NOÛS 28:2 (1994) 225–240

Explanation, Unification, and Content

KEN GEMES
Yale University

Synopsis: The following is an essay on the notion of explanation as unification. In it a new notion of (logical) content developed in Gemes (1993) is used to explicate Michael Friedman's notion of "k-atomicity," and to explicate the notion of the surplus content of hypothesis h relative to evidence e. From this basis an analysis of unification as theoretical reduction is advanced. A second notion of unification, unification as reconciling prima facie incompatible statements, is introduced again with the aid of this new notion of content. More generally, it is argued that rather than seek the essence of explanation we should carefully catalog the various distinct explanatory virtues.

1. Against A Definition of Explanation

What is the nature of explanation? In his "Explanation and Scientific Understanding" Michael Friedman advanced the idea that explanation is unification. The unification Friedman has in mind occurs when a vast number of apparently disparate phenomena are brought under the reign of a few compact laws. For instance, Friedman cites the kinetic theory of gases as effecting

> a significant unification in what we have to accept. Where we once had three independent brute facts—that gases approximately obey the Boyle-Charles gas 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. (Friedman 1974, pp. 14–15).

What makes for unification, according to Friedman, is a reduction of "the total number of independent phenomena that we have to accept as ultimate or given." (loc. cit.).

Unfortunately, Friedman's semi-formal account of this notion of explanation

as unification as reduction of independent phenomena is open to some telling criticisms, as demonstrated by Philip Kitcher (Kitcher 1976) and below (Cf. Section 3). Kitcher himself remains sympathetic to Friedman's identification of explanation and unification and proposes his own account of what is involved in such unification (Kitcher 1981).

Besides the idea that explanation is unification, Friedman and Kitcher share a tacit assumption common to many philosophers of science. They assume that there is a *nature* of explanations, as if explanations formed a natural kind to be demarcated by a set of necessary and sufficient conditions. Now perhaps this assumption is correct, but no one has yet provided *any* reason (good or bad) for making it. Therefore, until positive reasons for accepting it are forthcoming, I suggest the following more modest approach. Let us attempt to catalog and elucidate various explanatory virtues without assuming that they can somehow be fashioned into a set of necessary and sufficient conditions for explanation. Here we leave open the plausible possibility that different explanations contain different explanatory virtues to different degrees. Given a sufficient catalog of explanatory virtues we can then discuss the merits and demerits of various purported explanations. Where disagreements about explanatory power arise it will be interesting to find out whether the disputants disagree because they differ in their assessments of whether the relevant explanation has a particular explanatory virtue, as against, for instance, disagreeing because they place different emphasis on different virtues.

As an entree into this admittedly ambitious project I want to consider Friedman's and Kitcher's preferred explanatory virtue, namely, unification.

2. Unification as Reduction

Suppose we have a number of independent phenomena described by the independent sentences S1, S2, ..., Sn—what is involved in independence need not concern us just yet. It is tempting to think that if we find a sentence S which describes all the phenomena described by S1, S2, ..., Sn, we have thereby effected a reduction of S1, S2, ..., Sn. However this temptation soon fades when we realize that where S is simply the conjunction of S1, S2, ..., Sn, we have a sentence that describes all the relevant phenomena without effecting any reduction. The problem here is that we have merely replaced n independent sentences with a sentence with n independent parts. To make any progress here we need an understanding of how to determine how many independent parts a sentence has.

Friedman proposes that we count a sentence S as being acceptable independently of another sentence S′ if there are sufficient grounds for accepting S which are not sufficient for accepting S′. Independent acceptability, according to Friedman, satisfies the following conditions:

(1) If S⊢Q then S is not acceptable independently of Q.
(2) If S is acceptable independently of P and Q⊢P then S is acceptable independently of Q. (Friedman 1974, pp.16–17).

To this we may presumably add the condition

(3) If S is acceptable independently of P and S⊢Q then Q is acceptable independently of P.[1]

Friedman then sets about defining reduction as follows:

> Let a *partition* of a sentence S be a set of sentences Γ such that Γ is logically equivalent to S and each S′ in Γ is acceptable independently of S... . I will say that a sentence is *K-atomic* if it has no partition i.e. if there is no pair {S1, S2} such that S1 and S2 are acceptable independently of S and S1&S2 is logically equivalent to S... . Let a *K-partition* of a set of sentences Δ be a set Γ of K-atomic sentences which is logically equivalent to Δ (I assume that such a K-partition exists for every set Δ). Let the *K-cardinality* of a set of sentences Δ, K-card (Δ), be inf {card (Γ): Γ is a K-partition of Δ}... . Finally I will say that S *reduces* the set Δ iff K-card (Δ ∪ {S}) < K-card (Δ). (loc.cit.)

He then gives an account of explanation in terms of this definition of reduction. He introduces the term **con(S)** to stand for the set of independently acceptable consequences of S, and defines explanation as follows:

> S1 explains S2 iff S2 ∈ con(S1) and S1 reduces con(S1).

