Unification and the Myth of Purely Reductive Understanding

Michael J. Shaffer*

ABSTRACT: In this paper significant challenges are raised with respect to the view that explanation essentially involves unification. These objections are raised specifically with respect to the well-known versions of unificationism developed and defended by Michael Friedman and Philip Kitcher. The objections involve the explanatory regress argument and the concepts of reduction and scientific understanding. Essentially, the contention made here is that these versions of unificationism wrongly assume that reduction secures understanding.

KEYWORDS: Explanation; reduction; simplicity; scientific understanding, unification.

1. Introduction

In post-Hempelian discussions of the nature of scientific explanation it is commonplace to note that one of the main functions of such explanation is to yield understanding, more specifically, scientific understanding.[1] That

---

[1]   See (de Regt 2017) and (de Regt, et al. 2009).

*   St. Cloud State University
    🖉 Department of Philosophy, St. Cloud State University, CH365N, 720 4th Ave. S., St. Cloud, MN 56301, USA
    ✉ mjshaffer@stcloudstate.edu

Hempel's covering-law model(s) of explanation failed to adequately account for this desideratum is one of the reasons that motivated the general rejection of the Hempelian account(s) of explanation.[2] It is apparent then that a theory of scientific explanation that does not show how explanations of these sorts yield such understanding is *ipso facto* inadequate. We can refer to this constraint on theories of explanation as *the understanding thesis*.

This paper explores how this constraint on one of the most prominent theories of explanation is supposed to be satisfied. In particular, it focuses on one of the most widely held successor theories to Hempel's deductive-nomological and inductive-statistical models of explanation: the view that explanation is a sort of global unification.[3] This is the view that explanation is essentially achieved when a phenomenon or law is integrated into the simplest global system that organizes (or systematizes) our corpus of beliefs. In other words, explanation is a matter of showing how various, apparently unrelated, beliefs can be derived logically from a small set of premises or axioms, thereby reducing the number of beliefs that must be accepted as brute axiomatic beliefs. The argument presented here concludes that standard unificationist theories of explanation of this sort, like the most well-known versions proposed by Michael Friedman (1974) and Philip Kitcher (1981, 1989, 1993), fail to convincingly show how unification yields understanding in any interesting, non-trivial, sense.

Establishing this result involves seriously contending with a particular traditional objection to explanation that has often been blithely ignored. This is *the explanatory regress argument* (ERA). This wide-scoped objection to accounts of explanation is based on the observation that explanations

---

[2]   See (Salmon 1989) for the canonical recounting of the relevant history of the concept of explanation, and (Lambert 1980) for some additional criticisms of the view that Hempel's covering-law models yield understanding. (Woodward 2017) is also a useful survey of theories of explanation.

[3]   Kitcher (1981) points out that lurking behind the official, logical empiricist, doctrine espoused by Hempel was a version of the unificationist view he advocates. This implicit view can be found in (Hempel 1965, 1966) and in Feigl's classic (1970). Prominent contemporary defenders of this view include Friedman (1974) and Kitcher (1981, 1989, 1993).

involve relating the unfamiliar to the familiar, or relating that which is not understood to that which is.[4] According to the ERA bona fide explanation occurs only when that which is being explained—the explanandum—is explained by something—the explanans—that is already itself explained.[5] This point can then be related to scientific understanding by noting that understanding is, in an important sense, parasitic on explanation. Understanding can be achieved only by relating what we do not understand to what we do understand. But then (so the criticism goes) we are faced with what appears to be a regress that can be terminated only by the positing of some privileged body of explanatorily primitive unexplained explainers that themselves generate all understanding. As there do not appear to be such fundamental privileged explainers, it seems that if anything is explained at all and anything understood, explanation *must not* require that an explanandum can only be explained by an explanans that is itself already explained. Unificationists often portray themselves as being able to avoid this problem. For example, Friedman explicitly tells us that,

> This global view of scientific understanding [unificationism] also, it seems to me, provides the correct answer to the old argument that science is incapable of explaining anything because the basic phenomena to which others are reduced are themselves neither explained nor understood. According to this argument, science merely transfers our puzzlement from one phenomenon to another; it replaces one surprising phenomenon by another equally surprising phenomenon. (Friedman 1974, 18)

However, here it will be shown that unificationist views like Friedman's do not provide an adequate answer to the ERA, and do not provide an acceptable account of the relation between explanation and understanding.

---

[4]   This is especially true of (Friedman 1974).

[5]   The real worry behind the ERA appears to be that if we do not require that the explanans is itself explained, we will be forced to posit unexplained explainers to terminate the regress or we must argue that explanation is somehow generated *sui generis* by the explanation relation. In ignoring the ERA, and adopting a global view of explanation the unificationists seem to favor the second option.

It will be argued that in attempting to avoid the ERA and replace Hempel's covering-law model of explanation, unificationist views fail to connect explanation to understanding in any interesting, non-trivial, sense. The failure on the part of the unificationists to adequately link explanation to understanding non-trivially, is in part the result of the unrealistic epistemic nature of the unificationist view and the acceptance of two dogmatic views about the relationships between simplicity, truth and understanding. To put the point generally, the unificationist view of explanation requires epistemic agents and communities to possess computational resources that far exceed those that are available, and so cannot possibly yield a realistic account of scientific understanding or it leads to skepticism about explanation. Since neither of these views is acceptable, it will be suggested that the unificationist view should be rejected.

On a more positive note, it will be suggested that examination both of the ERA and of the criticisms of the unificationist view reveals that there are two distinct but equally important concepts of scientific understanding that should be distinguished. The first concept is that of *semantic understanding*. The second is that of *reduction*. When these different senses of 'understanding' are properly distinguished, it is possible to foresee the construction of a complete, non-skeptical and naturalistically acceptable account of explanation that can truly yield an acceptable account of scientific understanding.

