What Elements of Successful Scientific Theories Are the Correct Targets for "Selective" Scientific Realism?

Dean Peters*

Selective scientific realists disagree on which theoretical posits should be regarded as essential to the empirical success of a scientific theory. A satisfactory account of essentialness will show that the (approximate) truth of the selected posits adequately explains the success of the theory. Therefore, (*a*) the essential elements must be discernible prospectively; (*b*) there cannot be a priori criteria regarding which type of posit is essential; and (*c*) the overall success of a theory, or 'cluster' of propositions, not only individual derivations, should be explicable. Given these desiderata, I propose a "unification criterion" for identifying essential elements.

1. Introduction. In the face of the "pessimistic induction" (Laudan 1981), many proponents of scientific realism have embraced the idea of "selective realism." This states that not all the propositions of an empirically successful theory should be regarded as (approximately) true but only those elements that are essential for its success. It is, however, not obvious how a term like "essential" is to be understood. In this essay, after considering and rejecting several existing views, I will argue that the essential posits of a theory are those that unify the accurate empirical claims of that theory.

The basic argument for scientific realism is the "no-miracles argument" (NMA), versions of which have been presented, among others, by Putnam (1975a), Boyd (1980), and Musgrave (1988). The NMA is an example of inference to the best explanation and relies on the premise that the truth of a given theory's claims about unobservables is necessary to provide an adequate explanation of its success. In line with this argumentative strategy, all

Received August 2013; revised November 2013.

*To contact the author, please write to: University of Johannesburg, Corner Kingsway and University Road, Johannesburg, 2006, South Africa; e-mail: deandpeters@gmail.com.

Philosophy of Science, 81 (July 2014) pp. 377–397. 0031-8248/2014/8103-0003$10.00

the accounts surveyed below will be assessed by the criterion that the truth of the propositions they regard as essential should explain the success of the relevant theory. However, this criterion as stated is not by itself sufficient to determine which account is best, as there are various differing perspectives from which a theory's success might be explained.

There are at least three explanatory 'axes' upon which the various accounts of theoretical success might be located. The first might be labeled "retrospective-prospective." It distinguishes between accounts that attempt to explain the success of a theory from the perspective of a later theory and those that attempt to provide an explanation without introducing any scientific concepts not found in the earlier theory. The second axis might be labeled "special-neutral." It distinguishes between accounts that hold that there are a priori reasons to regard a particular type of theoretical element as having a special explanatory role and accounts that do not. The third axis might be labeled "particular-overall." It distinguishes between those accounts that focus on explaining particular empirical successes (i.e., the derivation of a particular result) from those that attempt to explain the overall success of a given theory or 'cluster' of theoretical propositions.

The unification account that I defend in this essay is prospectively applicable, is neutral regarding which type of theoretical element might have explanatory relevance, and seeks to account for the overall success of a set of theoretical propositions. That the correct account of selective realism is located in this section of explanatory 'space' is defended in sections 2–4, going through the axes in the order given above. The arguments given here will make reference to accounts of essentialness that have been proposed in the literature and that here serve as contrasting positions for the view defended. Following this discussion of contrasting positions, in section 5 I will give positive reasons for preferring the unification account.

2. Retrospective versus Prospective Accounts. Retrospective accounts of essentialness often depend on the "causal" account of reference, which was famously proposed by Putnam (1975b, 1978) and Kripke (1980). Under this account, the meaning of a term is identified with its actual referent, rather than the properties associated with the term by a theoretical description. The referent is whatever particular entity the people who coined the term were actually causally interacting with at the time they coined it. Importantly, the reference is fixed by this sort of causal association even if the association between those who initially use the term and the corresponding referent is mediated by a substantially false description.

The referential stability supposedly secured by the causal account is the basis for a type of selective realism. For instance, Hardin and Rosenberg (1982) argue that the posit of a luminiferous ether in Fresnel's wave theory of light "referred to the electromagnetic field all along" since the ether played the causal role ascribed to the electromagnetic field in modern physics. Regarding the ether as an “essential posit” of Fresnel’s theory thus explains the empirical success of the theory from the perspective of modern electromagnetic theory.

Laudan (1984), however, notes that under such an account practically any two theories that attempt to explain a common set of phenomena will be taken as referring to the same set of entities. For example, Aristotle’s notion of “natural place,” Descartes’s particle “vortices,” and Newton’s “action at a distance” all play the causal role of accounting for gravitational action. So, picking the “essential” elements of theories by reference to the causal roles specified by current theories has the result that virtually any past theory would be regarded as accurately describing the essential underlying elements. But many past theories have manifestly not been empirically successful. Thus, the account fails to distinguish successful from nonsuccessful theories and so is hardly adequate as an explanation of success.

For many selective realists who are sympathetic to retrospective accounts of essentialness, this conclusion is, in fact, not catastrophic, because they typically argue that both the causal role and some of the key descriptions associated with a term must be held constant for the reference to remain stable (see, e.g., Kitcher 1993; Psillos 1994). However, because the causal component of such “causal-descriptivist” accounts is so indiscriminate, it effectively does no work in distinguishing essential from inessential posits. Judgments of essentialness must depend on the role a posit plays within a theory, with judgments about reference (if these are required at all) following epiphenomenally in their wake.

Stanford (2006, chap. 6) has offered a specific refutation of the Hardin and Rosenberg application of the causal theory of reference that he calls the “trust argument.” However, he also gives an argument that would seem to refute more descriptivist accounts. He claims that the dependence on hindsight in picking out the essential elements of theories under such accounts results in problematic circular justifications: “One and the same present theory is used both as the standard to which components of a past theory must correspond in order to be judged true and to decide which of that theory’s features or components enabled it to be successful. With this strategy of analysis, an impressive retrospective convergence between judgments of the sources of a past theory’s success and the things it ‘got right’ about the world is virtually guaranteed” (166).

It is, indeed, frequently possible to explain the success of a past theory by taking as approximately true those theoretical posits that do happen to correspond to the posits of the later theory. But if convergence between old and new theory in fact is guaranteed, this cannot support realism. The convergence is adequately explained by the fact of retrospective selection itself, with no need to appeal to the truth of the theoretical claims in question.