Kitcher has shown that these definitions have the unpalatable consequence that only K-atomic sentences can explain. Briefly, the problem is that if S1 is not k-atomic con(S1) is equivalent to S1. To see this we need merely recall that if S1 is not k-atomic there are sentences S3 and S4 such that S3 and S4 are independently acceptable of S1 and ⌜S3&S4⌝ is logically equivalent to S1. In this case since both S3 and S4 are consequences of S1 and are independently acceptable of S1 they are members of con(S1). Therefore, since the conjunction of S3 and S4 is logically equivalent to S1, and both are members of con(S1), con(S1) entails S1. On the other hand since every member of con(S1) is a consequence of S1, S1 entails con(S1). Therefore (con(S1) ∪ {S1}) is equivalent to con(S1). Since (con(S1) ∪ {S1}) and con(S1) are equivalent and in determining the K-cardinality of a set we are to consider any set logically equivalent to the set in question, in determining the K-cardinality of (con(S1) ∪ {S1}) and con(S1) we will be considering the same sets. In other words, they have the same K-cardinality and hence S1 does not reduce con(S1) and hence S1 does not explain S2.

This technical difficulty is related to an important intuitive question: Why should a sentence, or set of sentences, S1 reduce some sub-class of its own consequence class, for instance con(S1), if it is to explain another sentence S2? If S2 is a member of a sub-class S, say {S2, S3, ..., Si}, of S1's consequence class and the K-cardinality of S1 is less than the K-cardinality of S it seems reasonable to say that S1 reduces S. In this case we have reduction because while S1 contains all of the content of S it contains fewer independently acceptable content

parts than does S. This may not give us much of a handle on what is involved in S1 explaining S2, but it gives a neat idea of what is involved in a sentence, or perhaps class of sentences, reducing, hence unifying, a class of sentences. And this indeed was one of the problems Friedman originally set out to illuminate.

I have previously suggested it is mistaken to take one explanatory virtue, for instance reduction, as the essence of explanation. When we couple this with the technical difficulties facing Friedman's account of explanation we have sufficient incentive to drop Friedman's proposed account of explanation and see if we can salvage from it an account of what we may here call r-unification—'r' for reduction. However, before proceeding to offer such an account we need to amend Friedman's account of K-atomicity. In amending Friedman's account of K-atomicity we will need to make a detour to consider what exactly should count as part of the contents of a statement. This detour will have a further pay-off in allowing us to explicate a new notion of reduction in the later parts of this paper.

3. Content and K-atomicity

Consider the following case: Die A has just been rolled and I have the following reports from Peter, Paul, and Luke all of whom I take to be perfectly reliable die reporters: Peter says A came up 1, 2 or 3, Paul says A came up 1, 4 or 5 and Luke says it came up 1. Now the claim S, 'A came up 1,' is equivalent to the conjunction of the claims S′, 'A came up 1,2 or 3,' and S″, 'A came up 1,4 or 5.' Clearly I have sufficient grounds for accepting each of these claims, on the one hand the testimony of Peter, on the other the testimony of Paul, which taken individually are not grounds for accepting S. Thus by Friedman's account S is not K-atomic. This, presumably, is extremely counter-intuitive. Moreover suppose there are sentences A and B such that you have sufficient reason for accepting A and sufficient reason for accepting B, however unbeknownst to you A and B are inconsistent. For instance, suppose A is a complex mathematical claim that a reliable book you have read cites as a theorem and B is another mathematical claim for which you yourself have fashioned a proof and that unbeknownst to you B entails not A. In this case for you, for any sentence S such that A and B are both independently acceptable of S, S is not K-atomic. This follows since for any such sentence S, S is equivalent to $\ulcorner(SvA)\&(SvB)\urcorner$ and both A and B, and hence $\ulcorner SvA\urcorner$ and $\ulcorner SvB\urcorner$ are independently acceptable of S.

That a statement S can be factored into independently acceptable disjunctive statements containing disjuncts with content completely extraneous to the content of S should have no bearing on the question of whether S is k-atomic. The question is not whether S can be partitioned by any old means into consequences independently acceptable of S. The important question is whether S can be partitioned into distinct *content parts* independently acceptable of S. The difference here is that while $\ulcorner SvA\urcorner$ and $\ulcorner SvB\urcorner$ will count as consequences of S for arbitrary S, B, and A, they will not generally count as content parts of S. The account of content invoked here is developed primarily in Gemes (1994). While

this is not the appropriate place to fully rehearse my new account of content we can introduce a simple enough surrogate.