2. Explanation as unification

Before turning to the critical appraisal of the unificationist view, we must look at the details of the unificationist theory of explanation. As already noted, the two most prominent versions of unificationism are those independently offered by Michael Friedman and by Philip Kitcher. So, first, we need to examine the basic details of these views with an eye to determining how Kitcher's and Friedman's specific unificationist views of explanation are supposed to yield scientific understanding, as per the understanding thesis. Note, however, that sections 2.3 and 2.4 contain presentations of the technical details, respectively, of Friedman's and Kitcher's views of explanatory unification and these sections can be skipped or skimmed over

for those who are not especially interested in the technicalities involved. Ultimately, as we shall see, what is most important is that we note that both Friedman and Kitcher subscribe to a core set of views about explanation that involves unification and simplification that is supposed to satisfy the understanding thesis. Let us then begin by looking at the informal versions of Friedman's and Kitcher's views of explanatory unification.

2.1. Friedman, Kitcher and unificationism

Friedman presents a version of the unificationist view that is based on the idea that what we are inclined to regard as being in need of explanation are laws and not ordinarily events, *pace* Hempel. More crucially, explanation of laws is achieved by showing that the phenomenon described by the target law are really just cases of some other phenomenon described by a more fundamental law. This is just the familiar relation of reduction, and Friedman candidly tells us that,

> The central problem for the theory of explanation comes down to this: what is the relation between phenomena in virtue of which one phenomenon can constitute an explanation of another, and what is it about this relation that gives understanding of the explained phenomenon. (Friedman 1974, 6)

So for unificationists like Friedman it would seem to be the case that explanation, and thereby understanding, is supposed to be achieved by reduction. Friedman motivates the consideration of unificationism as a serious option by rejecting as inadequate Hempel's covering-law model, and what he, respectively, calls the familiarity and intellectual fashions views of explanation.

The intellectual fashions view will be ignored here as it is irrelevant to the issue raised in this paper, but the familiarity view will play a more important role in the arguments to follow. Essentially the familiarity view holds that explanation is a matter of relating the unexplained to the explained by relating the unfamiliar to the familiar. Friedman's specific criticisms of these views will not be repeated here. However, in criticizing these two views, Friedman importantly establishes a set of three desiderata that any adequate theory of explanation should satisfy:

(DE1) The theory should be sufficiently general.

(DE2) The theory should be objective.

(DE3) The theory should connect explanation and understanding (Friedman 1974, 13-14).[6]

It is clear from the passage quoted in section 1 that Friedman is especially concerned to show that his own account of explanation is capable of satisfying DE3, but he is also careful to explain that,

> When I ask that a theory of scientific explanation tell us what it is about the explanation relation that produces understanding, I do not suppose that 'scientific understanding' is a clear notion. Nor do I suppose that it is possible to say what scientific understanding is in advance of giving a theory of explanation. It is not reasonable to require that a theory of explanation proceed by first defining 'scientific understanding' and then showing how its reconstruction of the explanation relation produces scientific understanding. We can find out what scientific understanding consists in only by finding out what scientific explanation is and vice versa. (Friedman 1974, 6)

This disclaimer is rather troubling, and this passage contains an anticipation of the criticism that unificationism only trivially satisfies DE3. Later we will return to this issue, but, at this point, it suffices to note that Friedman sees traditional views of explanation as all failing to satisfy one or more of DE1-DE3 and argues that unificationism does not suffer from these problems.

In the place of the more traditional theories of explanation, Friedman then offers us unificationism as a replacement. Friedman's unificationist theory depends on the following core intuition:

> 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 phenomena is, other things equal, more comprehensible than one with more. (Friedman 1974, 15)

---

[6]   DE3 is, of course, just the understanding thesis mentioned in the introduction.

Friedman's views about explanation then essentially include the following ideas. First, explanation is unification. Second, the best unification is the simplest one. Finally, explanation yields understanding because the simplest unification is the most understandable one. This last point is crucial to note. On Friedman's view the best unification is understandable because it requires us to accept the total systematization of our beliefs about some body of phenomena that involves the fewest primitive beliefs.

Having laid out the details of Friedman's view we can now turn our attention to Kitcher's view, and it is interesting to note that Kitcher's view was constructed as a sympathetic, but critical, reaction to the deficiencies of Friedman's unificationist analysis of explanation.[7] Kitcher's unificationism is then importantly similar to that of Friedman. Most crucially, it is clear that Kitcher and Friedman have (at least roughly) the same sort of idea in mind concerning the issue of how unification yields explanation, and thereby, understanding. Specifically, it is by the reduction of our set of our antecedently accepted beliefs, K, to the *simplest* systematization of those beliefs that is supposed to issue in greater understanding. This is ultimately due to the simplicity of the reducing set of beliefs. In his (1989) Kitcher explicitly introduces the following principle of acceptance for systematic unifications:

(U) S should be chosen over S' as the explanatory store over K, $E(K)$, just in case S has greater unifying power with respect to K than S' (Kitcher 1989, 477).

Essentially what Kitcher advocates is that we accept the global theory that *best* unifies, that best explains, the totality of our antecedent beliefs by showing that they can be derived from a small set of special argument patterns involving a small set of basic (axiomatically accepted) beliefs that, in part, constitute the reducing theory.

As with Friedman's view what we have here is a form of inference to the best explanation (IBE), where the quality, or 'bestness', of an explanation is to be measured in terms of its ability to simplify our global belief system and thus supposedly generate understanding. Kitcher explicitly tells us,

---

[7]   In fact, (Kitcher 1976) is intended to be an explicit criticism of Friedman's account of explanatory unification.

> I have sketched an account of explanation as unification, attempting to show that such an account has the resources to provide insight into episodes in the history of science and to overcome some traditional problems for the covering law model. In doing so, let me indicate very briefly how my view of explanation as unification suggests how scientific explanation yields understanding. By using a few patterns of argument in the derivation of many beliefs we minimize the number of *types* of premises we must take as underived. That is, we reduce, in a far as is possible, the number of types of facts we must accept as brute. Hence we can endorse something close to Friedman's view of the merits of scientific explanation. (Kitcher 1981, 529)

So, as with Fiedman's unificationism, such a reduction is supposed to make our global belief system more understandable by reducing the number of independent facts we must accept as basic, and so Kitcher also endorses DE3.