It is, however, not true that such convergence is "guaranteed." Post (1971) famously argued that it is a requirement on later theories that they explain the successes of past theories and that otherwise successful theories might fail in this requirement. Even in cases where there is theoretical continuity, the theoretical elements that are retained in the newer theory might fail to explain the success of the older theory. Nevertheless, it must be conceded that it will generally be far easier to construct a satisfactory explanation post hoc, whatever particular account of explanation is favored. Prospective selective realism avoids the possibility that hindsight has been used in picking out particular elements as essential, which renders the realist claim more credible if these elements in fact are retained (even if only approximately) in successor theories. Thus, although a retrospective account might represent a 'fallback position' for the selective realist, it is extremely undesirable relative to prospective accounts.

3. Special versus Neutral Accounts. In subsections 3.1 and 3.2, respectively, I focus on two particular "special" accounts of essentialness that have been prominent in the literature. The first is Worrall's mathematical structural realism, as presented in his (1989b) paper on the topic. The second is entity realism and Cartwright's development of this idea into "phenomenological realism." In the third subsection, I argue against special accounts in general, by reference to these particular examples.

3.1. Mathematical Structural Realism. Part of the influence of Worrall's account is no doubt due to his development of Fresnel's wave theory of light as a case study for selective realism. On Worrall's (1989a, 1989b) interpretation, Fresnel's achievements were twofold. First, he produced a mathematically rigorous wave theory of light that yielded quantitative predictions for diffraction phenomena, including hitherto unobserved phenomena such as the famous "white spot." Second, along with Arago, Fresnel suggested that light was specifically a transverse wave, accounting for the selective transmission of light through polarizing materials. This suggestion also resulted in the so-called Fresnel equations, which accurately predict the amount of light that will be refracted or reflected at the interface of two media.

This episode presents a prima facie problem for the realist because the propagation of a wave appeared to require the existence of some medium that permeates space—the "luminiferous ether." Positing the existence of this substance, however, turned out to be quite incorrect from the perspective of later science; in the mature formulation of Maxwell's theory, light is the manifestation of a "freestanding" electromagnetic wave. The now-standard selective realist response to this case, as established by Worrall, is to attempt to demonstrate that the appeal to a mechanical ether in fact was inessential to the theory's success but that the abstract wave equations (which were largely preserved in later theories) were essential. Worrall generalizes this outcome to argue that we should not be realists about the intrinsic character of the entities postulated by any given theory but rather about the form of the relations between these entities, as expressed by the mathematical formalism of the theory.

3.2. Entity Realism and Phenomenological Realism. The broad idea behind entity realism is that a realist attitude is warranted in respect of the entities proposed by successful scientific theories but not necessarily in respect of the properties attributed to them. Hacking's (1983) criterion for inferring the existence of entities—experimental manipulability—is an adaptation of the causal theory of reference above. However, this criterion tacitly commits Hacking to more than just a naked assertion of the entities' existence, specifically the truth of at least those theoretical claims that describe how these entities will behave within a certain type of experimental setup. This point is preemptively conceded by Hacking, when he claims that we presuppose knowledge of "a modest number of home truths about electrons" (265) when we design scientific apparatus relying upon manipulation of these entities.

Cartwright's account has a major advantage over Hacking's in that it gives some specific rationale for accepting, and means for picking out, these "home truths." She gives three arguments for adopting a realist attitude only toward claims about the existence of entities and a minimal set of their properties. I shall focus only on the third, which purports to pick out which theoretical posits "do the work" of deriving empirical predictions. This argument states that empirically successful phenomenological models or laws are not deductively derived from high-level theory but are "built upwards" by attempting to fit the phenomena directly (1983, chap. 8).

A major example used by Cartwright and her coworkers to illustrate this claim is the London model of superconductivity (Cartwright, Shomar, and Suarez 1995). They argue that there was no preexisting rationale for this model in the theory of electromagnetism. Rather, the model was constructed specifically to accommodate the observed phenomenon that materials rapidly expel magnetic fields as they undergo the transition to superconductivity. Abstract theoretical propositions are involved in this process but merely as heuristics that "guide" the formation of the phenomenological model. And such a tenuous connection to empirical success is not sufficient to warrant a realist attitude in respect of these propositions.

Cartwright (2009) has more recently described her views under the title of "phenomenological realism." Specifically, she adopts a realist attitude toward the descriptions of phenomenological laws and models found in theories and toward the entities and basic causal powers required for these laws and models to be true. One thing that must be conceded in favor of this view is that more ambitious forms of scientific realism are committed to at least phenomenological realism. Phenomenological models are more directly related to empirical findings than abstract theoretical laws. As such, if any part of a scientific theory can be inferred to accurately represent the world under explanationist arguments like the NMA, then surely phenomenological models are first in line.

3.3. Special versus Neutral. The first major objection to special accounts is that they are motivated by particular cases or types of cases that are not necessarily representative. We might concede that the mathematical structural realist and the phenomenological realist have successfully identified those theoretical elements that are intuitively essential to the Fresnel wave theory of light and the London model of superconductivity, respectively. And yet we can, and should, deny that similar analyses can be applied more generally.

Applying this criticism to phenomenological realism, we start by noting that fundamental laws could be related to phenomenological models in various different ways. At one end of the spectrum, a model that is deductively derived from the fundamental law (given certain 'natural' boundary conditions, auxiliary assumptions, etc.) has the correct form to serve as a phenomenological model without further modification. Empirical measurements are required only to fix the values of free parameters left open in the theoretical derivation. In the middle of the spectrum, both the form of the model and the value of the parameters are derived from the phenomena. Fundamental laws merely provide some helpful heuristics that guide this process. At the far end of the spectrum, the phenomena are sufficiently simple in structure that an appropriate model can be read directly from them. High-level theory, if it is involved at all, may serve merely to provide a post hoc 'explanation' of the observed regularity.

