

# Explanation = Unification? A New Criticism of Friedman's Theory and a Reply to an Old One

William Roche and Elliott Sober\*†

---

According to Michael Friedman's theory of explanation, a law  $X$  explains laws  $Y_1, Y_2, \dots, Y_n$  precisely when  $X$  unifies the  $Y$ 's, where unification is understood in terms of reducing the number of independently acceptable laws. Philip Kitcher criticized Friedman's theory but did not analyze the concept of independent acceptability. Here we show that Kitcher's objection can be met by modifying an element in Friedman's account. In addition, we argue that there are serious objections to the use that Friedman makes of the concept of independent acceptability.

---

**1. Introduction.** Is unification the key to understanding what a scientific explanation is? This idea has been developed creatively by Michael Friedman (1974) and Philip Kitcher (1981, 1989), both of whom take their inspiration from Kant's idea that scientific understanding is achieved via the unification of diverse phenomena. In this article, we examine Friedman's development of the idea that law  $X$  explains laws  $Y_1, Y_2, \dots, Y_n$  if and only if  $X$  unifies the  $Y$ 's. The special spin he gives to this idea is his use of the concept of one sentence's being acceptable independently of another.<sup>1</sup>

Received January 2016; revised July 2016.

\*To contact the authors, please write to: William Roche, Department of Philosophy, Texas Christian University, Fort Worth, TX; e-mail: w.roche@tcu.edu. Elliott Sober, Department of Philosophy, University of Wisconsin–Madison, Madison, WI 53706; e-mail: ersober@wisc.edu.

†We are grateful to Nicholaos Jones, Joshua Smith, and Neil Tennant for useful discussion. We also thank the anonymous referees of this journal for their helpful queries and comments.

1. Kitcher's very different theory of explanation as unification is based on the concept of an argument pattern. For criticisms of his theory, see Barnes (1992a), Humphreys (1993), Skipper (1999), Woodward (2003, 2014), Griesmaier (2005), and Gijsbers (2007).

Philosophy of Science, 84 (July 2017) pp. 391–413. 0031-8248/2017/8403-0001\$10.00  
Copyright 2017 by the Philosophy of Science Association. All rights reserved.

Kitcher (1976) develops an objection to Friedman's theory that is widely taken to be devastating, so it may be asked why new objections are worth examining. One reason is that Friedman's theory can be modified to evade Kitcher's objection (as we explain in sec. 3). Another is that philosophers sometimes think that Friedman's basic idea concerning explanation, unification, and independent acceptability is on the right track even though his technical implementation of that basic idea is flawed. This is true, for example, of Psillos, who writes: "There may be a way to defend the spirit of Friedman's model of unification. . . . I want to motivate my suggestion by noting that the basic idea behind Friedman's program is right: explanation *does* amount to reduction of the number of independently acceptable laws. The crux of this idea is that all apparently independently acceptable regularities are consequences of more basic laws" (2002, 271–72). Friedman (1983, 261–62) himself took that position in the aftermath of Kitcher's criticism. The concept of independent acceptability is therefore worth examining. Friedman depends on this concept but offers little by way of elucidation. Kitcher's criticism of Friedman does not focus on that concept at all.<sup>2</sup>

Before setting to work, we want to mention two reasons for being skeptical about unificationist approaches to explanation. First, it seems implausible that an explanation of two or more events must unify them. True, the motion of the moon and the motion of an apple falling to earth are explained by Newton's unifying theory, but that is just one favorable example. Consider, in contrast, the fact that this rose is red and so is that fire hydrant. The right explanation here is disunifying. The shared color is a coincidence, but that does not mean that it is inexplicable. The explanation is in terms of separate independent causes, not common causes (Sober 2012). Notice that the same point would apply if the explananda were generalizations ("all roses are red" and "all fire hydrants are red") rather than two token events.<sup>3</sup>

2. The extant literature on the idea of explanation as unification is vast. In addition to the references above, see, e.g., Salmon (1984, 1989), Barnes (1992b, 1994), Gemes (1994), Schurz and Lambert (1994), Jones (1995a, 1995b, 1997, 1998, 2008), Bartelborth (1996, 2002), Halonen and Hintikka (1999), Schurz (1999, 2014), Strevens (2004), de Regt and Dieks (2005), Jenkins and Nolan (2008), Douglas (2009), Weber and van Bouwel (2009), Gijsbers (2013), Petkov (2015), and Fry (2016). Very little of this considerable literature involves more than a cursory treatment of Friedman's theory. It is standard practice to ignore Friedman's theory entirely or else refer to Friedman's theory, note that Kitcher (1976) shows that Friedman's theory is problematic, and then turn to Kitcher's theory. Salmon (1989) and Gemes (1994) are, as far as we know, the lone exceptions. But even here, the focus is not on Friedman's use of the concept of independent acceptability. This is all somewhat ironic, since Friedman's theory is superior to Kitcher's theory in at least one important respect (as we explain in sec. 2) and since Friedman's theory can be modified to evade Kitcher's objection (as we argue in sec. 3).

3. It might be objected that even if red roses get one explanation and red fire hydrants another, the fact remains that the roses are unified with each other and so are the fire

The second skeptical idea we want to float is that equating explanation with unification derives its appeal, at least in part, from conflating an epistemological question with the question of what an explanation is. We agree that hypotheses that unify two or more bodies of phenomena are often epistemically “better” than hypotheses that disunify them. Sometimes unifying explanations have higher likelihoods, and sometimes they can be expected to be more predictively accurate (Sober 2015). However, these epistemic points do not address the problem that Hempel (1965) crystallized and that subsequent philosophers, including Friedman and Kitcher, have addressed. Hempel’s problem was this: given a set of propositions (or sentences), which of them explains which others? Unification can be a guide to belief without being the right solution to Hempel’s problem (Sober 2003).

Having noted these two skeptical thoughts, we freely admit that unification might be the key to understanding how laws explain other laws, even if it is misguided when the task is to understand what it is to explain events and accidental generalizations. Since Friedman’s theory concerns the explanation of laws, the possibility remains that his approach is correct.

**2. Friedman’s Theory of Explanation.** Friedman motivates the idea that explanation and unification are tightly linked by describing an example:

The kinetic theory effects a significant *unification* in what we have to accept. Where we once had three independent brute facts—that gases approximately obey the Boyle–Charles law, that they obey Graham’s law, and that they have the specific-heat capacities they do have—we now have only one—that molecules obey the laws of mechanics. Furthermore, the kinetic theory also allows us to integrate the behavior of gases with other phenomena, such as the motions of the planets and of falling bodies near the earth. This is because the laws of mechanics also permit us to derive both the fact that planets obey Kepler’s laws and the fact that falling bodies obey Galileo’s laws. From the fact that *all* bodies obey the laws of mechanics it follows that the planets behave as they do, falling bodies behave as they do, and gases behave as they do. Once again, we have reduced a multiplicity of unexplained, independent phenomena to one. I claim that this is the crucial property of scientific theories we are looking for: 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. (1974, 14–15)

---

hydrants. We agree that this may be so. However, it need not be: just as roses differ from fire hydrants, so too may roses differ from each other, and fire hydrants may do the same.

Friedman develops this idea by stating some assumptions:

First of all, I will suppose that we can represent what I have been calling *phenomena*—i.e., general uniformities or patterns of behavior—by law-like *sentences*; and that instead of speaking of the total number of independent phenomena we can speak of the total number of (logically) independent law-like sentences. Secondly, since what is reduced is the total number of phenomena we have to accept, I will suppose that at any given time there is a set  $K$  of *accepted* law-like sentences, a set of laws accepted by the scientific community. Furthermore, I will suppose that the set  $K$  is deductively closed in the following sense: if  $S$  is a law-like sentence, and  $K \vdash S$ , then  $S$  is a member of  $K$ ; i.e.,  $K$  contains all law-like consequences of members of  $K$ . (1974, 15–16)

The expression “independent” in these two passages is crucial.  $K$  is deductively closed, so it is infinite in size, and thus no law-like sentence can reduce the number of sentences in  $K$ . However, it does not follow that no law-like sentence can reduce the number of independent sentences in  $K$ . Friedman’s view, we take it, is that the number of independent sentences in  $K$  is finite; this makes it possible for a law-like sentence to reduce the number of independent sentences in  $K$ .