2.2. Friedman's formal account of unification[8]

For those interested in the technical details, Friedman gives this informal idea a formal treatment as follows. We are to regard a scientific community C, at a given time t, as accepting a set of law-like sentences K. In other words, K is the set of laws accepted in C at t. K is to be understood to be deductively closed such that if S is a law-like sentence and $K \vdash S$, then $S \in K$. We are then supposed to accept the systematization of K that reduces K to the smallest set of independently acceptable sentences, where independent acceptability is characterized as follows:

i. If $S \vdash Q$, then S is not acceptable independent of Q.

ii. If S is acceptable independently of P and $Q \vdash P$, then S is acceptable independently of Q. (Friedman 1974, 16)

Reduction is to be understood formally in the following manner. We first define a *partition* of the sentence S as the set of sentences Γ such that Γ is

---

[8]   As noted previously, readers who are not interested in the formal details of Friedman's view can skip this section.

logically equivalent to S and each S', where $S' \in \Gamma$, is acceptable independently of S. Moreover, a sentence will be *K-atomic* if it has no partition in the sense just defined, and a *K-partition* of a set of sentences Δ will be a set Γ of K-atomic sentences logically equivalent to Δ. The *K-cardinality* of a set Δ, abbreviated K-card (Δ), is defined as inf $\{$card (Γ): Γ is a K-partition of $\Delta\}$. We then get one of the key concepts of Friedman's unificationist view. *S reduces* the set Δ, if and only if, K-card $(\Delta \cap \{S\}) >$ K-card (Δ). He then tells us that what we really want reduced is the number of independently acceptable consequences of S, $\mathrm{con}_K(\mathrm{S})$.

Friedman then ultimately defines *explanation* as follows:

> (DI′) S_1 explains S_2 if and only if there exists a partition Γ of S_1 and an $S_i \in \Gamma$ such that $S_2 \in \mathrm{con}_K(S_i)$ and S_i reduces $\mathrm{con}_K(S_i)$. (Friedman 1974, 17)

On this basis, Friedman presents us with a relatively simple formal treatment of explanation as global unification that is supposed to yield understanding because global systematizations with fewer independent phenomena that we must understand primitively are more understandable. Presumably, we are supposed to accept the theory that best explains K for C at t, and the best explanation of K is that reduction of K in accord with DI´ that is globally the simplest reduction in the sense of simplicity just specified. Again, it is precisely because such unifications are simpler that they are supposed to be more understandable and this formal model shows how this works in terms of the concept of the simplest reduction.

2.3. Kitcher's alternative formal account of unificationism[9]

Kitcher too offers a formal account of the unificationist view of explanation, but it is presented in terms of a very different technical framework from that employed by Friedman. In any case, the main idea behind Kitcher's alternative version of unificationism is captured in the following passage:

---

[9]   As noted previously, this section can be skipped by those who are not interested in the formal details of Kitcher's view.

> The general problem I have set is that of specifying $E(K)$, the
> *explanatory store over K*, which is the set of arguments accepta-
> ble as the basis for acts of explanation by those whose beliefs are
> exactly the members of K. (For purposes of this paper I shall
> assume that, for each K there is exactly on $E(K)$.
>
> The unofficial view [unificationism] answers the problem: for
> each K, $E(K)$ is the set of arguments which best unifies K. My
> task is to articulate this answer. (Kitcher 1981, 512)

It is important to note that Kitcher is focusing on the unification of K by appeal to a set of general argument patterns, and argument patterns are not argument forms in the sense employed in modern formal logic. As Kitcher understands them, argument patterns are more context-specific and less abstract than logical argument forms. Moreover, the reducing theory is not merely a set of sentences as in the case of Friedman's view. According to Kitcher each theory is constituted, in part, by a set of accepted argument styles, or canonical argument patterns that are regarded as 'good'.

Kitcher elaborates on this idea, and tells us that a *generating set* Σ is a set of argument patterns Π such that each argument in Σ instantiates a pattern in Π. A generating set for Σ is *complete with respect* to K, if and only if, every argument which is acceptable relative to K and which instan-tiates a pattern in Π is a member of Σ. To select the explanatory store $E(K)$ we first narrow our choice to those sets of arguments which are acceptable relative to K, the possible systematizations of K. From among the various generating sets of argument patterns that are complete with respect to K, in accordance with U we select that set with the greatest unifying power and we call the selected set the *basis* of the set of arguments in question. $E(K)$, the explanatory store over is K, that systematization whose basis does best by the criteria of unifying power (Kitcher 1981, 519-20). Im-portantly, this includes the idea that the best systematization is that which is simplest. So, let us then turn our attention to the central concept Kitcher's account explanation, the concept of unifying power.

Kitcher begins his explication of the concept of explanation as unifica-tion by offering a series of related definitions that are to be used in defining what he refers to as a general argument pattern. A *schematic sentence* is an expression generated by replacement of at least one of the non-logical

expressions in the sentence with dummy letters. For each schematic sentence there must also be a set of *filling instructions* that specify how the dummy letters are to be replaced. A *schematic argument* is then a set of schematic sentences, and a *classification* for such an argument is a set of sentences describing the inferential features of the schematic argument. The classification, in effect, identifies which sentences are premises, which sentence is the conclusion, what logical rules of inference are used, etc. Kitcher then defines a *general argument pattern* as a sequence of sentences such that:

a. The sequence has the same number of terms as the schematic argument of the general argument pattern.

b. Each sentence in the sequence is obtained from the corresponding schematic sentence in accordance with the appropriate filling instructions.

c. It is possible to construct a chain of reasoning which assigns to each sentence the status accorded to the corresponding schematic sentence by the classification (Kitcher 1981, 516-17).

The task that Kitcher then undertakes is to explicate the concept of the unifying power of a set of argument patterns so that we can determine which set is the best explanatory unification of K.

After proposing that the unifying power of a set of argument patterns should be defined as the ability of those argument patterns to generate "…a large number of accepted sentences as the conclusions of acceptable arguments which instantiate a few, stringent patterns (Kitcher 1981, 520)," Kitcher points out that this suggestion will not work. Rather mysteriously, he then abandons the attempt to specify the concept of unifying power more precisely. Nonetheless, he does give us some idea of what he has in mind. First we define the conclusion set of a set of arguments Σ, as the set of sentences which occur as conclusions, $C(\Sigma)$, of some argument that is a member of Σ. Kitcher then suggests that the unifying power of a reduction base B_i with respect to K varies with the size of $C(\Sigma_i)$ and the stringency of the patterns of B_i, and inversely with the number of members of B_i (Kitcher 1981, 520). The idea is to select the smallest set of premises that generates the largest conclusion set with respect to our belief system.

