relatively comprehensible by the unification, that is, to the theory’s unifying power. Thus a theory that unifies more phenomena explains better than a theory that unifies fewer. Does it explain each of those phenomena better, or does its explanatory superiority lie solely in the quantity, not the quality of its explanations? Is intensive, as well as extensive, explanatory power proportional to unifying power? Only if the answer is *yes* is the incomprehension reduction view committed to the proportionality thesis, and so vulnerable to the arguments given in this paper.

The resolution of the issue depends, I think, on the sense in which a unification “reduces the number of facts that we have to accept as brute”. On a *strong* reading, the unified facts bear a metaphysical dependence relation to the unifying facts; on a *weak* reading, they bear only a logical relation—that is, the sense in which the unifying facts are more fundamental is nothing more than is implied by the fact that they entail, but are not entailed by, the unified facts. The weak reading, I will argue, collapses the distinction between the incomprehension reduction view and unification empiricism;

such a unificationism is committed to the proportionality thesis. By contrast, the strong reading allows for an account of explanation that is consistent with (though it does not mandate) the denial of the proportionality thesis. However, such an account introduces into explanation a notion of metaphysical dependence that is quite alien to the unification account and has no place in either Friedman's or Kitcher's definition of unifying power. The theory of explanation implied by the strong reading cannot, therefore, be attributed to Friedman or Kitcher, or to anyone who wishes to endorse a unification account of the sort that they offer.

To the strong reading, then. This reading identifies the difference between the "brute" or "irreducible" unifying facts and the "derived" or "reducible" unified facts as a genuine metaphysical difference. On such a view, the fundamental laws of nature have a reality or independent existence that the high-level laws they explain lack. When we derive the high-level laws from the fundamental laws, then, we see that where we might have thought there were several distinct existences, there is in fact only one fundamental (and therefore incomprehensible) existence. The derivation relation, then, demonstrates the presence of a strong metaphysical dependence relation: the high-level laws are somehow dependent for their very being on the fundamental laws.

Observe that this conception of the order of things reverses the empiricist or "Humean" conception sketched above: on the Humean view, it is individual matters of fact that have a fundamental, independent existence. The "fundamental" laws of nature are, metaphysically speaking, the high-level facts: they are patterns to be found in the fundamental facts. The existence of the fundamental laws therefore metaphysically depends on the existence and disposition of the particular matters of fact. There is no prospect, for a Humean, of reducing the number of brute facts. Indeed, on the metaphysical version of the incomprehension reduction view, a Humean universe is a maximally incomprehensible universe.

It is for these reasons that it is highly implausible to credit to either Friedman or Kitcher, or to anyone moved by their empiricist leanings to prefer the unificationist account of explanation, a metaphysical view of incomprehension reduction.

Two remarks on the metaphysical view before I move on. First, let me show, as promised above, that the metaphysical view militates against the proportionality thesis. In order to explain a fact, on the metaphysical view, you must show that its existence wholly depends on some other body of facts. Quite plausibly, dependence of this sort does not come in degrees: a fact either depends entirely on other facts, or it does not. It follows that the reduction of incomprehension that comes from appreciating such a dependence is also an all or nothing matter. Thus there are no degrees of explanation: all explanations are, intensively, equally strong. We should prefer more unifying theories because they explain more, not because they explain better. (That said, a metaphysical view might admit partial dependence relations and hold that partial understanding arises from appreciating such relations; the strong view is therefore not committed to rejecting the proportionality thesis.)

Second, Friedman's and Kitcher's empiricist inclinations aside, is there a way to be a "metaphysical" unificationist of the sort characterized by the strong reading? Yes; but then you are not really a unificationist. Rather, you are an *X*-ist, where *X* is the dependence relation that must hold between facts for one to reduce, and so to explain, the other. If the relation is one of causation, you are endorsing a causal theory of explanation. If it is one of metaphysical necessitation, you are endorsing a modal theory of explanation (or one variety of the modal approach). On any view of scientific explanation, unification will be to some extent a side-effect of successful explanation. An account is unificationist only if it is solely *qua* unifier that a theory explains. On the metaphysical reading, this criterion is not satisfied: a theory explains

not by unifying, but by bearing relation X towards, its explananda.⁴

If Friedman and Kitcher do not have the strong metaphysical reading of “incomprehension reduction” in mind, what are they thinking? In what sense do they hold that the facts in a unifying theory have an independence that the unified facts do not? I can only think that they have in mind logical independence, and thus a weak reading of incomprehensibility reduction: the unifying facts entail the unified facts, but not vice versa. In other words, the facts that we have to accept as “independent” are the axioms of our scientific theories; the facts that we see to be “dependent” are the consequences of those axioms.⁵ What is the explanatory significance of such a logical relation? For a unificationist, the logical relation demonstrates that derived facts are instances of a pattern specified by the unifying facts. A fact is explained to the degree to which it is seen to fit into, to be subsumed by, a wide-ranging yet simple pattern of phenomena. That is unification empiricism. On the weak reading, then, there is no gap after all between unification empiricism and the incomprehensibility reduction view, because the latter is more or less a restatement of the former. As I argued above, unification empiricism is committed to the proportionality thesis; it follows immediately that the weak reading of incomprehensibility reduction—the only reading compatible with genuine unificationism—is committed to the same.

But the proportionality thesis is false.

4. Perhaps Kitcher would argue that an all-or-nothing dependence relation is grounded in the graded notion of unifying power. I would then ask: does the account of what is explanatory cite this relation? If so, what you have is not a unification account of explanation but an X -ist account of explanation, where X is the dependence relation in question, along with a unification or “best system” account of X . If explanation does not go by way of the dependence relation, then the all-or-nothing aspect of the relation will have no consequences for explanatory power: explanatory power will be, like unifying power, graded.

5. In the case of particular matters of fact, they are the consequences of the axioms and initial or boundary conditions.

References

- Brush, S. G. (1983). *Statistical Physics and the Atomic Theory of Matter from Boyle and Newton to Landau and Onsager*. Princeton University Press, Princeton, NJ.
- Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Translated by P. P. Wiener. Princeton University Press, Princeton, NJ.
- Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy* 71:5–19.
- Hempel, C. G. (1965). *Aspects of Scientific Explanation*. Free Press, New York.
- Hempel, C. G. and P. Oppenheim. (1948). Studies in the logic of explanation. *Philosophy of Science* 15:135–175. Reprinted in Hempel (1965), chap. 10.
- Huxley, T. H. (1896). *Darwiniana*. Appleton, New York.
- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science* 48:507–531.
- . (1986). Projecting the order of nature. In R. E. Butts (ed.), *Kant's Philosophy of Physical Science*. D. Reidel, Dordrecht.
- . (1989). Explanatory unification and the causal structure of the world. In P. Kitcher and W. C. Salmon (eds.), *Scientific Explanation*, volume 13 of *Minnesota Studies in the Philosophy of Science*, pp. 410–505. University of Minnesota Press, Minneapolis.
- Lewis, D. (1986). *Philosophical Papers*, volume 2. Oxford University Press, Oxford.
- . (1994). Humean supervenience debugged. *Mind* 103:473–490.
- Strevens, M. (2009). *Depth: An Account of Scientific Explanation*. Harvard University Press, Cambridge, MA.