When does a given law-like sentence reduce the number of independent law-like sentences in  $K$ ? Friedman considers and rejects the idea that this reduction can be achieved just by replacing two or more independent sentences in  $K$  with a single sentence  $S$ . Suppose that  $K$ ’s membership is restricted to the Boyle-Charles law ( $B$ ), Galileo’s law ( $G$ ), and their consequences.<sup>4</sup> Let  $M$  be the laws of mechanics. It might seem that  $B$  and  $G$  can be replaced by  $M$  and that this replacement reduces the number of independent sentences in  $K$  from two to one.<sup>5</sup> Consider, though, the conjunction  $B \& G$ :  $B$  and  $G$  can be replaced by  $B \& G$ , but, clearly,  $B \& G$  does not reduce the number of independent sentences in  $K$ . It might be countered that the conjunction  $B \& G$  is really two independent sentences, because it is logically equivalent to the two-membered set  $\{B, G\}$ . This, however, will not do, for, as Friedman notes, any sentence is logically equivalent to a set of  $n$  sentences for any finite and positive number  $n$ . Law  $M$ , for instance, is logically equivalent to the one-membered set  $\{M\}$ , the two-membered set  $\{B,$

4. Here and throughout, by “consequences” we mean law-like consequences.

5. However, there is a problem. If bridge laws are needed in order to derive  $B$  and  $G$  from  $M$ , then it is not true that  $B$  and  $G$  can be replaced by  $M$ . And if a different bridge law is needed for each derivation, and if those bridge laws are independent of each other and of  $M$ , then replacing  $B$  and  $G$  by  $M$  and two bridge laws would increase, not decrease, the number of independent sentences in  $K$ .

$B \supset M$ }, the three-membered set  $\{B, B \supset M, G\}$ , the four-membered set  $\{B, B \supset M, G, G \supset M\}$ , and so on.

Friedman then points out that the situation is different when the issue concerns independently acceptable sentences as opposed to just (logically) independent sentences: “although every sentence is equivalent to a set of  $n$  independent sentences, it is not the case that every sentence is equivalent to a set of  $n$  *independently acceptable* sentences—e.g., the members of the set  $\{P, P \supset S\}$  may not be acceptable independently of  $S$ ; for our only grounds for accepting  $P \supset S$ , say, might be that it is a consequence of  $S$ ” (1974, 16). But how is the notion of independent acceptability to be understood? Friedman begins with a modest remark: “I don’t have anything very illuminating to say about what it is for one sentence to be acceptable independently of another. Presumably, it means something like: there are sufficient grounds for accepting one which are not also sufficient grounds for accepting the other” (16). He then advances the following two theses:

**Friedman on Independent Acceptability 1 (FIA1)** If  $S$  entails  $Q$ , then  $S$  is not acceptable independently of  $Q$ .

**Friedman on Independent Acceptability 2 (FIA2)** If (a)  $S$  is acceptable independently of  $P$  and (b)  $Q$  entails  $P$ , then  $S$  is acceptable independently of  $Q$ .

That is all that Friedman has to say on the subject.

FIA1 and FIA2, we take it, can be reformulated as follows:

**FIA1.** If  $S$  entails  $Q$ , then evidence  $E$  is sufficient for accepting  $S$  only if  $E$  is also sufficient for accepting  $Q$ .

**FIA2.** If evidence  $E$  is sufficient for accepting  $S$  but not for accepting  $P$ , and  $Q$  entails  $P$ , then  $E$  is sufficient for accepting  $S$  but not for accepting  $Q$ .

Here we assume that by “grounds” Friedman means *evidence*.

When is evidence sufficient for acceptance? A natural answer is this:

**High Probability (HP)** Evidence  $E$  is sufficient for accepting  $S$  if and only if  $p(S|E) > c$ .

Three clarifications are in order here. First, the right-hand side of HP should be understood as short for “either  $E$  is the total evidence available to an agent (or a scientific community) at a time and  $p(S|E) > c$  or  $E$  is a subset of the total evidence  $E^*$  available to an agent (or a scientific community) at a time and both  $p(S|E) > c$  and  $p(S|E^*) > c$ .” Second,  $c$  is a “caution

threshold"; it says how probable  $S$  must be, given the evidence, for accepting  $S$ . Third, we assume that  $0.5 \leq c < 1$ , but we do not assume any particular value for  $c$ .<sup>6</sup> If, as HP says, acceptability is a matter of high probability, then FIA1 and FIA2 are provably correct.<sup>7</sup> We assume in what follows that FIA1 and FIA2 are to be understood in terms of HP (although in sec. 6 we consider a strengthened definition of what acceptability is). Friedman never says otherwise, and his principles FIA1 and FIA2 are true when thus understood.<sup>8</sup>

In addition to using the concept of one sentence's being acceptable independently of another, Friedman also uses the concept of *reduction*. Although he says only a little about independent acceptability, Friedman provides more details about reduction. He writes:

Let a *partition* of a sentence  $S$  be a set of sentences  $\Gamma$  such that  $\Gamma$  is logically equivalent to  $S$  and each  $S'$  in  $\Gamma$  is acceptable independently of  $S$ . ... I will say that a sentence  $S$  is *K-atomic* if it has no partition; i.e., if there is no pair  $\{S_1, S_2\}$  such that  $S_1$  and  $S_2$  are acceptable independently of  $S$  and  $S_1 \& S_2$  is logically equivalent to  $S$ . ... Let a *K-partition* of a set of sentences  $\Delta$  be a set  $\Gamma$  of *K-atomic* sentences which is logically equivalent to  $\Delta$  (I assume that such a *K-partition* exists for every set  $\Delta$ ). Let the *K-cardinality* of a set of sentences  $\Delta$ , *K-card* ( $\Delta$ ), be  $\inf \{\text{card}(\Gamma) : \Gamma \text{ is a } K\text{-partition of } \Delta\}$ . ... Finally, I will say that  $S$  *reduces* the set  $\Delta$  iff  $K\text{-card}(\Delta \cup \{S\}) < K\text{-card}(\Delta)$ . (Friedman 1974, 17)

So  $S$  reduces  $\Delta$  if and only if the smallest *K-partition* of  $\Delta \cup \{S\}$  (i.e., the smallest set of *K-atomic* sentences that is logically equivalent to  $\Delta \cup \{S\}$ ) is smaller than the smallest *K-partition* of  $\Delta$  (i.e., the smallest set of *K-atomic* sentences that is logically equivalent to  $\Delta$ ).

6. We assume, for simplicity, that the value of  $c$  is constant from context to context. But nothing essential for our purposes hinges on this assumption.

7. With respect to FIA1, suppose that (a)  $S$  entails  $Q$  and (b) evidence  $E$  is such that  $p(S|E) > c$ . Given a, it follows that  $p(Q|E) \geq p(S|E)$ . Given b, it then follows that  $p(Q|E) > c$ . Thus, if acceptability is a matter of high probability, FIA1 is correct. As for FIA2, suppose that (a) evidence  $E$  is such that  $p(S|E) > c$  but  $p(P|E) \leq c$  and (b)  $Q$  entails  $P$ . Given b, it follows that  $p(P|E) \geq p(Q|E)$ . But then since, given a,  $p(P|E) \leq c$ , it follows that  $p(Q|E) \leq c$ . Thus, if acceptability is a matter of high probability, FIA2 is correct.

8. This reading also sits well with the fact that Friedman (1983, 243–44) clearly assumes HP (see also Friedman 1981, 8). Although the HP conception of acceptability is subject to the lottery paradox (Kyburg 1961, 197), wherein an agent can be entitled to believe  $A$  and to believe  $B$  but not to believe  $A \& B$ , since  $p(A|E)$  and  $p(B|E)$  are high enough but  $p(A \& B|E)$  is not, we do not think that this is a problem for Friedman's theory. Friedman can simply assume that every statement in  $K$  has a probability, given the total evidence, that is greater than  $c$ .

We now are in a position to state Friedman's theory of explanation. Let "con<sub>K</sub>(*X*)" be the set of independently acceptable consequences of *X*, where each member of con<sub>K</sub>(*X*) is acceptable independently of each of the others and of *X* as well. Here's the theory:

**Friedman on Explanation (FE)** *X* explains *Y* if and only if (i) *X* entails *Y*, (ii) *Y* is acceptable independently of *X*, and (iii) *X* reduces con<sub>K</sub>(*X*).<sup>9</sup>

We note that *X* is not a member of con<sub>K</sub>(*X*). This follows from the fact that independent acceptability is irreflexive. Take some sentence *X*. Since *X* entails itself, it follows, by FIA1, that *X* is not acceptable independently of itself.

