THE JOURNAL OF PHILOSOPHY
VOLUME C, NO. 11, NOVEMBER 2003

PYRRHIC VICTORIES FOR SCIENTIFIC REALISM*

Scientific realists suggest we have good reasons to believe that our best current scientific theories offer us literally true (or probably true, or approximately true, or probably, approximately true) descriptions of how things stand in otherwise inaccessible domains of nature. Their most influential argument for this position is the abductive, explanationist, or "miracle" argument: in Hilary Putnam's memorable phrase, that the success of science would be a miracle if our scientific theories were not true; more formally, that the (approximate) truth of our current scientific theories provides at least the best and perhaps the only explanation for the empirical successes those theories enjoy.

Such realism stands challenged most powerfully by the *pessimistic induction* over the history of scientific inquiry (PI): thinkers ranging from Pierre Duhem and Henri Poincaré to Larry Laudan have repeatedly pointed out that the history of science exhibits a parade of eminently successful theories that have nonetheless been subsequently rejected. And they ask why we should not believe that the same fate awaits our own successful theories.[1]

* My thanks to David Malament, Jarrett Leplin, Philip Kitcher, Stathis Psillos, Larry Laudan, Jeff Barrett, Pen Maddy, Anjan Chakravartty, Alan Nelson, Alexander Rosenberg, Arthur Fine, and the members of a number of graduate seminars in which I have addressed these issues for helpful discussion, suggestions, and comments. This material is based upon work supported by the National Science Foundation under Grant No. SES-0094001. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

[1] Two anonymous referees rightly note that I here follow the regrettably common practice of characterizing the relevant sorts of success at a painfully vague or intuitive level, though I am happy to accept the appealingly commonsensical account of such success (grounded in the potential for solving practical problems of prediction and intervention) offered by Kitcher (in "Real Realism: The Galilean Strategy," *The Philosophical Review*, cx, 2 (April 2001): 151–97, especially pp. 166–67). As a general

0022-362X/03/0011/553–72

Of course such a challenge cannot depend upon agreement with current theories as a test of truth, so historical examples offer classically inductive evidence of the *falsity* of our own successful theories only if we have some other ground for regarding these successful predecessors as definitively refuted. But the challenge need not be formulated in this way in any case. The ghostly historical procession shows that past theories have been quite successful in just the ways that impress contemporary realists while nonetheless making fundamental claims about nature *inconsistent* with those of present successful theories (as well as those of other past successful theories in the same domains). Since *at most* one theory in any such inconsistent set can be true, we are provided with abundant inductive evidence that empirical success cannot possibly be a reliable indicator of the truth about nature, no matter what the truth is: any general practice of inferring from such successes to the approximate truth and/or reference of the theories that enjoy them would routinely if not invariably have to lead us astray. Alternatively, we can see the realist inference from success to approximate truth and/or reference as *self-undermining*, for if the success of current theories leads us to conclude that they are approximately true and/or referential, this implies in turn that many past theories must have been radically false and/or nonreferential *despite* being successful, undermining our original ground for concluding that current theories are approximately true and/or referential in the first place.[2]

Recent decades have witnessed not only impressive realist attempts to block or blunt this historical challenge by engaging the details of the history of science itself, but also the first serious efforts to recruit those details to the realist cause. While this engagement is certainly welcome, I will here argue that the most promising and influential of these efforts (including that of Clyde Hardin and Alexander Rosenberg, and, more recently, those of Philip Kitcher, Stathis Psillos, Jarrett Leplin, and John Worrall) manage to achieve only Pyrrhic victories

---

matter, however, defenders of the PI have resisted tying themselves too closely to *any* particular conception of scientific success, seeking instead to argue that there is (at most) a difference of degree and not in kind between present and past theories with respect to *whatever* sorts of success realists suppose could only be explained by the truth of the theories which enjoy them (see, for example, Laudan's "A Confutation of Convergent Realism," reprinted in David Papineau, ed., *Philosophy of Science* (New York: Oxford, 1996), pp. 107–38).

[2] This last formulation is similar to that offered by John Worrall in "How to Remain (Reasonably) Optimistic: Scientific Realism and the 'Luminiferous Ether'," in D. Hull, M. Forbes, and R. M. Burian, eds., *P.S.A. 1994*, Volume 1 (East Lansing, MI: Philosophy of Science Association, 1994), pp. 334–42, here p. 334.

for realism—that is, "defenses" of the approximate truth of past theories and/or the referential status of their central theoretical terms that are forced to concede to the realist's opponent either just the substantive points that were in dispute between them or everything she needs for a convincing historical case against realism. Although the resulting positions might fairly be said to be realist in name only, the point is far from merely terminological: in each of the cases I will discuss, those who seek to use the historical record to defend realism seem unaware that they have sacrificed the substantive tenets of the realist position on the altar of its name.[3]