Cartwright's phenomenological realism is based on cases, including the London model, that happen to be in the middle of this spectrum. But there are also cases where models derived relatively directly from theory are considerably more successful. The Fresnel case in fact is one such example—part of the appeal of the theory is that it results in predictive phenomenological models (such as that for the white spot) derived from first principles. A satisfactory account of essentialness should be able to capture (at least) both types of case.

In the case of structural realism, there are intuitively convincing counterexamples to the general view that the essential parts of scientific theories in all cases can be identified with mathematical equations. In particular, such a view would seem to rule out any selective realism about theories that are not in any substantial way expressed mathematically (Gower 2000, 74;

Newman 2005, 1377–78). This seems difficult to square with the undeniable predictive successes of, for instance, theories in modern molecular biology that rely on lock-and-key models of molecular interaction.

Perhaps these putative counterexamples could be disregarded if there were some general principles that could be articulated in favor of one or other special account of essentialness. But the second major objection to such accounts is that the criteria that are actually used to argue for the specialness of one type of theoretical posit in fact could be applied more generally.

For instance, Cartwright (1983) claims that entity realism (and thence phenomenological realism) is warranted by "inference to the most probable cause." She directs this form of reasoning only at low-level causes. The successful manipulation of a track in a cloud chamber, for example, licenses an inference to the existence of the particular particle that causes the track. However, she explicitly states that high-level theoretical posits tell us about general causal powers. So it is unclear why inference to the most probable cause should not license realist commitment to the higher-level causes—such as electrostatic forces—implied by these posits. Such an inference might not give us reason to believe the truth of abstract theoretical laws, along traditional realist lines, but rather give us reason to believe in the existence of real categories of causal interaction 'associated with' these laws. But that may still be enough for a selective realism more expansive than that allowed by Cartwright.

Psillos (1995, 1999) makes similar criticisms of mathematical structural realism. Worrall seeks to draw a distinction between formulating the equations describing the behavior of an entity and coming to an understanding of what the entity is. But Psillos argues that to give a structural description of an entity is just the same as providing information about the "nature" of that entity. For instance, Psillos argues, "mass" in Newtonian physics is understood as just that entity that, among other things, obeys the second law of motion. This point in fact is substantially conceded by Worrall in later presentations of structural realism, which are described briefly in section 5.1. Moreover, Psillos points out, it is not only equations that are preserved across theory change. In the transition from Fresnel's theory to Maxwell's, there is also continuity in various "theoretical principles" such as the conservation of energy and the finiteness of the velocity of light waves. So it is not clear what is special about mathematical equations. Surely, Psillos suggests, it makes more sense to argue simply that Fresnel was right about some theoretical claims, and wrong about others, without committing in advance to the former being expressed as mathematical equations.

So special accounts of essentialness tend to be motivated by examples where the 'type' of theoretical posit in question is intuitively essential to empirical success. There is nothing wrong with using such examples to illustrate a broader account, but the proffered special accounts do not stand up on their own terms. For one, there are various intuitive counterexamples to these accounts. And the principles articulated to support them, when applied consistently, have more “neutral” results than advertised.

The points made above largely serve to undermine the existing arguments for particular special accounts of essentialness. It may yet be the case that a special account not examined here will avoid similar criticisms. While I concede that this is possible, it seems implausible. Given the enormous diversity of theoretical architecture in science, it would in fact be prima facie surprising if a particular ‘type’ of posit turned out to be essential in all cases of empirical success. In order to achieve full generality, an account of essentialness should take a rather more abstract, logical view of the role that theoretical posits play in a theory and how they relate to one another. These prima facie reasons to be skeptical of special accounts will be further supported by the attractiveness of the particular neutral view offered in section 5.

4. Particular versus Overall Success

4.1. The Working Posits Idea. One extremely intuitive variation of selective realism claims that the essential posits of a theory are just those that are actually involved in the derivation of particular successful predictions. Psillos (1999) has developed this idea in depth under the heading of the *divide et impera* strategy. However, following Kitcher (1993), I will refer to it as the “working posits” idea. An appealing feature of this idea, in light of the previous section, is that it is neutral regarding what type of theoretical posit might turn out to be essential. The fundamental problem with the idea is that the notion of which posits are “actually involved” in a given derivation is open to interpretation. This is especially so because a large number of theoretical posits are typically mentioned by scientists when they make a derivation, often including many, such as the luminiferous ether, that the selective realist would like to exclude.

Psillos proposes two quite different ideas for distinguishing those posits that are actually essential from those that are merely mentioned. I focus only on the second, which is that the essential elements of a theory are not those that are merely used in successful derivations but those that are in some sense required. Psillos (1999, 105) expresses this by saying that a hypothesis H is essential just in case it cannot be eliminated or replaced by another “available” posit in such a way that the successful derivation remains intact. However, as pointed out by Lyons (2006) and Vickers (2013), much of the weight of this suggestion is borne by the vague term “available.”

One interpretation of this term is that any posit that is logically derivable from H should be counted as available. This is a natural thought given the various examples we have seen. The "function" of the luminiferous ether posit in Fresnel's theory of light is to provide a material basis for waves that propagate through space, so why not eliminate the abstract posit and accept the wave posit in its own right? The problem with this thought is this: if it is legitimate to accept an intermediate result and regard as inessential the assumptions used to derive it, then very little actually is essential. The process of elimination can be continued until we take the empirical results themselves as assumed and regard all other theoretical posits leading up to them as inessential.

Psillos (1999) argues that this process of elimination should not result in a set of posits that cannot explain the empirical results in question. But this just presupposes some substantial account of explanation, which Psillos does not offer. In fact, all variations of the working posits idea seem caught between two models of explanation, either of which leads to highly counterintuitive consequences if applied consistently.

If we accept a simplistic deductive-nomological model of explanation, the truth of a particular subset of posits explains a theory's empirical success simply because it deductively entails these results. But it is a problem with the deductive-nomological model in general, and thus with the proposed variant of selective realism, that universal claims that subsume the explanandum can nevertheless be intuitively unsatisfactory as explanations. While the blackness of a particular raven can be explained by reference to the proposition that all ravens are black, it seems like a better explanation would give some more abstract explanation for why this generalization holds true. So this model of explanation results in intuitively trivial or underinclusive explanations.