To help make Friedman's theory comprehensible, we turn to a toy example. Initially, at time *t*<sub>1</sub>, *K*'s membership is restricted to *L*<sub>1</sub>, *L*<sub>2</sub>, and their consequences, where *L*<sub>1</sub> and *L*<sub>2</sub> are acceptable independently of each other. Suppose that the total evidence available at *t*<sub>1</sub> is restricted to *E*<sub>1</sub> and *E*<sub>2</sub>, where *E*<sub>1</sub> makes *L*<sub>1</sub> acceptable but does not make *L*<sub>2</sub> acceptable, and *E*<sub>2</sub> makes *L*<sub>2</sub> acceptable but not *L*<sub>1</sub>. In other words,  $p(L_1|E_1) > c$  but  $p(L_2|E_1) \leq c$ , and  $p(L_2|E_2) > c$  but  $p(L_1|E_2) \leq c$ .<sup>10</sup> Law *L*<sub>3</sub> entails each of *L*<sub>1</sub> and *L*<sub>2</sub>, but  $p(L_3|E_1) \leq c$ ,  $p(L_3|E_2) \leq c$ , and  $p(L_3|E_1 \& E_2) \leq c$ . Then later, at time *t*<sub>2</sub>, evidence *E*<sub>3</sub> is obtained and  $p(L_3|E_3) > c$ , so, at *t*<sub>2</sub>, *K*'s membership is restricted to *L*<sub>1</sub>, *L*<sub>2</sub>, *L*<sub>3</sub>, and their consequences. Does *L*<sub>3</sub>, at *t*<sub>2</sub>, explain *L*<sub>1</sub>?

Law *L*<sub>3</sub> entails *L*<sub>1</sub>, and since  $p(L_1|E_1) > c$  but  $p(L_3|E_1) \leq c$ , *L*<sub>1</sub> is acceptable independently of *L*<sub>3</sub>.<sup>11</sup> Thus, conditions i and ii in FE are satisfied. Now consider condition iii. This condition is satisfied if and only if the smallest *K*-partition of con<sub>K</sub>(*L*<sub>3</sub>) ∪ {*L*<sub>3</sub>} (i.e., the smallest set of *K*-atomic sentences that is logically equivalent to con<sub>K</sub>(*L*<sub>3</sub>) ∪ {*L*<sub>3</sub>}) is smaller than the smallest *K*-partition of con<sub>K</sub>(*L*<sub>3</sub>) (i.e., the smallest set of *K*-atomic sentences that is logically equivalent to con<sub>K</sub>(*L*<sub>3</sub>)). Suppose that each of *L*<sub>1</sub>, *L*<sub>2</sub>, and *L*<sub>3</sub> is *K*-atomic. Then, since *L*<sub>3</sub> is a member of con<sub>K</sub>(*L*<sub>3</sub>) ∪ {*L*<sub>3</sub>} but is not a member of con<sub>K</sub>(*L*<sub>3</sub>), the smallest *K*-partition of con<sub>K</sub>(*L*<sub>3</sub>) ∪ {*L*<sub>3</sub>} is {*L*<sub>3</sub>}, whereas

9. Friedman (1974, 17–18) modifies FE so that it implies that if *X* explains *Y*, then *X* & *Z* explains *Y* for any sentence *Z* that is acceptable independently of *X* and of *Y*. We find this principle that Friedman seeks to accommodate implausible, since *Z* could be entirely irrelevant to *X* and to *Y*. Because of this, and because all of our main critical points about FE unmodified carry over to FE when modified, we will focus on FE unmodified. We have more to say about explanatory relevance in the next section.

10. In addition, we will assume that  $p(L_1|E_1 \& E_2) = p(L_1|E_1)$  and  $p(L_2|E_1 \& E_2) = p(L_2|E_2)$ .

11. This illustrates the fact that independent acceptability is not symmetric. Law *L*<sub>1</sub> is acceptable independently of *L*<sub>3</sub>. But, as *L*<sub>3</sub> entails *L*<sub>1</sub>, it follows by FIA1 that *L*<sub>3</sub> is not acceptable independently of *L*<sub>1</sub>. It does not follow, of course, and it is not true, that independent acceptability is antisymmetric.

the smallest  $K$ -partition of  $\text{con}_K(L_3)$  is  $\{L_1, L_2\}$ . Given this, and given that  $\{L_3\}$  is smaller than  $\{L_1, L_2\}$ , it follows that  $L_3$  reduces  $\text{con}_K(L_3)$ . So, by FE,  $L_3$  explains  $L_1$ . This is just as it should be, according to Friedman.

Now consider a different question: Does  $L_1 \& L_2$  explain  $L_1$ ? The conjunction  $L_1 \& L_2$  entails  $L_1$ , and since  $L_1$  is acceptable independently of  $L_2$ , and since  $L_1 \& L_2$  entails  $L_2$ , it follows, by FIA2, that  $L_1$  is acceptable independently of  $L_1 \& L_2$ . So conditions i and ii in FE are satisfied. However, condition iii is not satisfied. It might seem otherwise, since it might seem that the smallest  $K$ -partition of  $\text{con}_K(L_1 \& L_2) \cup \{L_1 \& L_2\}$  is  $\{L_1 \& L_2\}$ , whereas the smallest  $K$ -partition of  $\text{con}_K(L_1 \& L_2)$  is  $\{L_1, L_2\}$ . However,  $L_1 \& L_2$  is not  $K$ -atomic, since  $L_1 \& L_2$  is logically equivalent to  $\{L_1, L_2\}$  each member of which is acceptable independently of  $L_1 \& L_2$ . Thus,  $\{L_1 \& L_2\}$  is not a  $K$ -partition of  $\text{con}_K(L_1 \& L_2) \cup \{L_1 \& L_2\}$ . The smallest  $K$ -partition of  $\text{con}_K(L_1 \& L_2) \cup \{L_1 \& L_2\}$  is simply  $\{L_1, L_2\}$ . Thus, the smallest  $K$ -partition of  $\text{con}_K(L_1 \& L_2) \cup \{L_1 \& L_2\}$  and the smallest  $K$ -partition of  $\text{con}_K(L_1 \& L_2)$  are one and the same set. So  $L_1 \& L_2$  does not reduce  $\text{con}_K(L_1 \& L_2)$ , and thus, by FE,  $L_1 \& L_2$  does not explain  $L_1$ . This too is just as it should be, as Friedman (1974, 18) notes: conjunctions never explain their conjuncts.

Our toy example is just an example, but our main points about it generalize. First, if  $X$  is  $K$ -atomic and has at least two independently acceptable consequences, then the smallest  $K$ -partition of  $\text{con}_K(X) \cup \{X\}$  is  $\{X\}$ , whereas the smallest  $K$ -partition of  $\text{con}_K(X)$  has at least two members, and so  $X$  reduces  $\text{con}_K(X)$ . Second, if  $X$  is an  $n$ -fold conjunction and each conjunct is acceptable independently of the others, then the smallest  $K$ -partition of  $\text{con}_K(X) \cup \{X\}$  and the smallest  $K$ -partition of  $\text{con}_K(X)$  are one and the same set—the set of  $X$ 's conjuncts—and thus  $X$  does not reduce  $\text{con}_K(X)$ .

The first of these points can be strengthened:  $X$  is  $K$ -atomic and has at least two independently acceptable consequences if and only if the smallest  $K$ -partition of  $\text{con}_K(X) \cup \{X\}$  is  $\{X\}$ , whereas the smallest  $K$ -partition of  $\text{con}_K(X)$  has at least two members, and so  $X$  reduces  $\text{con}_K(X)$ .<sup>12</sup> This means that the only way for  $X$  to reduce  $\text{con}_K(X)$  is for  $X$  to be  $K$ -atomic and have

12. The left-to-right part of this biconditional is simply the point to be strengthened. The right-to-left part can be established via its contrapositive. Suppose, first, that  $X$  is not  $K$ -atomic. Then there is a set  $\Gamma = \{A_1, A_2\}$  such that (i)  $\Gamma$  is logically equivalent to  $X$  and (ii) each of  $A_1$  and  $A_2$  is acceptable independently of  $X$ . Since  $\Gamma$  is logically equivalent to  $X$ , it follows that  $X$  entails each of  $A_1$  and  $A_2$ . But then, given that each of  $A_1$  and  $A_2$  is acceptable independently of  $X$ , it follows that each of  $A_1$  and  $A_2$  is a member of  $\text{con}_K(X)$ . Given this, and given that  $\Gamma$  is logically equivalent to  $X$ , it follows that  $\text{con}_K(X)$  entails  $X$ . It is also the case, of course, that  $X$  entails  $\text{con}_K(X)$ . Thus,  $X$  is logically equivalent to  $\text{con}_K(X)$ . But then any  $K$ -partition of  $\text{con}_K(X)$  is also a  $K$ -partition of  $\text{con}_K(X) \cup \{X\}$  and vice versa. Hence, the smallest  $K$ -partition of  $\text{con}_K(X) \cup \{X\}$  and the smallest  $K$ -

at least two independently acceptable consequences. It follows that FE can be reformulated as:

**Friedman on Explanation\* (FE\*)**  $X$  explains  $Y$  if and only if (i)  $X$  entails  $Y$ , (ii)  $Y$  is acceptable independently of  $X$ , (iii)  $X$  is  $K$ -atomic, and (iv)  $X$  has two or more consequences that are acceptable independently of  $X$  and of each other.