I. REFERENCE WITHOUT DESCRIPTIVE ACCURACY

One strategy of realist reply to the historical challenge suggests that Laudan's classic defense of the PI exaggerates the extent to which the central terms of the rejected past theories he considers should be judged nonreferential by present lights. Laudan seems to require, for example, that the descriptive claims theories make about their central posits be largely accurate in order for the central terms in those theories to refer successfully, and some critics have suggested that causal accounts of reference reveal this requirement to be prejudicially restrictive: Hardin and Rosenberg,[4] for instance, argue that because "[o]ne permissible strategy of realists is to let reference follow causal role," realists are free to regard the central terms of even those theories they regard as radically mistaken to have been referential

[3] Of course, even if I am right to think that recent realist efforts achieve only Pyrrhic victories for the realist cause, we will still be left with a difficult question about just how to regulate our beliefs about the world in response to our historical predicament. Although I have characterized the PI as a challenge to our scientific theorizing about "otherwise inaccessible domains of nature," the distinction between observables and unobservables which figures so centrally in other influential challenges to realism (including, most famously, Bas van Fraassen's in *The Scientific Image* (New York: Oxford, 1980)) does not seem especially apposite here, for our characterizations of many observable entities (for example, chemical elements, evolutionary adaptations, supernovae) would seem to be routinely grounded in just those sorts of theoretical conceptions of the natural world that stand challenged by the PI itself. In forthcoming work I argue that the historical challenge is correctly focused not on our beliefs about entities of a particular sort (for example, "unobservables") but instead on beliefs *arrived at* in a particular way: more specifically, beliefs about nature that are reached *eliminatively*, that is, by coming to embrace one belief among a set of possibilities by eliminating its competitors, whether we so reason about observable or unobservable aspects of the natural world. But we would do well to settle first whether *any* serious challenge survives the most sophisticated and dogged recent efforts to convince us that the history of science actually poses no significant threat to scientific realism at all, and I will therefore defer to another occasion the general question of precisely which sorts of contemporary scientific beliefs the challenge of history calls into question.

[4] "In Defense of Convergent Realism," *Philosophy of Science*, XLIX (1982): 604–15.

after all. For example, because we regard the electromagnetic field as playing the causal role attributed to the ether by earlier physical theories, the realist may hold, they suggest, "that 'ether' referred to the electromagnetic field all along" (*ibid.*, pp. 613–14).

As a response to the PI, however, this misses the forest for the trees. The sort of account envisioned by Hardin and Rosenberg secures a history of successful reference for terms in discarded theories only by divorcing their reference from the question of the accuracy of those theories and thus *abandoning* the specifically theoretical beliefs of the very sort for which the realist hopes to convince us to share her realism in the case of current theories. But this runs the realist afoul of what we might call the "trust" argument: after all, if the central terms in past theories are held to be referential *despite* the fact that the theories in which they are embedded repeatedly turn out to be radically misguided, then the historical record seems to entitle the antirealist to her claim that we would be foolish to *trust* either the theoretical accounts of inaccessible domains of nature offered by successful contemporary scientific theories or the descriptions associated with their central theoretical terms. And trusting the accounts of such domains and entities given by (some) current theories is just what the realist hoped to convince us to do!

Precisely because this shortcoming of Hardin and Rosenberg's account is so easily recognized,[5] it is surprising that a version of the same problem confronts more recently influential and sophisticated realist approaches to the reference of theoretical terms. Kitcher,[6] for instance, argues convincingly that the particular *tokens* of a speaker's use of a natural kind term must be separated and her dominant linguistic intentions considered in assigning reference to them: thus the reference of some of Priestly's tokens of 'dephlogisticated air' are fixed by his intention to refer to the substance emitted in combustion, while the reference of other tokens is fixed by his intention to refer to the substance whose inhalation was rendering his breathing particularly light and easy or the substance he "exploded together" with "inflammable air" to produce water or nitric acid. In the former cases, Priestly's tokens of 'dephlogisticated air' fail to refer (as there is no substance emitted in combustion in the way Priestly imagines), but the latter tokens instead refer to oxygen. Likewise, some of Fresnel's

[5] See Worrall, "Structural Realism: The Best of Both Worlds?" *Dialectica*, XLIII (1989): 99–124, especially pp. 116–17; Laudan, "Discussion: Realism without the Real," *Philosophy of Science*, LI (1984): 156–62, especially p. 161; and Psillos, *Scientific Realism: How Science Tracks Truth* (New York: Routledge, 1999), especially chapter 12.

[6] *The Advancement of Science* (New York: Oxford, 1993), chapter 4.

tokens of 'light wave' fail to refer because their reference is fixed by their theoretical description as oscillations of molecules of the ether, while the reference of others is fixed by Fresnel's dominant intention to talk about light, however it is in fact constituted, and therefore refer to electromagnetic waves of high frequency.[7] Kitcher concludes that claims of referential failure for past theories are overstated, and that the heterogeneity of the different modes of reference-fixing in different contexts of use permitted past theorists to make many referential and indeed true claims about the world.