The concept of stringency is itself left undefined, but Kitcher explains that the stringency of an argument pattern is determined by the strictness of the conditions governing the substitution of dummy letters in argument patterns and conditions governing acceptable logical structure (Kitcher 1981, 518). The former are fixed, for the most part, by the filling instructions, the latter by the classification. Considering the stringency of sets of argument patterns is introduced in order to rule out cases both where all argument patterns are acceptable relative to K and where only one unique argument is acceptable relative to K. In any case, Kitcher admits that this account of unifying is too simple and quite vague, but he offers nothing more well-developed as a replacement for this account of unifying power. Kitcher does note that, in addition to counting numbers of argument patterns, the concept of unifying power probably also ought to include in some way the similarities among such argument patterns, but he offers no deeper account of this concept.

2.4. Kitcher, Friedman and the common basis of unificationism

The unificationists we have considered (who are without any doubt the most well-known defenders of unificationism) hold that the best explanation, the one that we ought to accept, is the simplest systematization of our belief corpus. As it happens, this belief system is also supposed to be the one that affords us the most understanding. Kitcher and Friedamn are then clearly advocating a sophisticated form of IBE (or what Peirce variously called abduction or the method of hypothesis) on a global scale. More to the point, on their view the 'bestness' of an explanation is understood in terms of a kind of simplicity as it applies to reduction bases.[10] On this general point about the virtues of simplicity they are, in fact, in broad agreement with Peirce himself who often stressed the role of economy in abductive inference and in science in general, as is the case with most defenders of IBE.[11] But, they apply this notion of simplicity to complete belief systems as opposed to localized sets of beliefs. Consequently, the,

---

[10]   We are here ignoring the differences between Peirce's conceptions of abduction and retroduction as developed in his later views.

[11]   See, for example, Peirce (c. 1901/1931-1958, 6.529-6.5230 and 7.220).

the generic rule of theory acceptance such unificationists appear to endorse is:

(URA)   Accept the best explanation of a total system of beliefs where the best explanation is the globally simplest unification of a global system of beliefs/sentences.

It is important to note, however, Peirce does not in any obvious way, seem to be committed to the idea that abduction is global in this sense, and this is the source of considerable trouble for unificationism.[12] In what follows, we will find that there is some reason to suspect that this kind of global IBE is not a computationally feasible form of inference. However, before turning our attention to the critique of unificationism, let us first look at one of its supposed successes.

2.5. The explanatory regress argument and unification

One of the virtues that unificationists have claimed for their theory is the avoidance of the ERA. As we saw in section 1, Friedman is especially clear about this virtue of his version of unificationism. His main objection to familiarity views like that defended by Dray (1964) and, to some degree, by Scriven (1970), is that we often explain laws by appeal to other laws with which we are less familiar. If this is so and we believe that these sorts of cases are *bona fide* instances of explanation, then explanation cannot be relating the unfamiliar to the familiar. Unification does not require relating the unfamiliar to the familiar and supposedly secures the connection between explanation and understanding by linking explanation with unification, and thereby simplicity. The unificationists avoid the ERA by ignoring it and simply jettisoning the requirement that we are more familiar with the explanans material than we are with the explanandum material.

Friedman is correct on this count. However, for the unificationists global unification in the simplest system then replaces any consideration of the kind of understanding discussed by familiarity theorists, but it is here where the unificationists go wrong. We *can* explain phenomena by relating them

---

[12]   Some problems with treating inference to the best explanation globally are discussed in (Shaffer 2012) and in (Fodor 2000).

to unfamiliar theories. Simply consider most classical explanations of microscopic phenomena by quantum mechanics (QM). Friedman is correct that we are most assuredly *not* more familiar with the relevant aspects of the QM explanations. Nevertheless, we must possess an important kind of understanding of the reducing theory if such an explanation is to yield real understanding and so the familiarity theorists are correct to take the ERA seriously. In the particular case of QM, it is plausible to believe now that this principle is satisfied. *Pace* Friedman, it is not the *unfamiliarity* of reducing theories that is the real issue, but rather the issue is whether we *understand* the reducing theory in some relevant sense. On the one hand, all novel theories are unfamiliar, but his need not entail that they are not understood. On the other hand, familiar theories may not be such that they are understood. As such, it is a critical mistake to merely equate understanding with familiarity, and thus ignore the ERA as Friedman does.[13]

3. The failure of unificationism

So where does unificationism go wrong? As outlined in the introduction, unificationism fails for several closely connected reasons. First, in attempting to avoid the ERA unificationism places great, unwarranted, faith in the metaphysical view that our world is ultimately simple and in the view that the world can be successfully explained by a simple global theory. This is important because these particular articles of faith must be true for it to be possible for us to cognitively grasp entire, global, scientific belief systems (i.e. belief corpuses) that potentially apply to an ultimately comprehensible world. All of this must be the case because unificationists require that we be able to *compare* the simplicity of such explanatory systems and accept the simplest systematization of the observed facts about the world. But, it is not at all clear that our world is simple in this sense, and it is not computationally and epistemically feasible that we could (even as a community)

---

[13]   The unificationists do not really avoid the ERA anyway, as they do not offer an account of how the basic beliefs are themselves understood.

compare global belief systems in the manner that unificationists require, even if global belief systems turn out to be relatively simple.

Supposing that we could manage to avoid these metaphysical and computational problems, it will be shown here that, *pace* the unificationists, the simplest unified belief corpus may not provide us with understanding in the relevant sense of the term and that the resulting unified system may not even be likely to be true. So, firstly, it will be argued here that unification is not sufficient for understanding and that the unificationists' account of explanation does not track the truth. In other words, given the unificationist view of explanation, it is perfectly possible that the best explanation (i.e. the simplest global unification of our scientific beliefs) is not understandable at all and may well be false. By DE3 (and DE2), it then follows that unificationism should be rejected if actual human understanding is factive. In any case, let us now turn our attention to the details of these criticisms of unificationism.