Let α be variable for well-formed formulae (wffs) of the language in question. Let β be a variable for wffs and sets of wffs of the language in question. A set of wffs β is contingent if and only if there is some contradiction μ such that $\beta \nvdash \mu$ and some ϕ such that $\phi \in \beta$ and $\{/\} \nvdash \phi$. Then, presuming the notion of atomic wff is defined for the language in question, we define content as follows, using '$\alpha < \beta$' to abbreviate 'α is a content part of β',

D1 $\alpha < \beta =_{df}$ α and β are contingent, $\beta \vdash \alpha$, and there is no σ such that $\beta \vdash \sigma$, σ is stronger than α and every atomic wff that occurs in σ occurs in a.[2]

We say σ is stronger than α where $\sigma \vdash \alpha$ and $\alpha \nvdash \sigma$. Under this notion of content *Fa* is part of the content of *(Fa & Fb)* but *(Fa v ~Fb)* is not. *(Fa v ~Fb)* is not a part of *(Fa & Fb)* because *Fa* is a consequence of *(Fa & Fb)* that is stronger than *(Fa v ~Fb)* and every atomic wff that occurs in *Fa* occurs in *(Fa v ~Fb)*. Similarly, *(x)(Fx)* is a content part of *(x)(Fx & Gx)* but *(x)(Fx v Hx)* is not. *(x)(Fx v Hx)* is not a content part of *(x)(Fx & Gx)* because *(x)(Fx)* is a consequence of *(x)(Fx & Gx)* which is stronger than *(x)(Fx v Hx)* and contains only atomic wffs that occur in *(x)(Fx v Hx)*. Applied to ordinary English, with the presumption that we have some idea of what are to count as atomic sentences of English, our definition yields the results, for instance, that 'Ken is in Sydney or Melbourne' is not part of the content of 'Ken and Alan are in Sydney' but 'Ken is in Sydney' is part of its content.[3]

Where '$\vdash$' is read as logical entailment the above definition D1 suffices to capture the (syntactical) notion of logical content. However under such a reading it does not capture the ordinary notion of semantic content according to which 'Ken is an unmarried man' is part of the content of 'Ken is a bachelor', and 'Pecky is a animal' is part of the content of 'Pecky is a sheep'. To capture this we need to read '$\vdash$' as including analytic as well as logical entailments.[4]

Let us return now to our original problem for Friedman's notion of k-atomicity. If for k-atomicity we demand that there be no independently acceptable content parts, rather than merely no independently acceptable factors, our problem cases disappear. In particular, where A and B are mathematical statements immaterial to S, neither ⌜SvA⌝ nor ⌜SvB⌝ will be content parts of S. Similarly, 'Die A came up 1, 4 or 5' is not a content part of 'Die A came up 1.'

Indeed, if we exchange Friedman's demand that k-atomic sentences not be divisible into independently acceptable consequences for the demand that they not be divisible into independently acceptable content parts we render Friedman's account immune from counter-examples developed by Wes Salmon.[5] Salmon (1989, pp.96–99) claims that under Friedman's account law statements of the conditional form

(1) $(x)(Fx \supset Gx)$

will typically not be k-atomic since there will typically be some predicate H such that (1) is equivalent to the conjunction of the two members of the set

(2) $\{(x)(Fx\&Hx \supset Gx), (x)(Fx\&\sim Hx \supset Gx)\}$

where both members of the set are acceptable independently of (1). Now if we demand that k-atomic statements not be divisible into independently acceptable content parts Salmon's counterexample does not hold since neither of the members of (2) are content parts of (1).

This, then, is why the statement consisting of a conjunction of Boyle-Charles law and Graham's law is not k-atomic: It has a content part, for instance Graham's law, which is acceptable independently of the conjunction in question. Conjunctions of independently acceptable statements do not achieve any real unification of their conjuncts because they contain as independent content parts the very contents allegedly being unified. That is why the conjunction of the Boyle-Charles law, Graham's law, Galileo's law and Kepler's laws does not reduce the set of its conjuncts.

Here then is an alternative definition of K-atomicity:

D1 S is K-atomic $=_{df}$ There are no content parts S1 and S2 of S such that S1 and S2 are acceptable independently of S and each other and $\ulcorner$S1 & S2$\urcorner$ is logically equivalent to S.

We may keep Friedman's original definition of a K-partition,

D2 L is a K-partition of S $=_{df}$ L is a set of k-atomic sentences such that L is logically equivalent to S

and K-cardinality,

D3 K-card S $=_{df}$ inf{card (Γ):Γ is a K-partition of S}.

4. R-unification

Finally we define the notion of r-unification as follows:

D4 S r-unifies S′ $=_{df}$ There is some subset S* of S such that S*$\vdash$S′ and K-card S* $<$ K-card S′.

The reference to subsets occurs in order to allow the possibility, for instance, that S may contain distinct subsets S* and S** which respectively reduce different sets of sentences S1 and S2. In this case we need to focus on the K-cardinality of S* against S1 and S** against S2 rather than, for instance, S against S1.

Is what I have called r-unification equivalent to the plain old notion of reduction? While all cases of r-unification count as cases of reduction I believe there are some cases of reduction which will not satisfy the criteria for r-unification. Consider those cases where we discover that a number of seemingly independent laws can be deduced from a significantly smaller number of laws combined with a variety of facts about initial conditions. Such cases often qualify as bona fide reductions. However if the set of initial conditions is large enough it may be that the K-cardinality of the set of reducing laws and relevant initial conditions is larger than the K-cardinality of the set of reduced laws.[6] Of course, if we had clear criteria for distinguishing law statements from statements of initial conditions we could speak here of a reduction of laws. However, failing this we may still regard r-unification as a species of reduction and as such it is an explanatory virtue that speaks well for any theory that possesses it.