We hope that FE\* makes it clear, in our toy example, why Friedman's theory delivers the verdict that  $L_3$  explains  $L_1$  but  $L_1 \& L_2$  does not. The problem with  $L_1 \& L_2$  is that it is not  $K$ -atomic.

We mentioned in section 1 that Friedman's theory is superior to Kitcher's in at least one important respect. The point we had in mind involves a good objection that Woodward (2003, 2014) makes to Kitcher's theory: the theory entails that explanation is "winner-takes-all." That is, according to Kitcher's theory, if  $X$  explains  $Y$ , then there is no  $X^*$  distinct from  $X$  such that  $X^*$  also explains  $Y$ . Friedman's theory is not subject to this objection. This can be seen by adding to our toy example. Suppose that  $L_6$ , which is  $K$ -atomic, comes to be accepted, where  $L_6$  entails each of  $L_3$ ,  $L_4$ , and  $L_5$  and where  $L_3$ ,  $L_4$ , and  $L_5$  are each acceptable independently of  $L_6$ . It is still the case that  $L_3$  entails  $L_1$ ,  $L_1$  is acceptable independently of  $L_3$ ,  $L_3$  is  $K$ -atomic, and  $L_3$  has at least two independently acceptable consequences. Hence, by Friedman's theory,  $L_3$  still explains  $L_1$  after  $L_6$  makes its appearance. It is also true, according to Friedman's theory, that  $L_6$  explains  $L_1$ . This is because  $L_6$  entails  $L_1$  (given that entailment is transitive),  $L_1$  is acceptable independently of  $L_6$  (given FIA2),  $L_6$  is  $K$ -atomic (by hypothesis), and  $L_6$  has at least two independently acceptable consequences (given that  $L_1$ ,  $L_2$ ,  $L_3$ ,  $L_4$ , and  $L_5$  are each acceptable independently of  $L_6$ ). So, contra the idea that explanation is winner-takes-all,  $L_1$  is explained both by  $L_3$  and by  $L_6$ .

We turn now to Kitcher's critique of FE. The fact that Friedman's theory is superior to Kitcher's theory in at least one important respect makes it all the more important to ask whether Kitcher's critique of FE is merely technical or, instead, is a stake through the heart of Friedman's whole approach.

---

partition of  $\text{con}_K(X)$  are one and the same set. Hence,  $X$  does not reduce  $\text{con}_K(X)$ . Kitcher (1976, 209) gives an argument along these lines. Suppose, second, that  $X$  is  $K$ -atomic but does not have at least two independently acceptable consequences. Then either (a)  $\text{con}_K(X)$  is empty or (b)  $\text{con}_K(X)$  has just one member. If a, then, trivially, it is not the case that the smallest  $K$ -partition of  $\text{con}_K(X)$  has at least two members. If b, then there is a  $K$ -partition of  $\text{con}_K(X)$  consisting of just the sentence in  $\text{con}_K(X)$ , and thus it is not the case that the smallest  $K$ -partition of  $\text{con}_K(X)$  has at least two members. Either way, it is not the case that the smallest  $K$ -partition of  $\text{con}_K(X)$  has at least two members.

**3. Kitcher's Critique of FE.** Kitcher's critique of FE begins with a point that is clearly correct: FE, which is equivalent to  $FE^*$ , implies that only  $K$ -atomic sentences can be explanatory. Kitcher argues that this consequence is a problem for Friedman's theory, since some conjunctions of independently acceptable sentences have explanatory power. He gives as an example the conjunction of the Boyle-Charles law ( $B$ ) and the first law of thermodynamics ( $T$ ). Laws  $B$  and  $T$  are acceptable independently of each other. Given this, and given that  $B\&T$  entails each of  $B$  and  $T$ , it follows by FIA2 that each of  $B$  and  $T$  is acceptable independently of  $B\&T$ . This means that  $\{B, T\}$  is a partition of  $B\&T$ , and thus  $B\&T$  is not  $K$ -atomic. And yet  $B\&T$  has explanatory power; for example,  $B\&T$  explains the law of adiabatic expansion of an ideal gas.

Kitcher's critique of FE is decisive; FE is false. However, it does not follow that Friedman's core idea—that explanation is a matter of unification which in turn is a matter of reducing the number of independently acceptable sentences in  $K$ —is false. FE is just one way of fleshing out Friedman's core idea. Perhaps there is an alternative development of this core idea that allows conjunctions of independently acceptable sentences to have explanatory power. But not just any such alternative formulation will do. What is needed is an alternative formulation according to which conjunctions of independently acceptable sentences can be explanatory, but conjunctions—whether of independently acceptable sentences or not—cannot explain their conjuncts.

Below we give an alternative formulation of exactly that sort. But first we need to state a *stipulative* definition, one in which we define what we mean by “relevant entailment” (r-entailment) in terms of classical entailment (c-entailment):

**Relevant Entailment (RE)**  $X$  r-entails  $Y$  if and only if (i)  $X$  c-entails  $Y$ , and (ii) for each occurrence of a nonlogical term  $t$  in  $X$ , there is at least one term  $t^*$  (where  $t$  and  $t^*$  are of the same grammatical type) such that  $X^*$  (the result of substituting  $t^*$  for that occurrence of  $t$  in  $X$ ) fails to c-entail  $Y$ .

For example, let  $X = [(F \supset G)\&F]$  and  $Y = G$ . Condition i holds since  $X$  c-entails  $Y$ . If the first occurrence of “ $F$ ” in  $X$  is replaced by “ $H$ ,” then the result is  $X^* = [(H \supset G)\&F]$ , which does not c-entail  $Y$ . The same point holds for the occurrence of “ $G$ ” in  $X$  and also for the second occurrence of “ $F$ ” in  $X$ . So condition ii holds. Thus,  $X$  r-entails  $Y$ . For a second example, let  $X = [F\&G]$  and  $Y = F$ . Here too condition i holds. But the occurrence of “ $G$ ” in  $X$  is irrelevant in that, for any  $H$ ,  $X^* = [F\&H]$  c-entails  $Y$ . So condition ii does not hold, and thus  $X$  fails to r-entail  $Y$ . With r-entailments,

as opposed to mere c-entailments, there are no logically irrelevant, or idle, occurrences of nonlogical terms.<sup>13</sup>

With “r-entailment” thus defined, we can reformulate Friedman’s theory of explanation (FE):

**Friedman on Explanation\*\* (FE\*\*)**  $X$  explains  $Y$  if and only if (i)  $X$  r-entails  $Y$ , (ii)  $Y$  is acceptable independently of  $X$ , and (iii)  $X$  reduces  $\text{con}^*_K(X)$ .

Here  $\text{con}^*_K(X)$  is the set of independently acceptable sentences r-entailed by  $X$ .

No conjunction r-entails its conjuncts. So FE\*\* resembles FE in that both imply that no conjunction explains its conjuncts. But FE\*\*, unlike FE, implies that conjunctions of independently acceptable sentences can be explanatory. Suppose, modifying our toy example a bit, that  $L_4$ ,  $L_5$ , and  $L_6$  are each independently acceptable sentences r-entailed by  $L_1 \& L_2$ . Then the smallest  $K$ -partition of  $\text{con}^*_K(L_1 \& L_2) \cup \{L_1, L_2\}$  is  $\{L_1, L_2\}$ , whereas the smallest  $K$ -partition of  $\text{con}^*_K(L_1 \& L_2)$  includes at least  $L_4$ ,  $L_5$ , and  $L_6$ . It follows that  $L_1 \& L_2$  reduces  $\text{con}^*_K(L_1 \& L_2)$  and thus explains each of  $L_4$ ,  $L_5$ , and  $L_6$ . The same seems to be true of Kitcher’s example concerning the Boyle-Charles law ( $B$ ), the first law of thermodynamics ( $T$ ), and the law of adiabatic expansion of an ideal gas ( $A$ ). Conjunction  $B \& T$  r-entails  $A$ , which is acceptable independently of  $B \& T$ , and, assuming that  $\text{con}^*_K(B \& T)$  has at least two members in addition to  $A$ ,  $B \& T$  reduces  $\text{con}^*_K(B \& T)$ . Hence,  $B \& T$  explains  $A$  along with each of the other members of  $\text{con}^*_K(B \& T)$ .

This shows that there is a way to flesh out Friedman’s core idea so that a conjunction of independently acceptable sentences can explain and so that, at the same time, no conjunction can explain its conjuncts. The trick is to use r-entailment instead of c-entailment.