But for all its success in rescuing the terms in discarded theories from blanket assertions of referential failure, Kitcher's approach runs afoul of the trust argument in just the same way that Hardin and Rosenberg's much simpler appeal to pure causal theories of reference did: Kitcher manages to rescue the reference of the central terms in discarded theories only *on those occasions* in which the user's dominant linguistic intentions explicitly eschew those specifically *theoretical* descriptions (like 'the substance emitted in combustion' or 'the oscillations of molecules of the ether') associated with her terms. But surely it offers little comfort to the realist if we insist that some tokens of terms like 'dephlogisticated air', and 'light wave' in rejected theories referred after all while admitting that the relevant theoretical accounts and descriptions of those entities were mistaken about virtually everything except the fact that the entities in question played some causal role in producing observable phenomena, for it is (once again) ultimately those very theoretical accounts and descriptions which the realist hopes to defend in the case of current theories. Thus, Kitcher offers a welcome sophistication to our account of reference, but one that makes no progress whatsoever in defending realism from the historical challenge: his account shows how tokens of the central terms of past theories were referential (and past theorists were able to enunciate important truths) just where their being so (or doing so) did not depend upon those theories actually getting anything much right about the natural world.[8]

[7] This strategy of analysis is further developed in connection with a general causal-descriptivist account of reference for natural kind terms in Stanford and Kitcher, "Refining the Causal Theory of Reference for Natural Kind Terms," *Philosophical Studies*, XCVII (2000): 99–129.

[8] Of course, Kitcher's further argument (*Advancement of Science*, chapter 5) that those *parts* of superceded theories enabling them to be successful were also true (see also Psillos) would seem to require that at least some of the central terms of past theories referred successfully even on occasions when their reference was fixed in what he calls the "descriptive mode" *and the descriptions in question were theoretical* (and conversely, that where uses of past terms were sufficiently infected with false theory as to be nonreferential, they were not involved in or responsible for those theories' successes). I argue below that this argumentative strategy provides the realist's oppo-

By contrast, Psillos seems painfully aware that realists undermine their own cause in establishing mere referential continuity without descriptive accuracy for past theories. His own defense of realist reference (*op. cit.*, chapter 12) is carefully constructed to avoid this particular pitfall by requiring at least some of the descriptive information associated with terms by our theories to be accurate in order for those terms to refer. It is especially revealing, then, that his account nonetheless concedes to the realist's opponent everything she needs to make her historical case. Let us see why.

Psillos argues that because pure causal theories make referential continuity too easy to get (that is, terms cannot fail to refer), any convincing account of reference along these lines will have to be "causal-descriptivist"; that is, it will have to require not simply that a term refer to whatever causes the phenomena that occasion its introduction into the language, but in addition that some of the fundamental descriptions associated with the term actually be satisfied by the entity to which it refers. But not all the associated descriptions are equally important: he argues that "some descriptions associated with a term are less fundamental in view of the fact that the posited entity would play its intended causal role even if they were not true" (*op. cit.*, p. 297). The descriptions that really count, then, and which must be satisfied by an entity in order for a term to refer to it, are those making up what he calls the theory's "core causal description" of the entity in question: the descriptions that would have to be true in order for the entity to play the causal role the theory assigns to it. Thus, he argues, the term 'phlogiston' failed to refer because "there is nothing which fits a description which assigns to phlogiston the properties it requires in order to play its intended causal role in combustion" (*op. cit.*, p. 291), in particular "the property that it is released during the process of combustion" (*op. cit.*, p. 298).

It is crucial to Psillos's account, however, that the term 'ether' in nineteenth-century wave theories of light and electromagnetism turn out to be referential, for (in contrast to Kitcher) he acknowledges that the postulation of the ether (as a "dynamical structure" serving as a carrier for light waves) played a crucial role in the successes of those theories. He argues that this demand creates no problem for the realist, however, because our own term 'electromagnetic field' shares the core causal description associated with the 'ether' of nineteenth-century optics, and so, he claims, the latter term referred to

---

nent with all the resources she needs to make a convincing historical case against realism.

the electromagnetic field itself.[9] Although current physical theories regard the beliefs of nineteenth-century optical and electromagnetic theorists about the *nature and constitution* of the ether to have been quite radically and fundamentally misguided, Psillos insists that such beliefs did not form any part of the actual core causal description of the term 'ether'. That is, the false beliefs of nineteenth-century theorists about the material and mechanical character of the ether were not part of the description that an entity would have to satisfy in order to play the causal role assigned to the ether by nineteenth-century wave theories of light and electromagnetism.