The other model of explanation picks out those posits "actually used" in producing a derivation. Given this model, it is natural to adopt a causal or historical perspective and so ask about the actual chain of reasoning that led to a result. However, this perspective is anathema to the selective realist; applied consistently, it will tend to characterize as essential all the posits that theorists actually accept. Moreover, although some critics of selective realism can be read as adopting this perspective, the antirealist must shrink from its most extreme consequences. For instance, before proceeding down a successful theoretical path, it is almost inevitable that a scientist will explore several blind alleys. Thus, posits that the scientist herself considers to be false, and do not appear in the final theory, can nevertheless be causally implicated in its ultimate success. So this model results in intuitively overinclusive explanations.

To avoid the Charybdis of underinclusion and the Scylla of overinclusion, we need some more sophisticated account of what makes for a good explanation of a given empirical success. This goal is partly addressed in the following subsection.

4.2. Essentially Contributing Parts of Derivation-Internal Posits. In a recent paper in this journal, Vickers (2013) has provided a sophisticated elaboration of the working posits idea that attempts to avoid both Charybdis and Scylla by means of three new proposals.

First, to avoid an overinclusive commitment, one must distinguish between posits that are "used" in some general sense and those that are linked to empirical success in the right way. That is, one must drive a wedge between historical/causal and epistemic/logical accounts of essentialness. To this end, Vickers suggests a distinction between "derivation-internal posits" (DIPs) and "derivation-external posits" (DEPs). The DIPs are those that give rise to an impressive empirical result by a deductive, truth-preserving inference, whereas the DEPs are those that merely "influence" the derivation of a result.

Vickers' second proposal emerges from the recognition that a posit that is deductively linked to an empirical result may be logically stronger than is required to achieve that result. He therefore argues that we should apply an eliminability criterion at the local level, taking as essential only the "essentially contributing parts" (ECPs) of the DIPs, where the ECP is the minimal logical consequence of the DIP that has the same logical consequences as the DIP within the context of the particular derivation in question.

As he has endorsed an eliminability criterion, Vickers must articulate some principled reason to avoid regarding almost all posits as eliminable. This brings us to his third proposal. In explaining this proposal, it is useful to think of a derivation as a "river" of logical connections, with various "upstream" DIPs combining (through logical entailment) to result in the derivation of "downstream" posits and eventually an empirical result. Vickers's proposal is that the logical structure of the "confluence points" where two or more posits combine to yield a "downstream" proposition must be preserved. Call the DIPs that so combine CDIPs (for "confluence DIPs"). One might still want to preserve only the ECPs of these CDIPs, but we cannot eliminate them altogether. This proposal not only averts the disastrous consequence of a pure logical eliminability criterion but is also intuitively appealing. The places where several posits combine to yield a new proposition are intuitively where the logical "work" is occurring in a derivation.

Vickers's account stands as a plausible candidate for selective realism. Recall that the aim of this enterprise is to explain why a theory is able to predict some or other empirical result. And Vickers's three criteria together provide a recipe for picking out a minimal set of posits that deductively entails and therefore, if true, explains a given result. However, there remain some problems with this account.

First, one might question the tenability of the distinction between DIPs and DEPs. As pointed out by Duhem, empirical results do not usually follow deductively from individual theoretical propositions. Derivations require the postulation of additional assumptions, boundary conditions, and so forth. So whether a given posit is related by deductive entailment to any empirical results depends on what other assumptions are included. Thus, indefinitely many formal derivations of a given result are possible. Which derivation actually occurs may depend on various factors, including pure happenstance and how much 'physical intuition' the scientist concerned has in respect of the problem. Although Vickers (2013, 201) acknowledges the concern, he does not attempt to offer a resolution to it. Thus his distinction, as outlined above, is too much a hostage to the contingencies of the scientific process to bear the explanatory weight that he would place upon it.

This leads to a second criticism of Vickers's account, similar to that leveled at "special" accounts of selective realism—namely, that it fails to apply its own (implicit) standards of explanation consistently. Vickers's stated aim is to provide some criterion for identifying as essential "at least *some* posits in at least *some* cases" (personal communication). But the selective realist should accept as (approximately) true whichever posits figure in the best explanation for a theory's empirical success, not just "some" of these posits. It is plausible that, all else being equal, the truth of a particular theoretical posit is a better explanation of the success of an empirical prediction if there is a (relatively "natural") deductive relationship between the two than if there is a weaker logical relationship. But it could still be that the truth of a posit more weakly connected to empirical results explains the success of the theory sufficiently well that this explanation is an instance of the NMA.

A final problem with Vickers's account concerns the overall thesis of this section. Vickers follows Kitcher and Psillos in attempting to explain the success of particular derivations, and his account is the most successful attempt yet to illustrate how this might be done. Yet the goal itself is faulty; the goal of the selective realist should be to explain the overall empirical success of a theory or, at least, a 'cluster' of theoretical propositions. One argument for this is that a strategy focused on the explanation of particular successes will lead to very counterintuitive conclusions. For instance, suppose there is a theory that has made various successful predictions by means of different individual derivations. Now suppose a specific abstract posit is found in all of these derivations. It is possible, under Vickers's account, that this posit could be eliminated without altering the branching structure of the argument within the scope of any single derivation. Vickers would therefore have to regard this posit as inessential despite its contributing to many successful derivations.

As an example of such an abstract central posit, consider the hypothesis of the 'unified' atom. This hypothesis states that the world is largely composed of tiny atoms, each of which embodies several properties: atoms are resistant to permanent chemical and mechanical decomposition but can be broken down by nuclear forces; atoms can act as mechanically discrete objects; atoms combine with each other in fixed proportions to yield chemical compounds; atoms can gain or lose charge and thence be acted on by electrical and magnetic fields; and so on. For a given empirical result that appeals to 'the atomic hypothesis', generally only a subset of these properties in fact are called upon. The existence of Brownian motion, for example, is derivable from the existence of the mechanical atom (or molecule, to be more exact), and this derivation need make no reference to any chemical properties of atoms. The derivation of ratios of chemical combination, on the other hand, requires only the 'chemical atom', and so on.