5. Unification as the Reconciling of Incompatible Claims

Having suggested that unification through reduction is not all there is to explanation, I now want to argue that unification through reduction is not the sole form of explanatory unification.

Consider the following kind of (partial) explanation: The explanation of the fact that Ron is in Indiana and in Illinois is that he is in fact presently straddling a state line and that Indiana and Illinois share a state line. Here the explanans is effective because it reconciles prima facie conflicting facts in the explicandum. The tension between the two facts, that Ron is in Indiana and that Ron is in Illinois, is resolved by the putative explanation. Such explanations effect a unification of prima facie incompatible phenomena. This type of explanation occurs in the mathematical, the natural, and the social sciences, as well as in everyday common sense settings. For instance, the posit of the ether had the explanatory virtue of unifying conflicting claims from post-Newtonian mechanics and electrodynamics. Laws of statistical mechanics have the benefit of reconciling the prima facie incompatible claims that the fundamental laws of physics are time-reversal symmetric yet laws governing macroscopic observable phenomena are often time-reversal asymmetric. Early formulations of quantum theory gained much plausibility from their ability to reconcile the claims that light exhibited both a wave-like and a particulate nature. Perhaps the chief explanatory virtue of Freud's theory of unconscious desires is that it is capable of unifying apparently conflicting phenomena, for instance, a subject's apparently sincere claims to love his father with his repeated obsessive thought that his father is experiencing horrible torture. I suspect that it is chiefly this virtue, to the exclusion of just about every other explanatory virtue, that makes Freudianism so attractive. If so this should makes us ever more insistent in demanding that putative explanations display a robust mix of explanatory virtues rather than exhibiting only one or two while excluding all others.

The type of unification involved in the above cases, where prima facie conflicting claims are rendered compatible by the introduction of new data, is clearly

different from the type of unification considered by Friedman and Kitcher, where disparate phenomena are brought under a common description. For Kitcher's and Friedman's unification there need be no initial tension, no incompatibility, between those phenomena that are being unified. Where prima facie conflicting claims are rendered compatible there need be no reduction of independent phenomena.

6. Weak Incompatibility Versus Genuine Incompatibility

Before defining what it is to render incompatible claims compatible we first need an account of what it is for claims to be incompatible in the first place. If A is incompatible with B then believing A gives some reason for not believing B. In other words, A on the evidence of B is less credible than A in the absence of B. This relationship can be neatly captured in the language of the probability calculus as follows: $P(A/B) < P(A)$, that is, the probability of A given evidence B, is less than the probability of A in the absence of any such evidence. Indeed this seems to capture the incompatibility involved in our state line case. In that case the claim that Ron is in Indiana has a lower probability on the evidence that he is in Illinois than it has in the absence of such evidence. This suggests the following definition,

A is incompatible with B $=_{df}$ $P(B/A) < P(B)$.

It is encouraging to note that incompatibility so defined is a symmetric relation. It follows from the probability calculus that $P(B/A) < P(B)$ if and only if $P(A/B) < P(A)$. Symmetry of incompatibility is exactly what we would expect; if A is incompatible with B, then surely it follows that B is incompatible with A.

Unfortunately, this proposed definition of incompatibility is too weak. Let A be the statement 'Fred graduated in the top 60% of his class' and B be the statement 'Fred graduated in the bottom 50% of the class.' On our present definition A and B are incompatible since, prima facie, $P(A) = 6/10$ and $P(A/B) = 1/5$. The problem here is that while B decreases the probability of A, it is obvious how A and B may be both be true; it simply need be the case that Fred graduated in the 40–50% range. It seems specious to say this latter claim renders A and B compatible because it is so obvious a claim, given A and B, that it is spurious to say A and B are incompatible in the first place.

To render our claims about Ron compatible we had to introduce a new notion, namely, that of straddling a state line. Similarly, Freud introduces new notions, such as that of the Oedipal complex, in order to reconcile conflicting attitudes subjects display towards their parents. Similarly, the posit of the ether, and the recent posit of super-positions, involve the introduction of new concepts in order to reconcile prima facie conflicting claims. In our case of the student Fred we did not need to introduce any new conceptual material in order to see how both the

claim that he came in the top 60% of the class and the claim that he came in the bottom 50% could be true.

This suggests how we may strengthen our initial definition of incompatibility so that it covers our state line case but excludes the case of Fred the student. For A and B to be genuinely incompatible it is not sufficient that $P(A/B) < P(A)$. It is also necessary that A and B can only be rendered compatible by introducing new conceptual material, that is, conceptual material that is foreign to A and B.

Are we making any progress? At first blush it may seem that we are going around in circles: We deferred the problem of defining what it is to render conflicting claims compatible in favor of defining what it is for claims to be incompatible in the first place. Then, in attempting to define incompatibility, we found we needed recourse to the notion of rendering incompatible claims compatible. In fact I think we have been making progress, albeit somewhat obliquely. We have the beginnings of a notion of incompatibility in terms of unfavorable relevance. If we can use this admittedly weak notion of incompatibility to define the notion of rendering compatible, we may return and define the notion of genuine incompatibility in terms of weak compatibility and rendering compatible by the introduction of new conceptual material. The root idea is that A and B are genuinely incompatible if and only if A and B are weakly incompatible and they may only be rendered compatible by the introduction of conceptual resources foreign to both A and B.