Of course, this account of the matter invites the realist to choose the core causal descriptions she associates with the central terms of past theories rather carefully, with one eye on current theories' claims about nature, so there is more than a whiff of ad hoc-ery about the proposal. But even if we set this worry aside and assume that the realist can delicately titrate the core causal descriptions she associates with the crucial terms in successful past theories so as to render them referential by the lights of current theories, Psillos's victory will *nonetheless* remain a Pyrrhic one. The reason is that this case for the referential status of central terms in successful past theories simply invites from the historical record a renewed form of the pessimistic induction itself, this time concerning our ability to distinguish (at the time a theory is a going concern) which of our beliefs about an entity are *actually* part of its core causal description. To see why, let us look more closely at the one case of reference Psillos examines in detail: the luminiferous ether of nineteenth-century optics.

The core of the problem is that nineteenth-century theorists themselves strenuously disagreed with *the very assessment Psillos offers* of what would have to be true of an entity in order for it to play the ether's causal role, that is, with his very claim about *which* descriptions of the ether enter into its core causal description. They considered whether the ether could play the role ascribed to it in propagating light and other electromagnetic waves without consisting of a mechanical medium of some kind *and explicitly denied that it could*! Maxwell himself concludes *A Treatise on Electricity and Magnetism*[10] with the following

[9] Actually, Psillos's claim seems better suited to the early twentieth-century conception of the electromagnetic field than the contemporary one: it is at least contentious to describe the electromagnetic field recognized by contemporary quantum electrodynamics as one in which light waves *propagate* at all (c.f. Psillos, p. 296). But I will not rely on this point in what follows: instead I will show that Psillos's realist victory is hollow even if we treat this early twentieth-century conception of the field as our own.

[10] 3rd edition in two volumes (London: Oxford, 1955 [1873]).

resolute insistence that the ether must be a material medium *in order to play the causal role* of propagating energy waves:

> If something is transmitted from one particle to another at a distance, what is the condition after it has left the one particle and before it has reached the other? If this something is the potential energy of two particles, as in Neumann's theory, how are we to conceive this energy as existing in a point of space, coinciding neither with the one particle nor with the other? *In fact, whenever energy is transmitted from one body to another in time, there must be a medium or substance in which the energy exists after it leaves one body and before it reaches the other*, for energy, as Torricelli remarked, 'is a quintessence of so subtile a nature that *it cannot be contained in any vessel except the inmost substance of material things*'. Hence all these theories lead to the conception of a medium in which propagation takes place, and if we admit this medium as an hypothesis, I think it ought to occupy a prominent place in our investigations, and that we ought to endeavor to construct a mental representation of its action, and this has been my constant aim in this treatise (*ibid.*, Volume II, p. 493, my emphasis).

Maxwell's positive conception of the field is a matter of some legitimate controversy, but here he quite explicitly denies that anything *besides* a mechanical or material substance is capable of playing the causal role he assigns to the ether: he insists that to avoid the patent (or at least inconceivable) absurdity of "energy...existing in a point of space" we must recognize the existence of a material medium filling the space between bodies, for energy "cannot be contained in any vessel except the inmost substance of material things." Thus, Maxwell self-consciously insisted that the "core causal description" of the ether *included* the very beliefs about the ether's material and mechanical character that Psillos grants are not satisfied by the modern conception of the electromagnetic field as a distinct entity ontologically on par with matter itself.[11]

[11] Of course, the contemporary conception of the electromagnetic field ascribes to it a well-defined (local) mass-energy content and it may therefore be said to qualify as a "material substance" in *some* current sense of this term. But the sense in question is far removed from the requirements Maxwell had in mind in insisting that the ether must be a "medium or substance" or denying that energy can reside anywhere "except the inmost substance of material things" (and Psillos seems to grant as much in recognizing that nineteenth-century theorists' beliefs about the nature and constitution of the ether have turned out to be mistaken). Thus, we must understand Maxwell as insisting that nothing besides a material or mechanical substance as *he* understood these notions could play the causal role he assigned to the ether, and it is the inaccuracy of this judgment (by modern lights) which suggests a renewed PI concerning our ability to discern just which parts of the descriptions we associate with terms in our successful theories are really part of their "core causal descriptions." My thanks to David Malament for helpful discussion on this point.

Thus, even if Psillos can convincingly argue that the *actual* "core causal description" of the ether does not include claims about its material or mechanical character, he will be forced to concede that the carefully considered judgments of leading scientific defenders of the theory concerning which of the descriptions associated with its central terms must be satisfied by an entity in order for it to play the causal role associated with the term (that is, which features figure in the actual core causal description) have proved to be unreliable. What this suggests, of course, is that we cannot rely on our *own* judgments about which of the descriptions we associate with our own terms are *genuinely* part of their own core causal descriptions, delivering us back into the arms of the trust argument with respect to current theories.[12]

Once this underlying problem with Psillos's strategy is recognized, further historical examples are relatively easy to identify. In his celebrated defense of the "continuity of the germ-plasm," August Weismann[13] argues vigorously that the germinal or hereditary materials are located in the nuclei of cells and that they must have an internal structure in which different parts of this germ-plasm are responsible for the production of varying characteristics in different cells. Further, he argues explicitly that the germinal materials must be separated and reduced in each subsequent cell division, until each cell contains only the tiny amount necessary for determining its own characteristics. He argues that there simply is no other way to account for the fact that different parts of the body possess different characteristics, and, in particular, insists that the cells of the body could not possibly be heterogeneous if each cell carried precisely the same hereditary materials:

> As the thousands of cells which constitute an organism possess very different properties, *the chromatin* which controls them *cannot be uniform; it must be different in each kind of cell.*
>
> The chromatin, moreover, cannot *become* different in the cells of the fully formed organism; the differences in the chromatin controlling the cells must begin with the development of the egg-cell, and must increase

[12] Psillos admits that there is an element of rational reconstruction in settling on the appropriate core causal descriptions for terms in past theories, but this glides over the real problem we have noted: without justifiable confidence in our ability to specify accurately core causal descriptions for theoretical entities, we do not know which features of our *own* theories are rightly included in the core causal descriptions associated with *their* central terms, and thus have no way to pick out which of our own actual beliefs we can trust.

[13] *The Germ-Plasm: A Theory of Heredity*, W. Newton Parker and Harriet Rönnfeldt, trans. (New York: Scribner's, 1893 [1892]).

> as development proceeds; for otherwise the different products of the division of the ovum could not give rise to entirely different hereditary tendencies. This is, however, the case. Even the first two daughter-cells which result from the division of the egg-cell give rise in many animals to totally different parts.... The conclusion is inevitable that the chromatin determining these hereditary tendencies is different in the daughter-cells (*ibid.*, p. 32, original emphasis).

The point here is not simply that Weismann made (by current lights) some mistakes about the constitution of the germ-plasm, still less is it important whether these false beliefs prevented particular tokens of 'chromatin' or 'germ-plasm' from referring on particular occasions of use. Rather, the point is that Weismann argued quite explicitly that *no entity could possibly play the causal role that he assigned to the hereditary material without being separated and parceled out differently to different cells*, offering further evidence of our historical unreliability in judging which theoretical descriptions are genuinely part of a term's core causal description.

Psillos's strategy is similarly undermined by cases in which the term in question is presently judged to be uncontroversially *non*referential. Consider Antoine Lavoisier's claim, defending the caloric fluid theory of heat in his 1785 *Memoir on Phlogiston,* that

> One can hardly think about these [thermal] phenomena without admitting the existence of a special fluid [whose accumulation causes heat and whose absence causes cold]. It is no doubt this fluid which gets between the particles of bodies, separates them, and occupies the spaces between them. Like a great many physicists I call this fluid, whatever it is, *the igneous fluid, the matter of heat and fire.*[14]

The point is not, of course, that Lavoisier was wrong about the need to posit caloric fluid. Rather, Lavoisier here suggests that nothing *besides* a subtle fluid could play the role in causing thermal phenomena that he assigns to caloric (but that contemporary theorists assign to molecular motion). This assessment is drawn even more explicitly in terms of caloric fluid's causal role in Lavoisier's later work:

> It is difficult to conceive of these phenomena without admitting that they are the result of a real, material substance, of a very subtile fluid, that insinuates itself throughout the molecules of all bodies and pushes them apart.... This substance, whatever it is, is the cause of heat, or in

[14] "Mémoire sur phlogiston," as translated in A. Donovan, *Antoine Lavoisier: Science, Administration, and Revolution* (Cambridge: Blackwell, 1993), here p. 171, original emphasis, translation modified.

> other words, the sensation that we call heat is the effect of the accumulation of this substance....[15]

Lavoisier's nonreferential account of caloric shows, no less than Maxwell's supposedly referential discussion of ether and Weismann's putatively referential account of the germ-plasm or chromatin, that we have historically been unreliable (again, by present lights) in judging which descriptions must be satisfied by an entity or what characteristics it must have in order for it to play the causal role assigned to it by a particular theory. Like Kitcher's then, Psillos's defense of the referential status of some central terms in past theories comes only at the cost of leaving us unable to be confident that any particular descriptions we associate with a referring term (even ones we presently regard as absolutely central and/or indispensable to fulfilling its causal role) will be retained in the development of further theoretical science. He therefore achieves a victory for realism only at a price that realists cannot afford to pay.

II. DILUTING APPROXIMATE TRUTH

Related difficulties afflict the complementary strategy of suggesting that Laudan is too quick to deny that his examples of superseded theories were approximately true. There is a further complication here, in that Laudan takes advantage of an argumentative shortcut in making his case against the approximate truth of rejected theories: he assumes explicitly that the failure of a theory's central terms to refer (in a sense that requires descriptive accuracy) ensures that the theory is not approximately true, arguing that "the *realist would never want to say that a theory was approximately true if its central theoretical terms failed to refer.* If there were nothing like genes, then a genetic theory, no matter how well confirmed it was, would not be approximately true."[16] Hardin and Rosenberg (*op. cit.*) protest, insisting that a case

[15] *Traité de chimie*, in Volume 1 of *Oeuvres de Lavoisier*, 6 volumes, volumes 1–4, J. B. Dumas, ed. volumes 5–6, Edouard Grimaux, ed. (New York: Johnson Reprint Corporation, 1965 [originally published in Paris: Imprimerie Impériale, 1862–1893]) p. 19, my translation.