3.1. Two dogmas of simplicity

As the expositions of Friedman's and Kitcher's views show, the essential motivation behind unificationism appears to involve a particular notion of reduction that relates explanation to simplicity, and thereby explanation to understanding. This unificationist argument essentially involves three steps. The first step (S1) is to define explanation and best explanation in terms of the unification of belief systems. The second step (S2) is to note that such unification implies simplification. Finally, the third step (S3) is to argue that simpler theories are more understandable. The various elements of this line of reasoning are found in Kitcher's and Friedman's claims that we have examined in earlier sections and it is obvious that there is broad agreement between Friedman and Kitcher in this respect. This argument, however, depends on at least two crucial assumptions. First, the unificationists appear to accept the view that, other things equal, a numerically smaller system of beliefs and/or inference schemes is more understandable. Second, as Friedman and Kitcher both do not appear to be skeptics and seem to believe that we can successfully (i.e. truthfully) explain phenomena and thereby understand the world, it must be the case that they believe

that the world really is, at least relatively, simple in some ontological sense.[14]

These assumptions are what we will call the *two dogmas of simplicity*. This first view, that a numerically smaller belief system is more understandable, is the *dogma of epistemic simplicity* (DES). The second dogma, the view that the world really is simple, as the *dogma of metaphysical simplicity* (DMS). As we shall see, both are mere dogmas and neither is non-controversially true. Nevertheless, their general acceptance–tacit or otherwise–has made unificationism rather more appealing than it otherwise would be, especially in light of their ignoring the ERA. DES is a crucial assumption behind S3 of the unificationist argument and DEM is, similarly, crucial for justifying S1 and S2.

3.2. The dogma of metaphysical simplicity

The unificationists' implied view that the world is simple appears either to be a matter of blind faith or it is an *a priori* matter, for surely it is not clearly the case on empirical grounds. However, as it is obvious that the fact of the matter concerning the simplicity of the world cannot be an *a priori* issue, the unificationists' assumption that the world is fundamentally simple appears merely to be a dogmatic assumption. In fact, empirical considerations seem to support the view that our world is rather complex, and that the world is divided into more than one independent domain characterized by entirely different, non-reducible, laws. Therefore, as neither appealing to faith nor to *a prioricity* is, at least in this case, legitimate, it seems that there is then no real substantive reason to suspect that our world really is simple.[15] Hence, we should have serious doubts that simple theories are more likely to be true, and this raises serious problems for unificationists.

Consider the following conditions of adequacy concerning explanation and IBE, all of which the unificationists appear to accept (implicitly or otherwise):

---

[14]   If this were not the case, then it is unclear what the purpose of the many examples of allegedly successful explanation are.

[15]   See (Fodor 1974; Dupré 1983; Cartwright 1999; and Shaffer 2012) for related points.

(CA1) The explanans of a *bona fide* explanation must bear some relation to the truth.[16]

(CA2) There is some connection between an explanation being the *best* explanation and that explanation's truth; the best explanation is likely to be true and,

(CA3) The 'bestness' of explanation is a function of its simplicity.[17]

If there is some connection between an explanation being the best explanation and that explanation's truth in this way (i.e. if CA2 and CA3 are correct in a way that allows CA1 to be satisfied), then, unless there is some non-dogmatic reason to believe DMS, the unificationists' view of the connection between unification and truth of explanation cannot be adequate. This is because they offer no reason to accept CA2 and CA3 except that doing so allows us to trivially link explanation and understanding in order to satisfy DA3. However, as it is not obviously the case that DMS is true and we do not have any justifications for accepting CA2 and CA3, we have no reason to suspect that CA1 is satisfied by the unificationists' views of explanation. As such, it is neither obviously true that simpler, more unified, global belief systems are explanatory, nor is it obviously true that simpler, more unified, global systems of belief are more likely to be true. So S1 and S2 of the unificationist argument are thus impugned.

3.3. The dogma of epistemic simplicity

In order to support S3 the unificationists argue that we ought to accept the simplest systematization of our scientific beliefs where the simplest systematization is the belief system that reduces the number of phenomena or inference patterns describing the phenomena to the smallest number by unification. As we have seen, the putative reason for doing so is that the simplest globally unified system of this sort is supposed to be the most understandable. Essentially, they believe that a system that is reductively simplified by unification is supposedly more understandable. This is an

---

[16]   As is well known, Hempel and Oppenheim (1948) argue that the explanans must be true.

[17]   Lipton (1991) and Thagard (1988) defend these conditions of adequacy for IBE.

article of pure epistemological faith (another dogma), and it is a view that
is easily exploded if we take the ERA seriously, as the unificationists do
not.

It seems obvious that a system that is unified in the unificationists' sense
may be completely incomprehensible. Simply consider a belief system that
is unified in this manner, but which reduces the apparent plethora of phe-
nomena to a unified but relatively small (axiomatic) reduction base that is
not itself understood at all. Should we really say that such a system provides
us with an explanation that yields understanding? Surely not. The following
well-understood sort of case demonstrates this and history is replete with
examples of such non-explanatory reductive thinking. Suppose that a cog-
nitive agent or a cognitive community attributes to all phenomena a rela-
tively unified, but inscrutable, set of causes: say a sophisticated kind of *deus
ex machina* reduction base.[18] This is just the sort of familiar but bad reduc-
tive reasoning involved in forms of supernaturalism and mysterianism. It
amounts to the contention that some suitably unified inscrutable source(s)
of power accounts for all observable phenomena. Such reductions do not
explain and they do they generate understanding. Second, consider a case
where we have a unified global reduction base–for example that proposed
by defenders of string theory–the basic concepts of which are arguably of
greater internal (semantic) complexity than the disparate phenomena it is
supposed to unify. In this particular case, the various phenomena associated
with the general theory of relativity with those of quantum mechanics are
supposed to be unified by reduction, but the concepts/argument patterns
of string theory, although numerically fewer in number, appear to be of
much greater mathematical/conceptual complexity than either of the theo-
ries it reduces. It is at least plausible then to suggest that we should not
count such a reduction as yielding understanding in any serious manner.