From the perspective of each derivation, the posit that there exists a unified atom is an isolated logical "tributary," the truth of which explains very little. But from a broader perspective, its truth explains a great deal. Accepting that it is true still allows us to explain particular successful derivations, but failing to accept it means that we are unable to explain the success of the broader theory. Moreover, for the reasons cited above, the explanations of particular derivations premised on the truth of this more basic premise are more satisfactory than those that do not. As illustrated by this example, and using Vickers's metaphor of scientific reasoning as a river system, our criterion of essentialness would be more explanatory if it focused on logical "divergence points," as opposed to convergence points. And these considerations, I take it, count against not only Vickers's account but any account that emphasizes explaining individual successes in derivation at the expense of explaining broader patterns of empirical success.

I have made three criticisms of Vickers's proposal from the perspective of explanation. Criticisms one and two emerge from the fact that it is nowhere stated exactly what kind of explanation is sought in giving an account of essentialness. It is therefore unclear how the differential treatment of DIPs and DEPs is to be justified. One can therefore, as I have done in the second criticism, argue that the explanatory virtues that Vickers claims reside in his account are even more amply fulfilled by a less minimalist account. The third criticism is that his explanandum—namely, the fact of individual empirical successes, is simply the wrong target from the realist's perspective. We aim to explain empirical success in science as a general phenomenon. We are thus led to explain the success of particular theories, understood as complex networks of propositions that scientists often accept or reject (although frequently with caveats) as a "package." By focusing on individual derivations only, Vickers makes a mystery of the fact that such larger networks sometimes generate numerous empirical successes.

Of course, Vickers is quite right to worry at the prospect of spreading our explanatory net too widely and so accepting the truth of propositions that in fact are relatively likely to be false. A related worry, raised by an anonymous reviewer for this journal, is that the boundaries between "theories" may be altogether indistinct—several sets of hypotheses that are conventionally understood to comprise distinct theories, such as Newtonian mechanics and Coulomb's force law, might both be implicated in a given set of successful derivations. If our goal is to explain the success of "theories," we must give some account of how such entities are to be delineated.

The positive account of essentialness, the unification criterion, which I defend below, while it does not delineate "theories" as such, does pick out those elements of a network of propositions that warrant (approximate) realist commitment from a given base of empirical results. I suspect that, when applied to a given body of evidence, such a criterion will largely replicate our intuitive boundaries between theories by picking out "clusters" of propositions whose truth would explain particular sets of empirical results. However, when considering all the empirical successes of science as a whole, it may well turn out that such boundaries dissolve. Such an outcome is already visible to some extent in the case of the atomic hypothesis discussed above, where the explanatory appeal of the hypothesis cuts across the boundaries of entire disciplines, not merely theories. This dissolution of theories as objects of realist commitment in favor of widely distributed individual posits is not, I think, something to be regretted. In fact, it seems like a corollary to the basic rejection by selective realism of the view that theories should be treated as holistic units for the purpose of realist commitment.

5. A Positive Account

5.1. The Unification Criterion. In outlining my positive view of essentialness, it will be necessary to return briefly to first principles. Specifically, we might ask, what standard of empirical success must a theory achieve such that we would regard it as a "miracle" if it later turned out to be entirely false? In answering this question, many contemporary advocates of scientific realism focus on the value of "novel prediction." However, several authors (and here I should acknowledge my debt to Worrall [2000, 2002, 2006, 2010]) argue that "novelty" should not be understood in the most literal sense, as in predicting some phenomenon that has not yet been observed but rather in the sense of "use-novelty." A theory makes use-novel predictions just in case it entails the existence of empirical data that were not used in the construction of the theory. What is intuitively appealing about a theory satisfying this criterion is that we 'get out more than we put in'—it entails more empirical content than was used in constructing it.[1] Or, to put it another way, a successful theory is

1. As pointed out by an anonymous reviewer for this journal, the notion of quantifying "empirical content" presents serious technical difficulties, as a universally quantified claim may imply an infinite number of particular empirical observations. A fully satisfactory account of unification would require some principled way of treating this entire

one that is not ad hoc, where an ad hoc theory has had the required empirical results "put in by hand."

The intuitive appeal of the use-novelty criterion is more easily discerned if we notice that it is, in effect, a unification criterion. A unification criterion holds that a theory is empirically successful, and thus confirmed, to the extent that it predicts or explains a large number of empirical results by appealing to only a relatively small number of basic principles. However, basic principles are generally not arrived at a priori—they are usually postulated by an inductive 'leap' that generalizes from some base of empirical observations. Thus a theory cannot be counted as unifying if it is ad hoc, that is, if the empirical base used to derive the theoretical principles is just as large as the set of empirical results then derived from those principles. The use-novelty criterion can equally be thought of as cashing out the unification account of confirmation, not merely refining the notion of novel prediction.

Although this is not the occasion to argue in depth for this account of confirmation, it is possible to adduce a few additional intuitive arguments in its favor. First, although we have all become inured to this phenomenon, it is worth being reminded how remarkable it is that certain theories are able to give us more information about the world than we 'put into' them. A successful theory 'amplifies' our knowledge, and the fact that such a thing is possible is what leads us to conclude that the world embodies a regular structure and moreover that such theories accurately describe this structure. Second, we are not usually particularly impressed by theories that derive modest predictions from a large base of empirical observation. One need not be a talented scientist to predict the arrival time of the next train, if one takes the same train every day. We are, however, impressed by predictive (or explanatory) leaps that take us far beyond the information already embodied in the theory.