[16] "A Confutation of Convergent Realism," p. 33, original emphasis. Laudan avails himself of this shortcut in part because he argues that realists have failed to provide any account of approximate truth on which the presumption that the approximate truth of a theory implies or entails its likely success can be defended. He suggests that typically this presumption is uncritically assumed to follow from the unobjectionable fact that a perfectly true (and, we might add, complete) theory would be perfectly successful—an inference that he insists is patently invalid—and he challenges realists to provide an analysis of approximate truth on which the presumption of success for approximately true theories is defensible. Thus, Laudan cannot afford to pin his argument on any particular conception of approximate truth itself.

like classical Mendelian genetics illustrates how a theory might have central terms that fail to refer by Laudan's descriptive criteria (for they insist, perhaps implausibly, that contemporary genetics recognizes nothing like a Mendelian gene) but nonetheless earn a judgment of approximate truth by current practitioners. Thus, they claim, Laudan's case relies inappropriately on the reference of central terms as a minimal condition for approximate truth.[17]

Laudan's own response to this objection asks by what right Hardin and Rosenberg "take contemporary theories as benchmarks of what there is and how it behaves" once they have granted that a successful theory (like Mendel's) may be very wide of the ontological mark and thereby undermined the realist's abductive argument itself.[18] But this reply risks missing the point of the objection: Hardin and Rosenberg can respond that they are simply showing how a current theory, *if* true, could ground the judgment that a particular past theory was both nonreferential and approximately true, thereby invalidating Laudan's argumentative shortcut from failures of reference to failures of approximate truth, and thus undermining the case he makes against the realist's abductive inference in the first place.

More effective, therefore, is to ask what weight this sort of objection to Laudan's argumentative shortcut is ultimately supposed to carry. Even if we grant Hardin and Rosenberg that the relationship between Mendelian and contemporary genetics illustrates how one theory can be (by another's lights) both nonreferential and approximately true, *this is surely not what the verdict of our own actual contemporary theories would be concerning most, if not all, of the other theories on Laudan's historical hit parade.* If, as the realist would have it, our contemporary theories are true, then we are inclined to insist that the many eminently successful theories included on Laudan's list (including the phlogiston theory of chemistry, the caloric theory of heat, the vital forces theories of physiology and embryology, and various optical and electromagnetic ether theories) are *not* in fact true, not even approximately so: the relationship between current theories and those on Laudan's list is simply not, in general, the one that Hardin and Rosenberg claim obtains between Mendelian and contemporary molecular genetics. Furthermore, the claim that these successful past theories are not even approximately true by contemporary lights is indeed

[17] Not all of Laudan's examples proceed in this fashion, however: he takes explicit notice of several theories whose central terms *did* refer, but which he suggests were nonetheless not approximately true ("A Confutation of Convergent Realism," pp. 123–24).

[18] "Realism without the Real," p. 159.

strongly *supported* by the fact that current theories hold that there is nothing like the central posits of those superseded theories, even if this fact is not alone sufficient to guarantee their radical falsity by contemporary lights.

Of course, Hardin and Rosenberg might instead have in mind some sense of "approximate truth" in which the literal truth of current theories is indeed consistent with the approximate truth of such classic success stories as the phlogiston theory of chemistry, the caloric theory of heat and the theories of the optical and electromagnetic ethers, but if so they will quickly find themselves back in the jaws of the trust argument. This is because the "approximate truth" of past successful theories, in *any* sense that is consistent with being as fundamentally and profoundly mistaken about the constitution of nature as these famous predecessors were by the lights of current theories, simply does not cut against the antirealist's insistence that the historical record shows why it would be a mistake to trust or believe the theoretical accounts of nature they offer or that doing so would have routinely led us badly astray in the past. Thus the realist will win a battle over something she calls "approximate truth," but again lose the war over realism, for her opponent's skepticism about the accounts of nature offered by current successful theories will rightfully survive her concession that those theories may well be "approximately true" in the attenuated sense the realist has managed to defend for past theories.

This same point undermines Kitcher's suggestive analogy between the response of modest scientific realists to the historical record and "that of the author who confesses in her preface that she is individually confident about each main thesis contained in her book but equally sure that there's a mistake somewhere."[19] The analogy is clever, for it invites us to see the realist's opponent as simply carping over our inability to attain an unreasonable standard of accuracy in what is admittedly a difficult business. Nonetheless, even when we restrict our attention (as Kitcher insists) to those historical cases of past theories grounding "predictions and interventions that were numerous, diverse, and hard to achieve,"[20] the analogy proves to be seriously misleading, for contemporary theories are surely not rightly thought to hold such successful predecessors as the wave theory of light, Newtonian mechanics, and the caloric theory of heat to have been mistaken simply in matters of minor detail comparable to a misplaced

[19] "Real Realism: The Galilean Strategy," p. 171; see also Kitcher's *Science, Truth, and Democracy* (New York: Oxford, 2001), pp. 18–19.