---

[18]   Kitcher considers the theological version of this response in his (1980), and
concludes that it can be rejected on the basis that it fails to be adequately stringent.
However, he offers no substantive account of the stringency of argument patterns.
In any case, we can imagine arbitrary non-theological cases that employ sufficiently
stringent argument patterns the component terms of which we do not understand at
all. Behe's (1996) intelligent design theory might well be a theoretical unification of
this sort.

Finally, consider a reduction base that is composed of a small set of inconsistent propositions. From these inconsistent axioms we can trivially derive all other propositions in our belief corpus, but we surely would not want to say that such a set of propositions is coherently understood in some brute manner that explains the propositions derived from them and generates understanding. Such a reduction base cannot even be true, let alone allow us to understand the propositions derived from them.

In the first case we have a reduction base that is numerically small, but which, intuitively, yields no understanding due to its incomprehensibility, and in the second case we find that reduction in the number of factors accepted as primitive in a theory may, ultimately, make for a more conceptually complex (i.e. semantically complex) and less understandable global theory. In the third case we have a reduction in the number of brute facts accepted and a derivation of all other propositions we accept, but no understanding of them because the reduction base is inconsistent. Cases of these sorts can easily be multiplied and articulated, but the details are not really important here. What is important is that cases one and two can be avoided only if the unificationists, by stipulation, simply *equate* understanding with unification in order to obviate S1–S3 and thereby satisfy DE3 trivially as a matter of definition. Given that this is clearly not an acceptable way to justify the steps of the unificationist argument, it seems then, at very least, to be the case that unification is not clearly a sufficient condition for scientific understanding. Cases of the third sort are trickier to avoid and this can be done only by requiring that we eliminate inconsistent reductions from the set of potentially acceptable unifications in some principled manner. As we shall subsequently see, however, this is not realistically possible.

Unification cannot be a necessary condition of explanation for naturalists without leading to outright skepticism, and Kitcher (1981, 1983, 1992, 1993), in particular, is an avowed naturalist.[19] Non-naturalists like Friedman (2000, 2001) are faced with a related, but different, problem. Such, non-naturalists can treat unification as a necessary condition of explanation and avoid skepticism only by attributing *a priori* faculties to us that exceed

---

[19]   Friedman, however, is an avowed non-naturalist as his (2000) and (2001) indicate.

well-verified computational constraints on any feasible inference procedure. How do these particular problems arise? First, consider what the unificationists' view of explanation would require of cognitive agents or cognitive communities. What they must be able to do is to compare total, global, unifications of our knowledge corpus, and then accept the one that is the simplest (in terms of some specified criterion of reductive simplicity) from among the array of possibilities. It turns out that such a task is not computationally satisfiable in realistic times for even finite sets of finite and very small systems of beliefs, even with the aid of all available computational resources.[20] In fact, even if we accept the dubious assumption that the linguistic representations of particular belief corpuses are finite in size and that we need only consider a finite set of such systems, we cannot check the consistency of relatively simple systems with available computational resources in feasible times, let alone check the consistency of and compare *all logically possible* systematizations of our beliefs in terms of their overall simplicity. However, we must be able to check the consistency of such systems or URA might well–in violation of CA1 and CA2–sanction our accepting very simple but inconsistent systematizations of our beliefs as explanatory. As we have seen, this is because all of our beliefs can trivially be derived from inconsistent sets of such axioms. To avoid such cases we must first delimit the (infinite) set of possible alternative systematizations to those that are logically consistent, and then we are supposed to select that consistent systematization from among the remaining set that fares best in terms of URA. However, it is not even *physically possible* to check for the consistency of such systematizations, let alone assess the comparative global simplicity of an infinite set of consistent systematizations of our beliefs.[21]

Given these deeply troubling (but well-known) facts about the computational features of belief systems and URA, unificationists like Kitcher who accept naturalism must be skeptics. This is because we cannot apply URA

---

[20]   See (Kornblith 1989) and (Cherniak 1986) for explicit consideration of such computational restrictions on epistemic processes. Also, see (Harman 1986) for discussion of the computational problem that arise for belief systems of infinite size. It is crucial to note that any belief system closed under logical consequence will be infinite.

[21]   (Cherniak 1986) is the canonical source on this point.

in practice in such a way that it can be satisfied and so we cannot really ever explain anything. Given these same facts, unificationists like Friedman, who accept some form of non-naturalism, must attribute to us almost occult *a priori* epistemic faculties that exceed computational/mathematical constraints on feasible, or even physically possible, procedures in order to allow for the satisfaction of URA. Both approaches are obviously unacceptable, and so, at very least, unificationism is either utterly unrealistic or deeply committed to skepticism. Consequently, unification is not a plausible necessary condition for explanation.

4. Two types of understanding

As mentioned in the introduction, it appears to be the case that two senses of 'understanding' are often conflated in discussions of scientific explanation. Not surprisingly, the conflation of these two concepts has led to considerable misunderstanding on the part of the parties to the traditional debate concerning scientific explanation and, more recently, to the debate concerning how explanation relates to understanding. However, the ERA provides us with great insight into this error, and this is why it is important that we take the ERA seriously. It just will not do do to sweep it under the rug as traditional and unworthy of serious consideration as the unificationists seem to do.

Ignoring the ERA has made the unificationist view appear to be rather more plausible than it really is. Moreover, this myopia is the root cause of the kinds of problems that afflict the unificationist view raised here. In effect, what unificationists appear to have done by ignoring the ERA is to equate (unwarrantedly and by mere stipulation) scientific understanding with simplifying reduction and simplifying reduction with explanation, thereby trivially satisfying DA3 (i.e. the understanding thesis). In doing so they ignore the concept of *semantic understanding* and this is the concept that is the real core of scientific understanding. Simply put, scientific understanding is not wholly a matter of reductive explanation. Thus, the unificationist view does not satisfy one of the constraints on theories of explanation that the proponents of unificationism themselves accept.