The positive account of essentialness defended in this section is called the "empirically successful subtheory account" (ESSA). The central idea of the ESSA is that the unification criterion of empirical success in fact does not justify regarding entire theories as confirmed but rather provides a means for picking out the confirmed elements within theories. Let us assume that a scientific theory consists of a set of propositions and their deductive closure. Given this picture, it is always possible to pick out a 'subtheory' consisting of a subset of these propositions. So, to pick out the essential elements of the theory under the ESSA, start with a subtheory consisting of statements of its most basic confirmed empirical consequences

set as a "unit" of information for the purposes of assessing use-novelty. For reasons of brevity, a rough-and-ready intuitive notion of empirical content will have to suffice for now.

or perhaps its confirmed phenomenological laws. These, after all, are the parts of a theory that even empiricists agree we should be "realists" about. Further propositions are added to this subtheory by a recursive procedure. Consider any theoretical posit not in the subtheory. If it entails more propositions in the subtheory than are required to construct it, tag it as confirmed under the unification criterion, and so add it to the subtheory. Otherwise, leave it out. When there are no more theoretical posits to consider in this way, the subtheory contains the essential elements of the original theory.

This procedure can be illustrated by Fresnel's theory, a simplified version of which is depicted graphically in figure 1. The initial subtheory simply states all the particular observational consequences of the theory that happen to be confirmed. This corresponds to all the boxes on the lowest level of the diagram, minus predictions of the speed of light varying depending on the orientation of the measuring device (labeled "speed of light"). Now consider the propositions, on the next level up, which express phenomenological laws or models of the type discussed by Cartwright and colleagues. Since "polarisation" and the "Fresnel equations" both entail more confirmed empirical content than that required to construct them, these propositions are added to the subtheory. In the next round, the proposition that light consists of "transverse waves" is added, since this successfully unifies lower-level posits. As expected, however, the posit of a luminiferous ether is not added. The only reason one might have for introducing this posit is

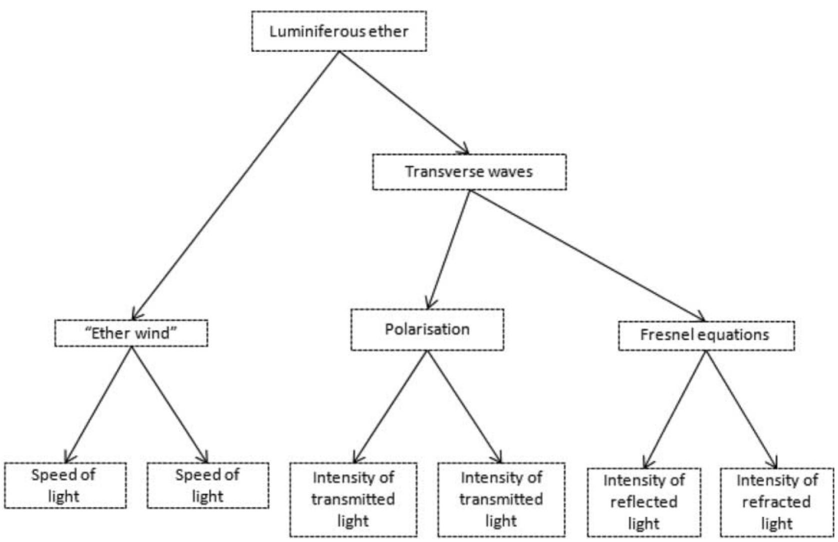

Figure 1. Some elements of Fresnel's theory of light. Posits are represented by boxes and deductive relations by arrows between them.

to account for the fact that light obeys wave equations. But the posit does not entail any additional verified content from the existing subtheory. It therefore does not satisfy the requirement that we "get out more than we put in."

This account of essentialness is appealing for several reasons. It is applicable prospectively, it is a priori neutral about which sorts of theoretical posits are essential, and it is concerned with explaining the overall empirical success of a cluster of theoretical propositions. Moreover, although I have articulated it here using full-blooded realist terminology—that is, in terms of "accepting theoretical posits"—it is also compatible with a more "neutral" variant of structuralism than that criticized in section 3. This variant can be described as the label "Ramsey-sentence realism" and has been endorsed by Cruse and Papineau (2002; Cruse 2005; Papineau 2010) and Worrall himself (2007; Worrall and Zahar 2001). Once a theory is reduced to its Ramsey sentence, basically the same procedure as that outlined above could be applied to pick out those theoretical relations and (metaphysically unspecified) entities that serve to unify confirmed empirical results in the right way. The unification model can, in other words, be thought of as providing the required account of "essential structure."

Finally, the unification criterion is, in its own right, an intuitively appealing account of what hypotheses we ought to believe, given a certain body of evidence. Many authors have emphasized the special epistemic status of unifying hypotheses in other contexts. For instance, in his theory of "explanatory unification," Kitcher (1981) has argued that a hypothesis is explanatory just in case it sets forth a general "argument pattern" that can be used in the derivation of many specific beliefs. Kitcher's theory of unification differs from the ESSA in that the latter emphasizes particular theoretical posits as opposed to more richly specified argument patterns. However, once the role of auxiliary hypotheses is taken into account (see the following subsection), this difference may narrow. More substantively, the ESSA is an account of confirmation, rather than explanation. As such, it explicitly considers the amount of empirical content required to construct a given hypothesis, not only the amount it successfully accounts for.

Thus, the unification criterion, rather pleasingly for the realist, picks out those posits that are confirmed by the empirical evidence—that is, that we have good reason for believing! The major precedent for this type of reasoning in philosophy of science is given by Whewell, in a form of inference he names the "consilience of inductions": "We have here spoken of the prediction of facts *of the same kind* as those from which our rule was collected. But the evidence in favour of our induction is of a much higher and more forcible character when it enables us to *explain* and determine cases of a *kind different* from those which were contemplated in the formation of our hypothesis. . . . Accordingly the cases in which inductions from classes of facts altogether different have thus *jumped together*, belong only to the best established theories which the history of science contains" (1858/1968, 153).

Despite this venerable pedigree, there are nevertheless various objections that might be raised against the ESSA. The following subsection will be concerned with responding to these before I conclude.