[20] *Science, Truth, and Democracy*, p. 19.

footnote, a speculative musing, or even an overstated conclusion. Instead, they hold these illustrious predecessors to have been deeply and thoroughly mistaken in their most central claims about the constitution and/or operation of nature, a situation more analogous to the author having been fundamentally mistaken in the principal thesis of her book or the central contentions she was concerned to advance.

Other recently influential efforts to defend realism from the historical record turn out to dilute approximate truth in ways that are similarly self-defeating. Leplin's extended defense of realism,[21] for example, champions the inference from a theory's success in making *novel* predictions (in a precisely specified sense of novelty) to what he calls its "partial truth" (*ibid.*, p. 127). Leplin modestly aims to defend only what he calls minimal epistemic realism: the claim that there *are* epistemic conditions that would warrant a realist attitude towards a theory, not that any present theory actually satisfies such conditions. But the partial truth Leplin argues we can infer from a theory's novel predictive success is nonetheless too meager to render the prospect of this inferential entitlement of any real significance in the debate over realism.

This is because Leplin rightly sees that he cannot infer any particular degree, kind, or respect of partial truth (in the sense of representational accuracy[22]) from any particular degree, kind or respect of success in novel prediction. He grants that he is "vague by default as to how much novel success merits what level of confidence in representational success," explicitly denies that "the amount of novel success provides a measure of the degree of representational accuracy achieved," and acknowledges that novel predictive success does not warrant attributing "a particularly high level of accuracy, because we have no way to determine what forms of inaccuracy might be irrelevant to the observable situation" (*ibid.*, p. 128). But with no inferential connection between degrees or kinds of novel predictive success and degrees or kinds of representational accuracy, we will never be able

[21] *A Novel Defense of Scientific Realism* (New York: Oxford, 1997).

[22] Leplin articulates two quite distinct senses of partial truth. The "pragmatic" sense seeks to capture what practicing scientists commit themselves to in accepting theories, both prospectively (contrasting the partial truth of present theories with falsity) and retrospectively (contrasting the partial truth of past theories with unqualified truth), while the "metaphysical" sense of partial truth is understood in terms of representational accuracy and is what we are supposedly entitled to infer from a theory's success in novel prediction. It is only this second, metaphysical sense of partial truth that will concern us here. My thanks to Leplin for clarifying this and other aspects of his work in correspondence, though he would not be at all satisfied by the conclusions I reach below.

to say anything more about any theory (even a merely possible future theory) than that it enjoys "some" degree of representational accuracy, no matter how much novel predictive success it has enjoyed. And this (unimprovable) claim is trivially satisfied: after all, Aristotelian mechanics, creationist biology, caloric thermodynamics, and phlogistic chemistry all enjoy "some" degree of representational accuracy, too—none of these theories is wrong or misleading about absolutely everything, not even everything fundamental.[23]

Thus, although he intends to eschew the traditional realist commitments he regards as indefensible, Leplin cannot, I suggest, defend even a minimal realism with such a feeble connection between novel predictive success and representational accuracy or partial truth. Instead his account dilutes the notion of partial truth to the point that the historical record of such partially true theories supports rather than opposes the antirealist's skepticism about the descriptive or representational accuracy of current (and even future) successful theories.[24]

Perhaps for this reason, Leplin sometimes seems tempted to assert a more fine-grained connection between novel predictive success and a particular degree or kind of representational accuracy: responding to the PI, he suggests that the fundamental theoretical mechanisms employed by theories that enjoyed novel predictive success have not been overturned by subsequent developments (*op. cit.*, p. 145). But he is fully aware of the problem that such a suggestion creates: the classic leading example of novel predictive success—the prediction of the Poisson bright spot—was made by a theory now regarded as

[23] For example, while none of creationism's causal or explanatory mechanisms are accepted in current biology, its theoretically motivated division of organisms into species is nearly identical to the leading approach in contemporary evolutionary theory (the biological species concept). The other three examples arguably enjoy some degree of representational accuracy by present lights even at the level of causal and explanatory mechanisms: for example, one of the important respects in which Kuhn famously suggests (in *The Structure of Scientific Revolutions*, 3rd edition (Chicago: University Press, 1996), pp. 206–07) that "Einstein's general theory of relativity is closer to Aristotle's [mechanics] than either of them is to Newton's" is presumably that general relativity (like Aristotelian mechanics) recognizes gravitational motion as itself a "natural" state of motion (that is, along a "straight" trajectory in curved spacetime) not requiring further causal explanation rather than a deflection from natural (inertial) motion as in Newtonian mechanics.