Here then is our first definition:

D5 A and B are weakly incompatible $=_{df} P(A/B) < P(A)$.

Now what is involved in rendering weakly incompatible A and B compatible? Well, if the incompatibility of A and B involves B counting against A, in the sense of being unfavorably relevant to A, could we not say that C renders incompatible A and B compatible if in the presence of C, B no longer counts against A?[7] In the language of the probability calculus we may express this as follows: $P(A/B\&C) \geq P(A/C)$. This seems to capture our state line case well enough. In that case the information that Jones is in Illinois, in the presence of the claim that Jones is straddling a state line and that Indiana and Illinois share a state line, serves to confirm that Jones is in Indiana. Here then is the beginnings of a definition:

C renders A and B compatible $=_{df}$ (i) $P(A/B) < P(A)$ and (ii) $P(A/B\&C) \geq P(A/C)$.

While this is a fair first step it is clearly not the whole story. Let A be 'Ron is in Indiana,' B be 'Ron is in Illinois' and C be 'If Ron is in Illinois then Ron is in Indiana.' In this case $P(A/B) < P(A)$ and $P(A/B\&C) > P(A/C)$, yet clearly C does not render A and B compatible. Part of the problem here is that C is itself a

consequence of A and hence does not tell us anything new to help reconcile A and B. If C is to render A and B compatible then it must add something new, some new content not contained in A or B, in virtue of which the content of B no longer counts against A. In seeing whether C really reconciles A and B we need to look not simply at C but at that part of C which goes beyond (the conjunction of) A and B. This might aptly be called the surplus content of C relative to A and B. The surplus content of C relative to A and B, that is (C-A&B), is the set of all content parts C′ of C, such that no content part of C′ is truth functionally dependent on any content part of ⌜A&B⌝.[8] For instance, where H is 'Ron is in Sydney & Alan is in Pittsburgh' and E is 'Ron is in Sydney' the surplus content of H relative to E is (equivalent to) 'Alan is in Pittsburgh.' In many typical cases the surplus content of C relative to A and B is simply C itself. Nevertheless we need to makes use of the notion of surplus content to rule out cases such as that of the conditional 'If Ron is in Illinois then Ron is in Indiana' as a potential unifier of 'Ron is in Indiana' and 'Ron is in Illinois'.

Now we may amend our definition as follows:

C renders A and B compatible $=_{df}$ (i) $P(A/B) < P(A)$ and (ii) $P(A/B\&(C\text{-}A\&B)) \geq P(A/(C\text{-}A\&B))$.[9]

According to this definition C, 'If Ron is in Illinois then Ron is in Indiana' does not render A, 'Ron is in Indiana', and B 'Ron is in Illinois' compatible. Since A, 'Ron is in Indiana,' entails C, 'If Ron is in Illinois then Ron is in Indiana,' there is no part of C that goes beyond A and B. Here, (C-A&B) = {/}. So, for the right hand side of the above definition to be fulfilled we need, per impossible, (i) $P(A/B) < P(A)$ and (ii) $P(A/B) \geq P(A)$. On the other hand where C is 'Ron is straddling a state-line and Indiana and Illinois share a state line,' (C-A&B) is identical to C and the right hand side of our definition is fulfilled.

Yet consider the following case: A is 'Smith has a chess rating of 2100 and Jones has a rating of 1800,' B is 'Jones just beat Smith in four consecutive Chess games,' and C is 'Smith has been humoring Jones in order to boost his confidence.' Prima facie, in this case A and B are incompatible and C renders them compatible. Yet here, arguably, the right hand side of our proposed definition is not satisfied since presumably $P(A/B\&(C\text{-}A\&B)) \not\geq P(A/(C\text{-}A\&B))$. The fact that Jones just beat Smith in four consecutive games does not, even in the presence of the claim that Smith is humoring Jones, cease counting against the chances that Smith is a much better player than Jones. In this case the point is not that in the presence of the new information provided by C, B ceases to count against A. Rather the point is that, in the presence of the new information in C, B loses much of its original sting against A. In other terms, $P(A/B\&(C\text{-}A\&B)) > P(A/B)$.

We need here to distinguish two different kinds of cases. In the first case the new information in C is itself favorable to A. In this case it is not enough that

$P(A/B\&(C\text{-}A\&B)) > P(A/B)$. For it may be that the new information in C is merely inductive evidence for A that in no way helps to reconcile A and B. Here our original definition in terms of $P(A/B\&(C\text{-}A\&B)) \geq P(A/(C\text{-}A\&B))$ seems to fit the bill. In the second case where the new information is not favorable to A it is sufficient (for reconciliation) that $P(A/B\&(C\text{-}A\&B)) > P(A/B)$. Actually, since the new information in C may have a different bearing on B than it has on A, we need to complicate the picture a little. Here then is our full definition, letting A, B and C range over sets of sentences:

D6 C renders A and B compatible $=_{df}$
(i) $P(A/B) < P(A)$ and
(ii) $P(A/C) \leq P(A)$ and $P(A/B\&(C\text{-}A\&B)) > P(A/B)$, or
$P(B/C) \leq P(B)$and $P(B/A\&(C\text{-}A\&B)) > P(B/A)$, or
$P(A/C) > P(A)$ and $P(A/B\&(C\text{-}A\&B)) \geq P(A/(C\text{-}A\&B))$, or
$P(B/C) > P(B)$ and $P(B/A\&(C\text{-}A\&B)) \geq P(B/(C\text{-}A\&B))$.