5.2. Clarifications and Responses to Objections. One possible complaint against the ESSA is that it is "riskier" than more modest forms of selective realism, such as Vickers's account or phenomenological realism. Because it, at least potentially, endorses realist commitment to a greater number of theoretical posits, it is more likely to be refuted by Laudan-type examples from the history of science. But this is, in fact, hardly a point *against* the ESSA but rather one in its favor. Like the scientist, it is arguable that the philosopher of science should adopt the Popperian advice to favor theories that "stick their neck out." So, provided there are some positive reasons to accept it, the case can be made that a more falsifiable theory is to be preferred over rivals, at least up until the point it is actually falsified!

The ESSA does not make a priori judgments about which 'levels' of a theory or 'types' of theoretical elements are essential. So, in different cases, the algorithm sketched out above will have very different results. One might nevertheless think of it as providing a useful "error theory" of the special accounts that have been proposed. There are indeed cases in which mathematical equations will be the most abstract theoretical posits regarded as essential under this account. And there are also cases in which the recursive procedure outlined above proceeds no further than phenomenological laws or models. But, in other cases, highly abstract 'metaphysical' claims may turn out to be essential under the ESSA. For instance, in the absence of observations of a differing observed speed of light depending on the observer's direction of motion, the posit of a luminiferous ether does not contribute any additional unifying power to the theory of light. However, if counterfactually such observations had been made, the ESSA would demand a realist attitude toward this posit.

Perhaps this last point, that the ESSA will occasionally recommend realism in respect of highly abstract posits, could be wielded as a complaint against it. It could be argued that this fails to reflect the humility that is required for any plausible realist position, especially in light of the pessimistic induction. This is a reasonable concern but is mitigated by two factors. First, if we think of the ESSA as imposing additional criteria on a basically structural understanding of theories, then these commitments appear considerably less worrisome. To accept that the posit of a luminiferous ether is essential (although this is not the judgment of the ESSA in the actual case) is simply to accept that there exists some entity that plays the specified role in unifying various lower-level phenomena. Second, recall that such relations are not accepted simply because they are posited as part of successful theories but because in these cases the particular logical structure of the theory gives us good reason to think that some-or-other factor plays the specified role.

Both Vickers and Cartwright argue that realist commitment should be withheld in cases where high-level theoretical posits merely "guide" or "influence" the formation of a model. It has been conceded that a relationship between a prospective unifying hypothesis and lower-level posits that requires relatively few (and "natural") auxiliary assumptions to satisfy the conditions of deductive entailment does give, all else equal, better reason to accept the prospective unifier as essential. But, since the 'strength' of a logical connection between two propositions is a matter of degree rather than kind, other factors may also be relevant in judgments of essentialness. Specifically, a posit that is more 'weakly' logically related to confirmed empirical results may nevertheless be counted as essential provided that there are enough of these results and they are sufficiently diverse. This is not, of course, to claim that any "unifying" hypothesis ought to be regarded as probably true—the connection between hypothesis and the various empirical results may nevertheless be so tenuous and post hoc that the truth of the hypothesis does not add anything to the explanation of why these results were achieved. Moreover, since both the 'strength' of logical connection and 'breadth' of influence vary on a gradient, there will always be cases where it is unclear to what degree a realist attitude is warranted. Nevertheless, contra Cartwright and Vickers, there will also be cases where a realist attitude is warranted toward a hypothesis that merely "inspires" empirical results simply because it has this relationship to many and varied such results.

Although this Duhemian criticism of Cartwright and Vickers is telling, there remains the suspicion that the ESSA will also encounter problems from this quarter. The algorithm outlined in the previous section presupposes that a theory consists of a set of propositions arranged in relations of deductive entailment. But, as pointed out repeatedly above, these relations will almost always be either more tenuous than this or be rendered deductive only by the addition of various auxiliary assumptions. A given theoretical posit therefore cannot be said to unify anything when taken by itself. If a realist commitment is to be sustained by the unification criterion, it seems it must therefore be directed at not only the 'headline' theoretical proposition but also at the various auxiliary assumptions. Perhaps this bullet can be bitten with gusto—why should we not, after all, be realists about auxiliary assumptions? These may too (albeit in approximate and/or structural form) tell us truths about the world and be preserved in successor theories. This concession does, however, complicate the project of identifying hypotheses that satisfy the 'more out than you put in' criterion. For now it seems we must evaluate entire clusters of hypotheses by this criterion, assessing how much empirical information is used in formulating boundary conditions and so on. However, while this may represent a substantial technical challenge, it does not seem to point toward any fundamental problem with the project.

A more fundamental worry is that which propositions are confirmed by the unification criterion will be relativized to the axiomatization of the theory. A proposition in the 'natural' axiomatization of the theory that acts to unify various empirical results, call it P, might, for instance, be replaced by Q and Q → P, in which case Q must be counted as part of the cluster of propositions that unify these same results. And thus any arbitrary proposition might be taken as an essential part of any give theory. Of course, this and similar problems affect all accounts of confirmation proposed so far, including the hypothetico-deductive model (Hempel 1945), Bayesianism (Earman 1992), and so on. One thing that can be said in favor of the unification criterion is that it is not obviously more affected by these problems than these other accounts of confirmation. Moreover, the worry that we lack a principled reason to prefer our natural axiomitization over possible alternatives seems ultimately to stand or fall alongside other skeptical worries about the justifiability of induction, the use of 'natural' predicates (Goodman 1983), and so on. In each case, we cannot fend off the committed skeptic but merely make the limited, Humean claim for the defensibility of our common practices.

6. Conclusion. In this essay, I have provided a general account of which posits are essential to the empirical success of a scientific theory and thus are the proper subjects of a realist attitude under the no-miracles argument. I have provided three conceptual 'dimensions' for evaluating accounts of essentialness and, using this framework, have argued that all existing accounts are deficient in one or other respect. Any satisfactory account of essentialness should be (1) applicable prospectively; (2) neutral as to what 'type' of theoretical posit is counted as essential; and (3) able to explain the overall empirical success of a given scientific theory or group of theories, as opposed to only explaining particular derivations. The empirically successful subtheory account satisfies these desiderata and moreover is intuitively appealing as an account of which theoretical posits we have good evidence for believing.