It should be noted that there is independent reason to use r-entailment when developing a deductivist theory of explanation. If  $X$  c-entails  $Y$  but does not r-entail  $Y$ , then  $X$  involves a kind of irrelevancy. That  $X$  can classically entail  $Y$  without explaining  $Y$  (owing to the fact that  $X$  includes explanatorily irrelevant information) is a familiar point in the literature on explana-

13. RE places no restrictions on  $Y$  when the question is whether  $X$  r-entails  $Y$ . However, if desired, RE could be modified so that there is a restriction on  $Y$  similar to the one on  $X$ . See Schurz and Lambert (1994, 89) on “relevant conclusions.” There are some similarities between r-entailment and entailment as understood in “relevance logic”; see Mares (2014) for an introduction to the latter. However, there are also some crucial differences. One concerns conjunctions and their conjuncts. Relevance logic (or at least standard varieties thereof) affirms that a conjunction entails its conjuncts, but our concept of r-entailment denies this.

tion. Think, for example, of Kyburg's (1965) discussion of a hexed piece of salt that dissolves in water, Salmon's (1971, 34) discussion of a man who takes birth control pills and never becomes pregnant, and Putnam's (1975, 295–97) discussion of a peg that fits through one hole in a board but not through another.<sup>14</sup>

Of course, it is possible that FE\*\* has a technical flaw of some sort. However, if that turns out to be so, we see no reason to assume that there can be no technical fix in the offing. Regardless, we now want to return to FE and focus on cases in which  $X$  is not a conjunction. If FE is problematic in the way it applies to those cases, this not only tells against the specifics of FE but also, and more importantly, against the underlying idea that explanation is a matter of unification which in turn is a matter of reducing the number of independently acceptable sentences in  $K$ .<sup>15</sup>

**4. A New Objection to Friedman's Theory.** Consider two situations in which you might find yourself. In both,  $K$ 's membership is restricted to  $L_1, L_2, L_3$ , and their consequences, where  $L_1, L_2$ , and  $L_3$  are each  $K$ -atomic. Law  $L_3$  entails each of  $L_1$  and  $L_2$  (but not vice versa). The situations are different in the following respect:

**Situation 1.** Laws  $L_1$  and  $L_2$  are acceptable independently of each other and independently of  $L_3$ , in that there is evidence  $E_1$  that makes  $L_1$  acceptable but does not have this impact on  $L_2$  or on  $L_3$ , and there is evidence  $E_2$  that makes  $L_2$  acceptable but does not have that impact on  $L_1$  or on  $L_3$ . Law  $L_3$  is also a member of  $K$ , and so there is evidence  $E_3$  that renders it acceptable.

**Situation 2.** The only evidence on hand is  $E_3$ , which renders  $L_3$  acceptable and therefore has the same impact on  $L_1$  and on  $L_2$ .

14. However, FE\*\* goes contrary to an idea that Friedman considered, which we discussed in n. 9. We are not the first to have proposed a modification of FE. Gemes (1994) distinguishes between a sentence's consequences and a sentence's content parts, where all content parts are consequences but not vice versa, and argues that it would better serve Friedman's purposes to define  $K$ -atomicity in terms of a sentence's content parts. This may be true, but, by itself, Gemes's proposed modification would not blunt Kitcher's critique. For, even on Gemes's suggested alternative definition of  $K$ -atomicity, no conjunction is  $K$ -atomic. The key feature of our proposed modification of FE is that it allows for  $X$  to explain  $Y$  even if  $X$  is not  $K$ -atomic.

15. We want to return to FE, and set aside FE\*\*, because FE is less complicated and because our discussion below in secs. 4 and 5 of how FE applies to cases in which  $X$  (a putative explanans) is  $K$ -atomic carries over to alternative ways of fleshing out Friedman's core idea. The point of constructing FE\*\* was to show that Kitcher's critique of FE, although decisive, does not get to the heart of the matter.

FE implies that whether  $L_3$  explains  $L_1$  depends on whether you are in situation 1 or situation 2. In situation 1,  $L_1$  is acceptable independently of  $L_3$ , and  $L_3$  has two independently acceptable consequences, so  $L_3$  explains  $L_1$ . In situation 2,  $L_1$  is not acceptable independently of  $L_3$ , and  $L_3$  has no independently acceptable consequences, so  $L_3$  does not explain  $L_1$ . According to FE, whether  $X$  explains  $Y$  depends on the evidence you have. Of course, whether you are justified in believing that  $L_3$  explains  $L_1$  depends on your evidence. However, our contention is that whether  $L_3$  in fact explains  $L_1$  should not depend on the evidence you happen to have.<sup>16</sup> There is a related point. Suppose you are in situation 1, whereas your twin is in situation 2. FE says that  $L_3$  explains  $L_1$  for you but not for your twin. This too is problematic for FE. Whether  $L_3$  explains  $L_1$  should not be agent relative in this sense.

Here we concur with a good point that Friedman himself makes—that an adequate theory of explanation should have the consequence that whether  $X$  explains  $Y$  should be an objective matter of fact: “It should be objective—what counts as an explanation should not depend on the idiosyncracies and changing tastes of scientists and historical periods. It should not depend on such nonrational factors as which phenomena one happens to find somehow more natural, intelligible, or self-explanatory than others. . . . If there is some objective and rational sense in which scientific theories explain, a philosophical theory of explanation should tell us what it is” (1974, 14). Friedman is right, but we think he should have gone farther and insisted that whether  $X$  explains  $Y$  depends not at all on what sort of evidence an agent happens to have for  $X$  and  $Y$ .<sup>17</sup>

16. Our point is related to Salmon’s (1989) worry that there are no  $K$ -atomic sentences. He develops his idea by focusing on sentences of the form  $(x)(Fx \supset Gx)$ . The worry is that  $\Gamma = \{(x)([Fx \& Hx] \supset Gx), (x)([Fx \& \sim Hx] \supset Gx)\}$  is a partition of  $(x)(Fx \supset Gx)$ , and thus  $(x)(Fx \supset Gx)$  is not  $K$ -atomic. Salmon reports that Friedman responded to this worry (in personal correspondence) as follows: “The problem you raise . . . is just the kind of problem I had in mind in requiring that a partition of a sentence  $S$  consist of sentences that are acceptable independently of  $S$ . In your case, it would seem in general that our ground for accepting the two conjuncts is just that they follow from ‘All  $F$  are  $G$ .’ If, on the other hand, we have grounds for accepting the two conjuncts independently—by testing for the two conjuncts directly, say—then it would seem that ‘All  $F$  are  $G$ ’ is in no way an explanation of the two, but just a summary of what we already know” (97). Salmon argues in response that Friedman’s proposed solution makes  $K$ -atomicity unacceptably subjective. We agree. But there is more. Friedman’s proposed solution, it seems, is that if the members of  $\Gamma$  are not acceptable independently of  $(x)(Fx \supset Gx)$ , then  $(x)(Fx \supset Gx)$  explains the members of  $\Gamma$ , whereas if the members of  $\Gamma$  are acceptable independently of  $(x)(Fx \supset Gx)$ , then  $(x)(Fx \supset Gx)$  does not explain the members of  $\Gamma$ . This is inconsistent with Friedman’s own theory, for if the members of  $\Gamma$  are not acceptable independently of  $(x)(Fx \supset Gx)$ , then by FE it follows that  $(x)(Fx \supset Gx)$  does not explain the members of  $\Gamma$ .

17. There is a different argument for thinking that Friedman’s use of the concept of acceptability introduces an objectionably subjective element into the theory of explana-

We have faulted Friedman's theory for making the question of whether law  $X$  explains laws  $Y$  and  $Z$  depend on the kind of evidence an agent has for  $X$ ,  $Y$ , and  $Z$ . The sort of agent independence we endorse is a consequence of most theories of explanation, starting with Hempel's, but is there an argument for thinking that this usual assumption is right? In the case of causal explanations of token events, the case for agent independence seems pretty clear. Whether one token event caused another depends not at all on whether anyone has evidence that these events occurred or were causally related. Causal facts existed aplenty long before scientists discovered them, and there are causal facts that scientists have yet to discover and some that scientists probably never will. And yet, in all these cases, it is true that one token event explains another, whether or not anyone knows that the events are so related. We feel that this fact about causal explanation provides a *prima facie* argument for the thesis that whether one law explains two or more other laws does not depend on anyone's having evidence that the laws are true.

This argument can be supplemented with another. Consider any case in which it is uncontroversial that a given law explains one or more other laws. When you look at the explanation, you will usually see that it takes the form of a derivation. The derivation need not mention the evidence that supports the explaining law and the evidence that supports the laws that are being explained. In the derivation you see that the one law entails the others and that the character of the derivation is what allows the one law to explain the others. This does not mean that every derivation is an explanation. However, some are, and the fact that explanatory derivations do not need to cite evidence for the laws that figure in premises and conclusions provides a strong reason for thinking that the explanation relation does not depend on anyone's having evidence for the laws' being true.