We may now define genuine incompatibility as follows:

D7 A and B are genuinely incompatible $=_{df}$ A and B are weakly incompatible and any C that renders them compatible involves the introduction of new conceptual content.

While I do not here propose a formal mechanism for deciding what exactly is involved in the introduction of new conceptual content I presume the general idea is fairly clear. For instance, the introduction of the concept of straddling a state line represents new conceptual content relative to the conceptual content of the claims that an individual is in Indiana and an individual is in Illinois, whereas the introduction of the concept of being within the 40–50% range does not represent new conceptual content with respect to the claims that an individual is in the 0–50% range and an individual is in the 40–60% range. In the actual historical cases discussed below the fact that they involve new conceptual content will be readily apparent.

7. C-Unification with Examples

A claim may be said to provide c-unification ('c' for compatibility) of our set of beliefs if it renders some genuinely incompatible members of that set compatible. More formally,

D8 C c-unifies A and B if A and B are genuinely incompatible and C renders A and B compatible

Many of the greatest scientific hypotheses, including both successful and unsuccessful hypotheses, have had the explanatory virtue of providing c-unifica-

tion for sets of beliefs widely accepted at the time of their proposal. Actually, this overstates the case a little. Usually such cases do not simply involve the c-unification and continued acceptance of genuinely incompatible theories. Rather they involve the acceptance of new claims intended to render the conflicting theories compatible plus a slight modification of the original theories. For instance, in the case of the explanatory posit of the ether it is not the case that in light of this posit all of the old post-Newtonian mechanics could still be accepted. The ether theory led to the abandonment of the claim that there are no absolute velocities. In such cases we may say that the new posit provides c-unification in the sense that it renders compatible substantial content parts of the original theories.

Here, in brief, are some examples which illustrate the different ways new conceptual content may provide c-unification by rendering genuinely incompatible statements compatible.[10]

The fist case is loosely drawn from Breuer's and Freud's case study of Anna O (Cf. Breuer and Freud 1960). Let A be the statement 'Anna O refuses to drink water' and B be the statement 'Anna O sincerely claims to have no reason for disliking water'. Here $P(A/B) < P(A)$. Now let C be the conjecture 'Anna O has a strong unconscious association of water with an unpleasant event in her life'. Here, presumably, $P(B/C) \leq P(B)$ and $P(B/A \,\&\, (C\text{-}A\&B)) \geq P(B/A)$. In particular, while the claim that Anna O has a a strong unconscious association of water with an unpleasant event in her life does not make it more probable that she sincerely claims to have no reason for disliking water, the former claim lessens the sting of her refusal to drink water against the later claim. Here the introduction of the new concept of unconscious associations helps reconcile the observed behavior of refusing to drink water and the sincerity of the disavowal of reasons for refusing to drink.

Our second case concerns the use of statistical mechanics to reconcile claims about time symmetry. Let A be the claim that the fundamental laws of physics are time-symmetric and B be the claim that the observed behavior of macroscopic entities exhibits time-asymmetry. Here $P(A/B) < P(A)$. Let C be some standard version of statistical mechanics. Then, presumably, $P(A/C) > P(A)$ and, arguably, $P(A/B \,\&\, (C\text{-}A\&B)) \geq P(A/(C\text{-}A\&B))$. In particular, while statistical mechanics makes it more probable that the fundamental laws of physics are time-symmetric, in the presence of statistical mechanics, the observed time-asymmetry in the behavior of macroscopic objects has no sting against the claim that fundamental laws are time-symmetric. For instance, Boltzmann's H-Theorem demonstrates that the arrow of time is nothing but a representational feature of the coarse graining of confirmational variables. Here the introduction of the concepts of statistical mechanics, for instance the notion of coarse graining, allows us to reconcile the observed time-asymmetric behavior of macroscopic objects with the claim about the time-symmetry of fundamental laws.

Our third case, which concerns a conflict between known properties of elementary particles and claims about those properties derived from quantum elec-

trodynamics, might be considered a more controversial example of c-unification. Indeed, by discussing its controversial nature I hope to dramatize some of the claims about explanatory virtues made above.