REFERENCES

Boyd, R. 1980. "Scientific Realism and Naturalistic Epistemology." In *PSA 1980: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 1, ed. Peter D. Asquith and Thomas Nickles, 613–62. East Lansing, MI: Philosophy of Science Association.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Oxford University Press.

———. 2009. "Entity Realism versus Phenomenological Realism versus High Theory Realism." Unpublished manuscript, London School of Economics Scientific Realism Revisited Conference, April.

Cartwright, N., T. Shomar, and M. Suarez. 1995. "The Tool Box of Science: Tools for the Building of Models with a Superconductivity Example." In *Theories and Models in Scientific Processes: Proceedings of AFOS '94 Workshop, August*, ed. W. E. Herfel, 15–26. Madralin and IUHPS 1994 Conference, August 27–29, Warsaw. Amsterdam: Rodopi.

Cruse, P. 2005. "Ramsey Sentences, Structural Realism and Trivial Realization." *Studies in History and Philosophy of Science* A 36 (3): 557–76.

Cruse, P., and D. Papineau. 2002. "Scientific Realism without Reference." In *The Problem of Realism*, ed. M. Marsonet, 174–89. Aldershot: Ashgate.

Earman, J. 1992. *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Goodman, N. 1983. *Fact, Fiction and Forecast*. 4th ed. Cambridge, MA: Harvard University Press.

Gower, B. 2000. "Cassirer, Schlick and 'Structural' Realism: The Philosophy of the Exact Sciences in the Background to Early Logical Empiricism." *British Journal for the History of Philosophy* 8 (1): 71–106.

Hacking, I. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hardin, C. L., and A. Rosenberg. 1982. "In Defense of Convergent Realism." *Philosophy of Science* 49 (4): 604–15.

Hempel, C. G. 1945. "Studies in the Logic of Confirmation." *Mind* 54 (213): 1–26.

Kitcher, P. 1981. "Explanatory Unification." *Philosophy of Science* 48 (4): 507–31.

———. 1993. *The Advancement of Science*. Oxford: Oxford University Press.

Kripke, S. A. 1980. *Naming and Necessity*. Oxford: Blackwell.

Laudan, L. 1981. "A Confutation of Convergent Realism." *Philosophy of Science* 48 (1): 19–49.

———. 1984. "Realism without the Real." *Philosophy of Science* 51 (1): 156–62.

Lyons, T. D. 2006. "Scientific Realism and the *Stratagema de Divide et Impera*." *British Journal for the Philosophy of Science* 57 (3): 537–60. doi:10.1093/bjps/axl021.

Musgrave, A. 1988. "The Ultimate Argument for Scientific Realism." In *Relativism and Realism in Science*, ed. R. Nola, 229–52. Dordrecht: Kluwer.

Newman, M. 2005. "Ramsey-Sentence Realism as an Answer to the Pessimistic Meta-induction." *Philosophy of Science* 72:1373–84.

Papineau, D. 2010. "Realism, Ramsey Sentences and the Pessimistic Meta-induction." *Studies in History and Philosophy of Science* A 41 (4): 375–85.

Post, H. 1971. "Correspondence, Invariance and Heuristics: In Praise of Conservative Induction." *Studies in History and Philosophy of Science* A 93 (3): 213–55.

Psillos, S. 1994. "A Philosophical Study of the Transition from the Caloric Theory of Heat to Thermodynamics: Resisting the Pessimistic Meta-induction." *Studies in History and Philosophy of Science* A 25 (2): 159–90.

———. 1995. "Is Structural Realism the Best of Both Worlds?" *Dialectica* 49 (1): 15–46.

———. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Putnam, H. 1975a. *Mathematics, Matter and Method*. Vol. 1 of *Philosophical Papers*. Cambridge: Cambridge University Press.

———. 1975b. *Mind, Language and Reality*. Vol. 2 of *Philosophical Papers*. Cambridge: Cambridge University Press.

———. 1978. *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Stanford, P. K. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Vickers, P. 2013. "A Confrontation of Convergent Realism." *Philosophy of Science* 80 (2): 189–211.

Whewell, W. 1858/1968. *William Whewell's Theory of Scientific Method*. Ed. R. Butts. Pittsburgh: University of Pittsburgh Press.

Worrall, J. 1989a. "Fresnel, Poisson and the 'White Spot': The Role of Successful Prediction in Theory-Acceptance." In *The Uses of Experiment—Studies of Experimentation in Natural*

Science, ed. D. Gooding, T. J. Pinch, and S. Schaffner, 135–58. Cambridge: Cambridge University Press.

———. 1989b. "Structural Realism: The Best of Both Worlds?" *Dialectica* 43:99–124.

———. 2000. "The Scope, Limits, and Distinctiveness of the Method of 'Deduction from the Phenomena': Some Lessons from Newton's 'Demonstrations' in Optics." *British Journal for the Philosophy of Science* 51 (1): 45–80.

———. 2002. "New Evidence for Old." In *In the Scope of Logic, Methodology and Philosophy of Science*, ed. P. Gardenfors, J. Wolenski, and K. Kijania-Placek, 191–209. Dordrecht: Kluwer.

———. 2006. "Theory-Confirmation and History." In *Rationality and Reality: Conversations with Alan Musgrave*, ed. C. Cheyne and J. Worrall, 31–61. Dordrecht: Springer.

———. 2007. "Miracles and Models: Why Reports of the Death of Structural Realism May Be Exaggerated." *Royal Institute of Philosophy Supplements* 82 (61): 125–54.

———. 2010. "Theory Confirmation and Novel Evidence: Error, Tests, and Theory Confirmation." In *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, ed. D. G. Mayo and A. Spanos, 125–54. Cambridge: Cambridge University Press.

Worrall, J., and E. Zahar. 2001. "Appendix IV: Ramseyfication and Structural Realism." In *Poincare's Philosophy: From Conventionalism to Phenomenology*, ed. E. Zahar, 236–51. Chicago: Open Court.