4.1. Reductive understanding and semantic understanding

So, in ignoring the ERA and accepting both the DES and the DMS without justification, the unificationists have overlooked one of the key concepts constitutive of scientific understanding, semantic understanding. This point is related to Heisenberg's observation that,

> For an adequate understanding of the phenomena, the first condition is the introduction of adequate concepts; only with the help of the correct concepts can we really know what has been observed. When we enter a new field, very often new concepts are needed, and these new concepts come up in a rather unclear and underdeveloped form. Later they are modified, sometimes they are almost completely abandoned and replaced by better concepts which then, finally, are clear and well-defined. (Heisenberg 1989, 19)

The relevant point in this passage (that we must have adequate concepts to achieve understanding) is made more poignant when coupled with Heisenberg's view that theoretical formalisms that describe phenomena are distinct from the concepts that, at least in part, are required for understanding of the phenomena in question.[22] For Heisenberg, scientific understanding involves some form of understanding of the concepts that allow us to conceive of the phenomena in terms of some theoretical (i.e. mathematical) formalism. However, from this passage it is clear that Heisenberg believes that understanding comes in degrees and that, often, we adopt a theory without adequate concepts in hand sufficient to generate complete understanding. Ultimately, inadequate and imprecise concepts are replaced by adequate and precise concepts in order to secure scientific understanding of the phenomena in question. We *can* employ a theory in practice without complete understanding of the theory, but then the theory should not be considered to be fully explanatory. It is only when we come to grasp the complete *meanings* of the fundamental terms of a theory adequately, no matter how globally simple the theory is, that we have a *bona fide* explanation of the phenomena it describes.[23]

---

[22]   See (Heisenberg 1930), especially chapters 1-3 and the appendix.

[23]   See (Radder 1991; Shaffer 2008a; and 2008b) on this matter.

So, it appears to be the case that a form of semantic understanding is needed here and that semantic understanding involves a form of truth-conditional semantics.[24] As such, a theory of explanation that incorporates this concept seems as if it might do better in accounting for the connection between truth, explanation, and understanding than unificationism does. That the unificationists have overlooked this concept is, again, primarily due to their ignoring the ERA and due to their tacit acceptance of a holistic version of conceptual, or linguistic, role semantics.[25] Given these sorts of views about semantics, the unit of meaning is the complete linguistic or conceptual system and the holistic meaning of such a system is exclusively a function of the logical relations between the constitutive elements of such a system. Both Friedman and Kitcher appear to believe that understanding is a matter of reduction and involves a particular, broadly logical, relation between the sentences that constitute a total theory. They do not appear to accept that understanding involves anything more.[26]

This is a dubious assumption from the perspective of semantics, and it is a plausible way to link explanation to understanding *only* if it is legitimate to ignore the ERA.[27] If we resist this temptation and–recalling our the discussion of DES–recognize that numerical smaller systems of sentences need not be more *semantically* understandable, then we should conclude that unificationism does not, non-trivially and non-skeptically, connect explanation and understanding. Given this observation about the incompleteness of the unificationists' views of understanding, we can then see, at least schematically, what an adequate theory of how explanation relates to scientific understanding will look like.

First, as explanation in general and IBE in particular, must have some truth connection in accord with CA1-CA3, and, as the ERA suggests, we must include *both* semantic understanding (which is truth-conditional) and reductive understanding (which is more holistic) in our account of scientific

---

[24]   Note, however, that there are problems with the computational aspects of some forms of truth-conditional semantics as well. See, for example, (Shaffer 2019).

[25]   See, for example, (Harman 1982).

[26]   If it does, then they *must* face the ERA with respect to the basic, axiomatic, beliefs in a unified systematization.

[27]   See (Fodor and Lepore 1991; and 1992).

understanding. Second, as we are primarily concerned with *human* understanding (which is limited) and should take pains to avoid skepticism about explanation, we must be sure that our theories of both reductive and semantic understanding are *both* computationally tractable.[28] As such, naturalistic study of human computational abilities seems to be a required for the construction of a realistic theory of scientific understanding. Considerations of global simplicity in the sense employed by the unificationists should then be rejected in favor of consideration of local notions of simplicity and this must be conjoined with consideration of the meanings of the basic facts in the reducing theory. So a local body of phenomena is explained by reducing it to a relatively unified but cognitively graspable theory, the basic terms of which are semantically understood. But, science so understood, is a piecemeal operation which often involves integrating local explanatory theories with other previously established explanations. Understanding and explanation are then contextual, come in degrees and are often partial.[29] In real scientific practice, we often only partially grasp and apply a given explanatory theory to a restricted phenomenal domain. Of course, there is much more to be said about what semantic understanding is and how explanatory integration works, but the theory of explanation suggested here promises the possibility of success where unificationism fails.

5. Return to Peirce: local abduction and feasible explanation

There are some clear lessons to be learned from the failure of unificationism, and chief among these are the following results: (a) naturalistic epistemologists ought to reject *global* accounts of explanation, (b) explanation and scientific understanding are, at least in part, fundamentally semantic phenomena, and (c) all realistic accounts of explanation must involve serious consideration of the concept of simplification. A theory of explanation that rejects occult epistemology ought then to be constructed in such a way that it conforms to these desiderata, and it will look quite a bit

---

[28]   See (Shaffer 2019) about some concerns on this point with respect to possible worlds semantics.

[29]   See (Shaffer 2012) on the partiality and contextuality of explanation.

more like Peirce's original theory of local abduction than it will look like contemporary unificationist views such as those defended by Friedman and Kitcher. Some foundational work has been done here with respect to such an account, but there is much more work to be done in fleshing out the core concepts of an adequate account of explanation and scientific understanding that reflects these considerations. Amongst the most important tasks then to be undertaken are the tasks of fleshing out of a truth-conditional concept of semantic understanding, the determination of the limits of realistic globalization in our belief systems, and determining how simplification, explanation, and scientific understanding are related with respect to potentially integratable local explanations.

References

Behe, Michael. 1996. *Darwin's Black Box*. New York: Simon and Schuster.

Cartwright, Nancy. 1999. *The Dappled World*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139167093

Cherniak, Christopher. 1986. *Minimal Rationality*. Cambridge: M.I.T. Press.

De Regt, Henk. 2017. *Understanding Scientific Understanding*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190652913.001.0001

De Regt, Henk, Sabina Leonelli, and Kai Eigner, eds. 2009. *Scientific Understanding*. Pittsburgh: University of Pittsburgh Press.

Dray, William. 1964. *Laws and Explanation in History*. New York: Oxford University Press.