According to the first principles of quantum electrodynamics (Q.E.D.), as formulated in Dirac's Lagrangian realization, perturbative calculations of the mass and charge of elementary particles, for instance electrons, yield infinite or indeterminate (e.g. ∞, $-\infty$, ∞/∞) quantities. This conflicts with, among other things, basic electrostatics, for instance the fact that electrons are capable of stable orbits in atoms. Renormalization Theory, as developed by Feynman, Schwinger and Tomonaga, gives an interpretation of these infinities through proper rescaling and in so doing provides a prescription for calculating with them. In this renormalized representation the indeterminacies in charge and mass of elementary particles are removed and upon proper scaling the actual calculations of mass and charge yields finite quantities in agreement with basic electrostatic experimentation including measurements of the electron charge as obtained using the Wilson cloud chamber. Simplifying somewhat, let A be the statement 'As determined by Dirac's Lagrangian realization, there are electrons with infinite or indeterminate mass and charge', B be the statement 'As determined by electrostatic experimentation, electrons have only finite and determinate mass and charge', and C be the statement 'The scaling involved in the claim that electrons have infinite or indeterminate charge is radically different from the scaling involved in the claim that electrons have only finite and determinate mass and charge'. Here again $P(A/B) < P(A)$, $P(A/C) \leq P(A)$ and $P(A/B \ \& \ (C\text{-}A\&B)) > P(A/B)$. In other words, while the claim that electrons as determined by Dirac's theory have infinite and indeterminate mass and charge counts against the claim that electrons as determined by electrostatic experimentation have finite and determinate mass and charge, in the presence of the claims that different scalings are involved in these two claims, the claim that they have finite and determinate mass and charge loses much of its sting against the claim that they have infinite and indeterminate mass and charge.

What is controversial in this example is the claim that Renormalization Theory really reconciles the cited aspects of quantum electrodynamics and basic electrostatics. Against this claim it might be argued that Renormalization Theory is merely a mathematical tool which only apparently reconciles conflicting claims but effects no real reduction in the tension between them. Here, I think, the basis of complaint is the belief that Renormalization Theory is seen merely as a way of allowing certain quantities to be calculated using various sophisticated mathematical techniques without diminishing the basic ontological conflict between Q.E.D. and basic electrostatics. To those who would therefore deny that Renormalization Theory really helps reconcile incompatible claims I would make the following response: You are confusing two different virtues, namely the virtues of reconciling incompatible claims and the virtue of having an ontological grounding. Renormalization Theory really has the first virtue but lacks the

second. Those who are greatly struck by its lack of the second virtue are prone to say it is not really an explanatory theory at all. In so doing they are prone to misextrapolate from the lack of an ontological grounding and claim that it therefore can have no explanatory virtue such as that of being able to reconcile conflicting claims. I think it far better for the advocates of the virtue of ontological grounding to concede that it has the explanatory virtue of reconciling conflicting claims and then argue that since it lacks the, by their lights, all important virtue of having a clear ontological grounding it should not count as a good, or, at least, full explanation. In so doing they will come to the heart of the conflict between the opponents and advocates of Renormalization Theory, that is, their differing answers to the question of the importance of the explanatory virtue of having an ontological grounding.

Finally, and perhaps most importantly, we should note that the conflict between those who see this as a genuine case of c- unification and those who do not may be neatly reflected in the formalisms developed above. To that extent it reflects well on the formalisms themselves. In particular, those who still insist that Renormalization Theory (R.T.) does not serve to reconcile Q.E.D. claims about infinite and indeterminate masses and charges (Q.E.D.1) with the evidence of electrostatics (E.E.) will presumably claim that P(Q.E.D.1/E.E. & (R.T.-Q.E.D.1&E.E.)) $\not\geq$ P(Q.E.D.1/E.E). That is to say, they will claim that the probability of the claims about infinite and indeterminate masses and charges are no more likely given the evidence of electrostatics and the new content added by Renormalization Theory then they are given simply the evidence of electrostatics alone. On the other hand those who do see Renormalization Theory as effecting the alleged unification will presumably claim that P(Q.E.D.1/E.E. & (R.T.-Q.E.D.1&E.E.)) $\geq$ P(Q.E.D.1/E.E). In other words, both the opponents and proponents of this alleged case of c-unification can accept those formalisms (notably D6, D7 and D8) however they will differ in the values given to the variables of the formalisms in this particular case. Indeed the use of those formalisms allows us to help pinpoint exactly where the disagreement occurs.

8. Conclusion

In looking for explanatory virtues rather than any fabled essence of explanation we open the path for serious discussion of wherein lie the merits and demerits of various putative explanations. If we are opponents of a particular alleged explanation it allows us to concede that the explanation has certain virtues while emphasizing the lack of other virtues we take to be highly important; we need not simply brand it a non-explanation. Perhaps in certain cases we will find opposing camps for and against a particular putative explanatory theory in complete agreement about exactly which explanatory virtues the theory has. They may even agree in their assessments of the degrees to which it has each such virtue. Then the disagreement may simply come to the question of how to weigh the importance of particular explanatory virtues. Here we may reach a genuine impasse.

But before we get there we need to get a clearer notion of exactly what are the explanatory virtues. R-unification and c-unification are prima facie estimable explanatory virtues even if the semi-formal accounts provided here prove to be flawed.