It may be objected that this criticism of Friedman's theory misses the mark, since Friedman talks about the law-like sentences that are accepted "by the scientific community," not by an individual scientist. This objection fails, however, since our point can be frameshifted. Whether one proposition explains another should not depend on how a given scientific community obtained its evidence for those propositions. In addition, there need be no uniform pattern by which the members of a scientific community obtain their evidence; when some scientists in the community are in situation 1 while others are in situation 2, it may be an unacceptable simplification to talk about "the" situation in which the community finds itself.

---

tion. For a proposition  $H$  to be acceptable in the light of evidence  $E$ ,  $p(H|E)$  must be "high enough"— $H$ 's probability given  $E$  must exceed the value of some caution threshold  $c$ , as HP says. If the value for  $c$  in a given context is subjective, then so is rational acceptability and thus so is Friedman's theory.

**5. A New Interpretation of Independent Acceptability.** So far we have understood Friedman's concept of independent acceptability as follows:

**Independent Acceptability 1 (IA1)**  $S$  is acceptable independently of  $Q$  for agent  $A$  at time  $t$  if and only if  $A$  has evidence  $E$  at  $t$ ,  $p(S|E) > c$ , and  $p(Q|E) \leq c$ , where  $p$  is  $A$ 's probability function at  $t$ .<sup>18</sup>

In the light of our criticism of IA1 in the previous section, we want to consider an alternative characterization of independent acceptability:

**Independent Acceptability 2 (IA2)**  $S$  is acceptable independently of  $Q$  if and only if, for some  $p$  and some  $E$ ,  $p(S|E) > c$  and  $p(Q|E) \leq c$ .

IA2 is like IA1 in that both are based on the high probability criterion for rational acceptability (HP). The key difference is that the right-hand side of IA1 is stronger than the right-hand side of IA2.<sup>19</sup>

IA2 has the consequence that independent acceptability is not relative to an agent, a time, and the agent's probability function. To see why, let us return to the case of situation 1 and situation 2. In situation 1, there is evidence  $E_1$  that makes  $L_1$  acceptable but does not have this impact on  $L_3$ , and there is evidence  $E_2$  that makes  $L_2$  acceptable but does not have that impact on  $L_3$ . It therefore follows by IA2 that in both situations each of  $L_1$  and  $L_2$  is acceptable independently of  $L_3$ . This means that if Friedman's theory were understood in terms of IA2, the theory would imply that in both situations  $L_3$  explains  $L_1$  and thus would evade the problem of subjectivity that plagues Friedman's theory when understood in terms of IA1. But is this new formulation of Friedman's theory otherwise adequate?

If FIA1 and FIA2—Friedman's two theses about independent acceptability—are understood in terms of IA2, they say this:

**FIA1<sub>IA2</sub>** If  $S$  entails  $Q$ , then, for each  $p$  and each  $E$ , either  $p(S|E) \leq c$  or  $p(Q|E) > c$ .

**FIA2<sub>IA2</sub>** If, for some  $p$  and some  $E$ ,  $p(S|E) > c$  but  $p(P|E) \leq c$ , and  $Q$  entails  $P$ , then, for some  $p$  and some  $E$ ,  $p(S|E) > c$  but  $p(Q|E) \leq c$ .

18. Recall that the right-hand side of HP should be understood as short for "either  $E$  is the total evidence available to an agent (or a scientific community) at a time and  $p(S|E) > c$  or  $E$  is a subset of the total evidence  $E^*$  available to an agent (or a scientific community) at a time and both  $p(S|E) > c$  and  $p(S|E^*) > c$ ." IA1 should be understood similarly.

19. We do not claim that IA2 is a plausible interpretation of Friedman's remarks about independent acceptability. The reason we consider IA2 as an alternative conceptualization of independent acceptability is to see whether IA2 would help justify Friedman's core idea.

FIA1<sub>IA2</sub> and FIA2<sub>IA2</sub> are both true, given IA2 and the high probability criterion of acceptance (HP).<sup>20</sup> This is good news for Friedman's theory. There is, however, some bad news. It is straightforward to show that:

**FIA1\***<sub>IA2</sub>  $S$  is not acceptable independently of  $Q$  in the sense of IA2 if and only if  $S$  entails  $Q$ .

(See app. A for proof.) FIA1\*<sub>IA2</sub> is similar to but stronger than FIA1<sub>IA2</sub>; FIA1\*<sub>IA2</sub> says that a fact about entailment is necessary and sufficient for a failure of independent acceptability, whereas FIA1<sub>IA2</sub> says only that the entailment is sufficient.

FIA1\*<sub>IA2</sub> poses a problem for Friedman's theory. Let  $S = (x)(Fx \supset Gx)$  and  $\Gamma = \{(x)([Fx \& Hx] \supset Gx), (x)([Fx \& \sim Hx] \supset Gx)\}$ . Neither member of  $\Gamma$  entails  $S$ , so by FIA1\*<sub>IA2</sub> it follows that each member of  $\Gamma$  is acceptable independently of  $S$ . But then, since  $S$  is equivalent to  $\Gamma$ ,  $S$  is not  $K$ -atomic, and so Friedman's theory (FE) entails that  $S$  explains nothing. This means that FE, understood in terms of IA2, implies that no universal generalization is  $K$ -atomic, and thus no universal generalization explains anything.

This problem generalizes. Any law-like sentence  $L$  is equivalent to a set  $\Gamma = \{L \vee M_1, L \vee M_2, \dots, L \vee M_n\}$ , where  $\{M_1, M_2, \dots, M_n\}$  is a partition (of law-like sentences), such that no member of  $\Gamma$  entails  $L$ . Thus, by FIA1\*<sub>IA2</sub>, each member of  $\Gamma$  is acceptable independently of  $L$ . It follows that no law-like sentence is  $K$ -atomic, and so FE, understood in terms of IA2, implies that no law-like sentence explains anything.<sup>21</sup>

Recall Friedman's claim that "although every sentence is equivalent to a set of  $n$  independent sentences, it is not the case that every sentence is equivalent to a set of  $n$  *independently acceptable* sentences—e.g., the members of the set  $\{P, P \supset S\}$  may not be acceptable independently of  $S$ ; for our only grounds for accepting  $P \supset S$ , say, might be that it is a consequence of  $S$ " (1974, 16). This claim is right when understood in terms of IA1, but it is wrong when understood in terms of IA2. For, when IA2 is in play,  $S$  is acceptable independently of  $Q$  if and only if  $S$  is independent of  $Q$  in that  $S$  does not entail  $Q$ . This means that FE, understood in terms of IA2, is tan-

20. The arguments given above in n. 7 for FIA1 and FIA2 can be readily adapted to establish FIA1<sub>IA2</sub> and FIA2<sub>IA2</sub>.

21. We noted above that our discussion of FE in secs. 4 and 5 carries over to alternative ways of fleshing out Friedman's core idea. There we had in mind, e.g., FE\*\*. Take any two law-like sentences  $X$  and  $Y$ . Suppose that (i)  $X$  r-entails  $Y$  and (ii)  $Y$  does not entail  $X$  and thus is acceptable independently of  $X$ . Then, by FE\*\*,  $X$  explains  $Y$  only if  $X$  reduces  $\text{con}^*_K(X)$ . It is not the case, though, that  $X$  reduces  $\text{con}^*_K(X)$ . For, if, as is implied by FE\*\* understood in terms of IA2, no law-like sentence is  $K$ -atomic, then there are no  $K$ -partitions, and thus it is false that the smallest  $K$ -partition of  $\text{con}^*_K(X) \cup \{X\}$  is smaller than the smallest  $K$ -partition of  $\text{con}^*_K(X)$ . Thus, by FE\*\*,  $X$  does not explain  $Y$ .

tamount to the proposal that Friedman initially considers and immediately rejects that explanation is a matter of reducing the number of independent sentences in  $K$ .