Dupré, John. 1983. "The Disunity of Science." *Mind* 42 (367): 321-46. https://doi.org/10.1093/mind/XCII.367.321

Feigl, Herbert. 1970. "The 'Orthodox' View of Theories: Remarks in Defense as Well as Critique." In *Minnesota Studies in the Philosophy of Science* Vol. 4, edited by Michael Radner and Stephen Winokur, 3-16. Minneapolis: University of Minnesota Press.

Fodor, Jerry. 1974. "Special Sciences, or the Disunity of Science as a Working Hypothesis." *Synthese* 28 (2): 77-115. https://doi.org/10.1007/BF00485230

Fodor, Jerry. 2000. *The Mind Doesn't Work That Way*. Cambridge: M.I.T. Press.

Fodor, Jerry and Ernest Lepore. 1991. "Why Meaning (Probably) Isn't Conceptual Role." *Mind and Language* 6 (4): 328-43. https://doi.org/10.1111/j.1468-0017.1991.tb00260.x

Fodor, Jerry and Ernest Lepore. 1992. *Holism: A Shopper's Guide.* Oxford: Black-well.

Friedman, Michael. 1974. "Explanation and Understanding." *The Journal of Philosophy*, 71 (1): 5-19. https://doi.org/10.2307/2024924

Friedman, Michael. 2000. "Transcendental Philosophy and A Priori Knowledge: A Neo-Kantian Perspective." In *New Essays on the A Priori*, edited by Paul Boghossian and Christopher Peacocke, 367-83. New York: Oxford University Press. https://doi.org/10.1093/0199241279.003.0015

Friedman, Michael. 2001. *The Dynamics of Reason.* Stanford: CSLI.

Harman, Gilbert. 1982. "Conceptual Role Semantics." *Notre Dame Journal of Formal Logic* 23 (2): 242-56. https://doi.org/10.1305/ndjfl/1093883628

Harman, Gilbert. 1986. *Change in View.* Cambridge: M.I.T. Press.

Heisenberg, Werner. 1930. *The Physical Principles of the Quantum Theory.* Chicago: University of Chicago Press.

Heisenberg, Werner. 1989. *Encounters with Einstein.* Princeton: Princeton University Press.

Hempel, Carl. 1965. *Aspects of Scientific Explanation.* New York: The Free Press.

Hempel, Carl. 1966. *Philosophy of Natural Science.* Englewood Cliffs, N.J.: Prentice-Hall.

Kitcher, Philip. 1976. "Explanation, Conjunction and Unification." *Journal of Philosophy* 73 (8): 207-12. https://doi.org/10.2307/2025559

Kitcher, Philip. 1980. "A Priori Knowledge." *The Philosophical Review* 86 (1): 3-23. https://doi.org/10.2307/2184861

Kitcher, Philip. 1981. "Explanatory Unification." *Philosophy of Science* 48 (4): 507-31. https://doi.org/10.1086/289019

Kitcher, Philip. 1983. *The Nature of Mathematical Knowledge.* New York: Oxford University Press. https://doi.org/10.1093/0195035410.001.0001

Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World" In *Minnesota Studies in the Philosophy of Science* Vol. 13, edited by Wesley Salmon and Philip Kitcher, 410-505. Minneapolis: University of Minnesota Press.

Kitcher, Philip. 1992. "The Naturalists Return." *The Philosophical Review* 101 (1): 53-114. https://doi.org/10.2307/2185044

Kitcher, Philip. 1993. *The Advancement of Science.* New York: Oxford University Press. https://doi.org/10.1093/0195096533.001.0001

Kornblith, Hilary. 1989. "The Unattainability of Coherence." In *The Current State of the Coherence Theory*, edited by John Bender, 207-14. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-2360-7_19

Lambert, Karel. 1980. "Explanation and Understanding: An Open Question?" In *Rationality in Science*, edited by Risto Hilpinen, 29-34. Boston: D. Reidel.

Peirce, Charles S. c.1901/1931-1958. *Collected Papers of Charles Sanders Peirce.* C. Hartshorne, edited by Paul Weiss, and Arthur Burks, 8 vols. Cambridge: Harvard University Press.

Radder, Hans. 1991. "Heuristics and the Generalized Correspondence Principle." *The British Journal for the Philosophy of Science* 42 (2): 195-226. https://doi.org/10.1093/bjps/42.2.195

Salmon, Wesley. 1989. "Four Decades of Scientific Explanation." In *Minnesota Studies in the Philosophy of Science* Vol. 13, edited by Wesley Salmon and Philip Kitcher, 3-195. Minneapolis: University of Minnesota Press.

Scriven, Michael. 1970. "Explanations, Predictions and Laws." In *Minnesota Studies in the Philosophy of Science* Vol. 3, edited by Herbert Feigl and Grover Maxwell, 170-230. Minneapolis: University of Minnesota Press.

Shaffer, Michael. 2001. "Bayesian Confirmation of Theories that Incorporate Idealizations. *Philosophy of Science* 68 (1): 36-52. https://doi.org/10.1086/392865

Shaffer, Michael. 2002. "Coherence, Justification, and the AGM Theory of Belief Revision." In *Perspectives on Coherentism*, edited by Yves Bouchard, 139-60. Ontario, Canada: Aylmer-Éditions du Scribe.

Shaffer, Michael. 2008a. "Idealization, Counterfactuals and the Correspondence Principle." In *The Courage of Doing Philosophy: Essays Dedicated to Leszek Nowak*, edited by Jerzy Brzezinski, Anderzj Klawiter, Theo Kuipers, Krzystof Lastowski, Katarzyna Paprzycka, and Piotr Przybysz, 179-204. Amsterdam: Rodopi.

Shaffer, Michael. 2008b. "Re-formulating the Generalized Correspondence Principle: Problems and Prospects." *Polish Journal of Philosophy* 2: 99-115. https://doi.org/10.5840/pjphil2008217

Shaffer, Michael. 2012. *Counterfactuals and Scientific Realism.* New York: Palgrave-MacMillan.

Shaffer, Michael. 2019. "Safety, the Preface Paradox and Possible Worlds Semantics." *Axiomathes* 29 (4): 347-61. https://doi.org/10.1007/s10516-018-9413-3

Woodward, James. 2017. "Scientific Explanation." *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/