The point of seeking formal or semi-formal accounts of alleged explanatory virtues is that it gives us something we can sink our teeth into. Such accounts give a clearer indication of what is being talked about—we can test with some degree of precision through the use of examples and counter-examples. Further it helps us pin-point disagreements about particular alleged cases. For instance, if two disputants disagree over whether some theory T has the explanatory virtue of rendering genuinely incompatible claims C1 and C2 compatible, we can use D6 to see where their disagreement lies. Is it because they disagree about whether P(C1/C2) is greater than P(C1)? Or is it because they disagree about whether P(C1/C2&T) is greater than P(C1/C2)? Finally, it is nice to have formal accounts because they increase our ability to program our successors, be they humans or machines, with our own, undoubtedly excellent, tastes.[11]

Notes

[1]Suggested by Clark Glymour.

[2]Note, for ease of comprehension, D1 does not have the desirable property that the content part relationship is closed under logical equivalence. A more adequate, but more difficult to comprehend, version having this property runs as follows,

D1′ $\alpha < \beta =_{df}$ α and β are contingent, $\beta \vdash \alpha$, and for some ψ, ψ is logically equivalent to α and there is no σ such that $\beta \vdash \sigma$, σ is stronger than ψ and every atomic wff that occurs in σ occurs in ψ.

For a full specification of this new notion of content for a number of langauges both propositonal and quantificational see Gemes (1994).

[3]We presume here that the English sentence 'Ken and Alan are in Sydney' is identifiable as the conjunction of the two atomic sentences 'Ken is in Sydney' and 'Alan is in Sydney'.

[4]The relevant notion of analytic content is explicated in Gemes (1994a). Of course analytical entailment, unlike logical entailment, cannot be given a merely syntactical treatment. Hence the notion of analytic content is a frankly semantic notion. Note, to save Friedman's notion of k-atomicity we need recourse only to the notion of logical content; however, later we will need recourse a wider conception of content to explicate the relation of the surplus content of a statement in relation to other statements.

[5]Cf. Salmon (1989), pp. 96–99.

[6]This was brought to my attention by Jim Woodward.

[7]Suggested by Clark Glymour

[8]The notion of truth-functional independence is developed in Salmon (1969) and applies to statements p and q where "in a truth-table adequate to represent the truth patterns of the statements involved, the proportion of true cases for q, given that p is true, is equal to the proportion of true cases for q, given that p is false." Note, while the notion of truth functional independence invokes the notion of truth table and hence applies directly to propositional rather than quantificational statements, it can be applied to quantificational sentences through their propositional content parts. For instance, $(x)(Fx \supset Gx)$ is not part of the surplus content of $(x)(Hx \vee Fx \supset Gx)$ relative to $(x)Gx$ since it has a content part, namely $(Fa \supset Ga)$, that is not truth-functionally independent of Ga which itself is a content part of $(x)Gx$. Again, in applying this notion to ordinary English sentences we are assuming that an appropriate notion of atomic English sentence is available.

[9]Any awkwardness about the expression B & (C-A&B) felt on the grounds that (C-A&B) names a set rather than a statement can be overcome by letting (C-A&B) stand for the untested part of C relative to ⌜A&B⌝ which, as defined in Gemes (1994b), is itself a statement.

[10]While the following are all cases of putative scientific explanations it is worth noting that the virtue of providing something like c-unification is something much prized in philosophical as well as scientific explanations. To cite just one of many examples, Kant clearly prizes his posit of the distinction between the phenomenal and noumenal realms on the grounds that it allegedly reconciles the prima facie incompatible claims that nature is determined and that we have free will. This is a theme I hope to pursue elsewhere.

[11]The idea of applying my notion of content to Friedman's analysis of the notion of explanation as unification is due to Clark Glymour. Thanks our due to Jim Woodward for helpful discussions of Friedman (1974). Clark Glymour, Wes Salmon, Philip Kitcher and referees from this journal provided comments on previous drafts which greatly improved the end result. Ariel Fernandez provided invaluable suggestions for finding cases of c-unification in physics and David Albert helped give precision to some of the case studies. Lamentably, I must take responsibility for any remaining errors.

References

Breuer, J. & Freud, S. (1960) *Studies in Hysteria*, Vol 3. of *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (London: Hogarth Press, London, 1960).

Friedman, M. (1974) "Explanation and Scientific Understanding", *Journal of Philosophy* LXXI (1974), 5–19.

Gemes, K. (1994) "A New Theory of Content I: Basic Content", *Journal of Philosophical Logic*, 23 (1994).

_____. (1994a) "A New Theory of Content III: Semantic Content", in preparation.

_____. (1994b) "The Precise Formulation of Inductive Scepticism and some Pseudo-Refutations", in preparation.

Kitcher, P. (1976) "Explanation, Conjunction. and Unification," *Journal of Philosophy*, LXXIII (1976), 2077–212.

Kitcher, P. (1981) "Explanatory Unification," *Philosophy of Science* 48 (1981), 507–531.

Salmon, W. (1969) "Partial Entailment as a Basis for Inductive Logic", in *Essays on Honor of Carl G. Hempel*, ed. N. Rescher, D. Reidel, Dordrecht, 1969.

Salmon, W. (1989) *Four Decades of Scientific Explanation* (Minneapolis: University of Minnesota Press, 1989).