Recall the problem with Friedman's initial proposal. It might seem that reduction can be understood in terms of replacing two or more independent sentences with a single sentence. This does not work, though, since  $B$  and  $G$  can be replaced by  $B\&G$ , and yet, clearly,  $B\&G$  does not reduce the number of independent sentences in  $K$ . A natural reply is that  $B\&G$  is really two independent sentences, because it is logically equivalent to the two-membered set  $\{B, G\}$ . But, as Friedman notes, any sentence is logically equivalent to an  $n$ -membered set of independent sentences. The problem, then, is that if reduction is understood in terms of replacing two or more independent sentences with a single sentence, and if a sentence is really  $n$  sentences provided that it is logically equivalent to an  $n$ -membered set of independent sentences, then no sentence reduces the number of independent sentences in  $K$ , and thus there is no explanation. The objection that Friedman makes to the first proposal he considers is really the same as our objection to FE when it is understood in terms of IA2: if FE is so understood, then no sentence reduces the set of its independently acceptable consequences, and thus there is no explanation.<sup>22</sup>

Would it help Friedman's theory if HP were replaced by an alternative Bayesian conception of acceptability and thus if IA2 were replaced by an alternative Bayesian conception of independent acceptability? We are skeptical. Consider, for example, the following alternative to HP:

22. An anonymous reviewer noted in effect that although  $S = (x)(Fx \supset Gx)$  entails each member of  $\Gamma = \{(x)([Fx\&Hx] \supset Gx), (x)([Fx\&\sim Hx] \supset Gx)\}$ , certain occurrences of nonlogical terms in  $(x)([Fx\&Hx] \supset Gx)$  and  $(x)([Fx\&\sim Hx] \supset Gx)$  are idle, i.e.,  $Hx$  and  $\sim Hx$ . Let us say that although  $S$  entails each member of  $\Gamma$ , it is not the case that  $S$  r\*-entails each member of  $\Gamma$ . A proponent of FE could modify FE so that a sentence  $S$  is  $K$ -atomic only if there is no set  $\Gamma = \{S_1, S_2\}$  such that (i)  $\Gamma$  is logically equivalent to  $S$ , (ii)  $S$  r\*-entails each member of  $\Gamma$ , and (iii) each of  $S_1$  and  $S_2$  is acceptable independently of  $S$ . If FE were thus modified, then although  $\Gamma$  is logically equivalent to  $S$ , and although each of  $(x)([Fx\&Hx] \supset Gx)$  and  $(x)([Fx\&\sim Hx] \supset Gx)$  is acceptable independently of  $S$ , it would not follow that  $S$  is not  $K$ -atomic. However, in this context, it seems ad hoc to appeal to a kind of nonclassical entailment. After all, the issue in this context is in effect how to count sentences. Also, there are alternative ways of modifying FE in terms of a kind of nonclassical entailment. One is to modify FE so that a sentence  $S$  is  $K$ -atomic only if there is no set  $\Gamma = \{S_1, S_2\}$  such that (i)  $\Gamma$  is logically equivalent to  $S$ , (ii)  $S$  r-entails and r\*-entails each member of  $\Gamma$ , and (iii) each of  $S_1$  and  $S_2$  is acceptable independently of  $S$ . Take the conjunction  $B\&G$ . Set  $\Gamma = \{B, G\}$  is logically equivalent to  $B\&G$ , and each of  $B$  and  $G$  is acceptable independently of  $B\&G$ . But, since it is not the case that  $B\&G$  r-entails each member of  $\Gamma$ , it might be that  $B\&G$  is nonetheless  $K$ -atomic and so counts as a single sentence. But that, we take it, would not be the right result.

**High Probability and Increase in Probability (HPIP)** Evidence  $E$  is sufficient for accepting  $S$  if and only if  $p(S|E) > c$  and  $p(S|E) > p(S)$ .

This leads to the following alternative to IA2:

**Independent Acceptability 2\* (IA2\*)**  $S$  is acceptable independently of  $Q$  if and only if, for some  $p$  and some  $E$ ,  $p(S|E) > c$ ,  $p(S|E) > p(S)$ , and  $p(Q|E) \leq c$  or  $p(Q|E) \leq p(Q)$ .

It is straightforward to show that:

**FIA1\***<sub>IA2\*</sub>  $S$  is not acceptable independently of  $Q$  in the sense of IA2\* if and only if  $S$  and  $Q$  are logically equivalent to each other.

(See app. B for proof.) The points above concerning the case of  $S = (x)(Fx \supset Gx)$  and  $\Gamma = \{(x)([Fx \& Hx] \supset Gx), (x)([Fx \& \sim Hx] \supset Gx)\}$  carry over straightforwardly to FE understood in terms of IA2\*. Furthermore, FIA1 and FIA2 are false when interpreted in terms of IA2\*. (See app. C for proof.)

**6. Concluding Comment.** After showing how Friedman's theory of explanation as unification can be reformulated so as to evade Kitcher's (1976) objection, we explored three Bayesian interpretations of the concept of independent acceptability. Each of these interpretations is problematic for Friedman's theory. When Friedman's theory is coupled with IA1, the theory has the implausible consequence that whether one proposition explains another depends on the kind of evidence an agent (or a scientific community) has for the two propositions. IA2 and IA2\* avoid this pitfall, but they saddle Friedman's theory with a separate implausibility since Friedman's theory, conjoined with IA2 or IA2\*, implies that universal generalizations never explain anything. We acknowledge that there may be other Bayesian conceptions of independent acceptability and that non-Bayesian epistemologies may have suggestions that are worth considering as well. So our conclusion is conditional: if independent acceptability is interpreted in any of the ways we have examined, Friedman's theory is in trouble.

## Appendix A

### Proof of FIA1\*<sub>IA2</sub>

Suppose, first, that  $S$  entails  $Q$ . Then, by FIA1<sub>IA2</sub>, it follows that  $S$  is not acceptable independently of  $Q$  in the sense of IA2.

Suppose, second, that  $S$  does not entail  $Q$ . Then (i)  $Q$  entails  $S$  or (ii)  $Q$  does not entail  $S$ . If i holds, then there are probability distributions on which  $p(S|E) > c$  but  $p(Q|E) \leq c$ . For example:

$E$	$S$	$Q$	$p$
T	T	T	0
T	T	F	17/35
T	F	T	0
T	F	F	0
F	T	T	15/89
F	T	F	0
F	F	T	0
F	F	F	1,077/3,115

It follows on this distribution that

$$p(S|E) = 1 > c > 0 = p(Q|E).$$

If ii holds, then there are probability distributions on which  $p(S|E) > c$  but  $p(Q|E) \leq c$ . For example:

$E$	$S$	$Q$	$p$
T	T	T	0
T	T	F	1/2
T	F	T	0
T	F	F	0
F	T	T	0
F	T	F	0
F	F	T	1/2
F	F	F	0

It follows on this distribution that

$$p(S|E) = 1 > c > 0 = p(Q|E).$$

Thus, either way,  $S$  is acceptable independently of  $Q$  in the sense of IA2.

Hence, FIA1\*<sub>IA2</sub> holds. QED

## Appendix B

### Proof of FIA1\*<sub>IA2\*</sub>

Suppose, first, that  $S$  and  $Q$  are logically equivalent to each other. Then, for any  $p$  and any  $E$ ,  $p(S) = p(Q)$  and  $p(S|E) = p(Q|E)$ . Hence, for any  $p$  and any  $E$ ,  $p(S|E) > c$  and  $p(S|E) > p(S)$  if and only if  $p(Q|E) > c$  and  $p(Q|E) > p(Q)$ . Hence,  $S$  is not acceptable independently of  $Q$  in the sense of IA2\*.

Suppose, second, that  $S$  and  $Q$  are not logically equivalent to each other. Then (i)  $S$  entails  $Q$  but not vice versa, (ii)  $Q$  entails  $S$  but not vice versa, or (iii) neither  $S$  nor  $Q$  entails the other. If i holds, then, for some  $p$  and some  $E$ ,  $p(S|E) > c$  and  $p(S|E) > p(S)$ , but  $p(Q|E) = p(Q)$ . For example:

$E$	$S$	$Q$	$p$
T	T	T	2/7
T	T	F	0
T	F	T	0
T	F	F	0
F	T	T	2/43
F	T	F	0
F	F	T	201/301
F	F	F	0

It follows on this distribution that

$$p(S|E) = 1 > c \text{ and } p(S|E) = 1 > 0.332 \approx p(S).$$

$$p(Q|E) = 1 = p(Q).$$

If ii holds, then, for some  $p$  and some  $E$ ,  $p(S|E) > c$  and  $p(S|E) > p(S)$ , but  $p(Q|E) = p(Q)$ . For example:

$E$	$S$	$Q$	$p$
T	T	T	2/7
T	T	F	10/63
T	F	T	0
T	F	F	0
F	T	T	5/14
F	T	F	0
F	F	T	0
F	F	F	25/126

It follows on this distribution that

$$p(S|E) = 1 > c \text{ and } p(S|E) = 1 > 0.802 \approx p(S).$$

$$p(Q|E) = p(Q) \approx 0.643.$$

If iii holds, then, for some  $p$  and some  $E$ ,  $p(S|E) > c$  and  $p(S|E) > p(S)$ , but  $p(Q|E) = p(Q)$ . For example:

$E$	$S$	$Q$	$p$
T	T	T	2/7
T	T	F	7/47
T	F	T	0
T	F	F	0
F	T	T	5/29
F	T	F	0
F	F	T	5,783/29,029
F	F	F	1,302/6,721

It follows on this distribution that

$$p(S|E) = 1 > c \text{ and } p(S|E) = 1 > 0.607 \approx p(S).$$

$$p(Q|E) = p(Q) \approx 0.657.$$

Hence, regardless of which of i–iii holds,  $S$  is acceptable independently of  $Q$  in the sense of IA2\*.

Thus, FIA1\*<sub>IA2\*</sub> holds. QED

## Appendix C

### Proof That FIA1 and FIA2 Are False When Interpreted in Terms of IA2\*

First, suppose that  $S$  entails  $Q$  but not vice versa. Then it is not the case that  $S$  and  $Q$  are logically equivalent to each other. Hence, by FIA1\*<sub>IA2\*</sub>,  $S$  is acceptable independently of  $Q$  in the sense of IA2\*. Hence, FIA1 is false when interpreted in terms of IA2\*.

Second, note that FIA2 when interpreted in terms of IA2\* is logically equivalent to:

**FIA2**<sub>IA2\*</sub> If  $S$  is not acceptable independently of  $Q$  in the sense of IA2\*, and  $Q$  entails  $P$ , then  $S$  is not acceptable independently of  $P$  in the sense of IA2\*.

Suppose that  $S$  and  $Q$  are logically equivalent to each other and that  $Q$  entails  $P$  but not vice versa. Then, by  $FIA1^*_{IA2^*}$ , it follows that  $S$  is not acceptable independently of  $Q$  in the sense of  $IA2^*$  but is acceptable independently of  $P$  in the sense of  $IA2^*$ . So  $S$  is not acceptable independently of  $Q$  in the sense of  $IA2^*$ , and  $Q$  entails  $P$ , but  $S$  is acceptable independently of  $P$  in the sense of  $IA2^*$ . Hence,  $FIA2_{IA2^*}$  is false. Hence,  $FIA2$  is false when interpreted in terms of  $IA2^*$ . QED

## REFERENCES

- Barnes, Eric. 1992a. "Explanatory Unification and the Problem of Asymmetry." *Philosophy of Science* 59:558–71.
- . 1992b. "Explanatory Unification and Scientific Understanding." In *PSA 1992: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1:3–12. East Lansing, MI: Philosophy of Science Association.
- . 1994. "Explaining Brute Facts." In *PSA 1994: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1:61–68. East Lansing, MI: Philosophy of Science Association.
- Bartelborth, Thomas. 1996. "Scientific Explanation." In *Structuralist Theory of Science: Focal Issues, New Results*, ed. Wolfgang Balzer and Carlos Moulines, 23–43. Berlin: de Gruyter.
- . 2002. "Explanatory Unification." *Synthese* 130:91–107.
- de Regt, Henk, and Dennis Dieks. 2005. "A Contextual Approach to Scientific Understanding." *Synthese* 144:137–70.
- Douglas, Heather. 2009. "Reintroducing Prediction to Explanation." *Philosophy of Science* 76:444–63.
- Friedman, Michael. 1974. "Explanation and Scientific Understanding." *Journal of Philosophy* 71:5–19.
- . 1981. "Theoretical Explanation." In *Reduction, Time and Reality: Studies in the Philosophy of the Natural Sciences*, ed. Richard Healey, 1–16. Cambridge: Cambridge University Press.
- . 1983. *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*. Princeton, NJ: Princeton University Press.
- Fry, Richard. 2016. "Backwards Explanation and Unification." *European Journal for Philosophy of Science* 6:55–70.
- Gemes, Ken. 1994. "Explanation, Unification, and Content." *Noûs* 28:225–40.
- Gijssbers, Victor. 2007. "Why Unification Is Neither Necessary Nor Sufficient for Explanation." *Philosophy of Science* 74:481–500.
- . 2013. "Understanding, Explanation, and Unification." *Studies in History and Philosophy of Science* 44:516–22.
- Griesmaier, Franz-Peter. 2005. "Kitcher-Style Unificationism and Explanatory Relevance." *Dialectica* 59:37–50.
- Halonen, Ilpo, and Jaakko Hintikka. 1999. "Unification: It's Magnificent but Is It Explanation?" *Synthese* 120:27–47.
- Hempel, Carl. 1965. "Aspects of Scientific Explanation." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, 331–496. New York: Free Press.
- Humphreys, Paul. 1993. "Greater Unification Equals Greater Understanding?" *Analysis* 53:183–88.
- Jenkins, C. S., and Daniel Nolan. 2008. "Backwards Explanation." *Philosophical Studies* 140:103–15.
- Jones, Todd. 1995a. "How the Unification Theory of Explanation Escapes Asymmetry Problems." *Erkenntnis* 43:229–40.
- . 1995b. "Reductionism and the Unification Theory of Explanation." *Philosophy of Science* 62:21–30.
- . 1997. "Unification, Reduction, and Non-ideal Explanations." *Synthese* 112:75–96.

- . 1998. “Unification, Deduction, and History: A Reply to Steel.” *Philosophy of Science* 65:672–81.
- . 2008. “Unification.” In *The Routledge Companion to Philosophy of Science*, ed. Stathis Psillos and Martin Curd, 489–97. London: Routledge.
- Kitcher, Philip. 1976. “Explanation, Conjunction, and Unification.” *Journal of Philosophy* 73:207–12.
- . 1981. “Explanatory Unification.” *Philosophy of Science* 48:507–31.
- . 1989. “Explanatory Unification and the Causal Structure of the World.” In *Scientific Explanation*, ed. Philip Kitcher and Wesley Salmon, 410–505. Minneapolis: University of Minnesota Press.
- Kyburg, Henry. 1961. *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press.
- . 1965. “Salmon’s Paper.” *Philosophy of Science* 32:147–51.
- Mares, Edwin. 2014. “Relevance Logic.” In *The Stanford Encyclopedia of Philosophy*, ed. Edward Zalta. Stanford, CA: Stanford University. <http://plato.stanford.edu/entries/logic-relevance/>.
- Petkov, Stefan. 2015. “Explanatory Unification and Conceptualization.” *Synthese* 192:3695–3717.
- Psillos, Stathis. 2002. *Causation and Explanation*. Montreal: McGill-Queen’s University Press.
- Putnam, Hilary. 1975. “Philosophy and Our Mental Life.” In *Mind, Language, and Reality*, 291–303. Cambridge: Cambridge University Press.
- Salmon, Wesley. 1971. “Statistical Explanation.” In *Statistical Explanation and Statistical Relevance*, ed. Wesley Salmon, 29–87. Pittsburgh: University of Pittsburgh Press.
- . 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- . 1989. “Four Decades of Scientific Explanation.” In *Scientific Explanation*, ed. Philip Kitcher and Wesley Salmon, 3–219. Minneapolis: University of Minnesota Press.
- Schurz, Gerhard. 1999. “Explanation as Unification.” *Synthese* 120:95–114.
- . 2014. “Unification and Explanation: Explanation as a Prototype Concept; A Reply to Weber and van Dyck, Gijsbers, and de Regt.” *Theoria* 29:57–70.
- Schurz, Gerhard, and Karel Lambert. 1994. “Outline of a Theory of Scientific Understanding.” *Synthese* 101:65–120.
- Skipper, Robert. 1999. “Selection and the Extent of Explanatory Unification.” *Philosophy of Science* 66 (Proceedings): S196–S209.
- Sober, Elliott. 2003. “Two Uses of Unification.” In *The Vienna Circle and Logical Empiricism: Re-evaluation and Future Perspectives*, ed. Friedrich Stadler, 205–16. Dordrecht: Kluwer.
- . 2012. “Coincidences and How to Reason about Them.” In *EPSA Philosophy of Science Association: Amsterdam, 2009*, ed. Henk de regt, Stephan Hartmann, and Samir Okasha, 355–74. Amsterdam: Springer.
- . 2015. *Ockham’s Razors: A User’s Manual*. Cambridge: Cambridge University Press.
- Strevens, Michael. 2004. “The Causal and Unification Approaches to Explanation Unified—Causally.” *Noûs* 38:154–76.
- Weber, Erik, and Jeroen van Bouwel. 2009. “Causation, Unification, and the Adequacy of Explanations of Facts.” *Theoria* 24:301–20.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- . 2014. “Scientific Explanation.” In *The Stanford Encyclopedia of Philosophy*, ed. Edward Zalta. Stanford, CA: Stanford University. <http://plato.stanford.edu/entries/scientific-explanation/>.