[24] This also seems the appropriate response to Larry Sklar's contention (in *Theory and Truth* (New York: Oxford, 2000) especially section 4.1) that our best current theories should be viewed in light of history as "on the road to truth," "pointing towards the truth," or "heading in the right direction"; however, it is far from obvious that Sklar would disagree with this assessment and his central contention is that the most interesting and important issues are simply obscured at this level of abstraction in any case.

radically misguided, namely, Fresnel's formulation of the wave theory of light, with its conception of light waves as oscillations of the molecules of a material ether. Therefore, Leplin's "direct" response to the skeptical induction (*op. cit.*, p. 146f.) goes on to appeal hopefully to Kitcher's suggestion in *The Advancement of Science* that those *parts* or *aspects* of rejected theories actually responsible for their successes (in this case, particularly their successes in novel prediction) have been preserved in or ratified by subsequent theorizing about nature. That is, Kitcher argues (as does Psillos, *op. cit.*) that past practitioners had *selective* confirmation for only those *parts* of past theories that have turned out to be *true*.[25]

This same "selective confirmation" approach also proves central to Kitcher's more recent defense of realism using what he calls the "Galilean Strategy" of generalizing the inference from success-to-truth in everyday contexts like card games and attempts to use the subway system (where its reliability can be checked) to that of theorizing about the natural world.[26] Kitcher's defense of this generalization depends, as he recognizes, on establishing convincingly that theories' "past successes stem from parts of the theories that are approximately correct,"[27] that is, on the ability of the strategy of selective confirmation to turn apparently failed past instances of the success-to-truth inference strategy into successful ones; otherwise we have compelling reasons to doubt that the reliability of the everyday success-to-truth inference survives its Galilean importation into the quite different context of scientific theorizing.

Of course, to evade the trust argument, the appeal to selective confirmation will have to provide what we might call "prospectively applicable" criteria of selective confirmation: that is, criteria that could have been applied to past theories at the time and can now be applied to our own theories *in advance of any future developments* to say just which parts of past theories were (and just which parts of present theories are) genuinely confirmed by the successes they enjoy. Else-

[25] In private correspondence, Leplin has indicated that he is disinclined "to try to turn what is clearly a problem for realism into a positive argument" in the way I suggest here, because even when the theoretical mechanisms responsible for novel predictive success survive in subsequent theories they are sometimes "so radically reconceived...that I do not hold much hope for founding upon their retention greater specificity as to [what] descriptive content novel success warrants a commitment to." He thus rejects the response I explore here and will have to accept instead that success in novel prediction warrants only an inference to "some" degree of representational accuracy, an inference which I have argued trivializes the realism it seeks to defend.

[26] See "Real Realism: The Galilean Strategy," and *Science, Truth, and Democracy.*

[27] "Real Realism: The Galilean Strategy," p. 170.

where,[28] I have argued in detail that the actual criteria of selective confirmation on offer fail this test: the grounds on which Kitcher proposes to distinguish "working" from "idle" posits fail to discriminate the rejected posits of past theories (like 'ether') from the posits he hopes to defend in present ones (like 'genes'), while Psillos's claim that scientists' own judgments of selective confirmation have proved to be reliable itself depends upon a highly selective reading of the historical evidence.

We need not review the details of these arguments, however, to see why this strategy of selective confirmation offers us yet another instance in which the realist reply concedes everything her opponent needs to make her historical case, for a central aspect of the problem can be extracted just from the historical sources noted above. The passages cited from Maxwell, Weismann, and Lavoisier illustrate that we have repeatedly misidentified those parts, features, or aspects of our theories that (by realist lights) were genuinely implicated in or required for their successes: as in the case of Psillos's core causal descriptions, then, we here have the materials for a renewed version of the pessimistic induction itself, this time concerning our ability to determine, at the time a theory is a going concern, which parts, features, or aspects are *actually required for the successes* of that theory. And as the case of Maxwell illustrates in striking detail, our ability to draw this distinction accurately is suspect even in the case of theories' successes in making novel predictions: Maxwell was as clear and explicit as he could be in insisting that the very *coherence* of the wave theory's successful application required the existence of a material, mechanical, substantival ether. Without confidence in our ability to pick out the parts of theories needed for their successes while those theories are live contenders, however, we are in no position to identify those parts or features of our own theories we may safely regard as accurate descriptions of the natural world (even though we know that not all are) and the realist's opponent again seems entitled to the conclusion she wanted all along.

A version of this same problem threatens to undermine the contention of Worrall's "Structural Realism"[29] that *structural* commitments typically survive the demise of the predictively successful theories in which they occur (by contrast with the theory's "content,"[30] "ontol-

[28] "Selective Confirmation: No Refuge for Realism," forthcoming in *Philosophy of Science*, supplemental volume to *P.S.A. 2002*.

[29] Developed most fully in "How to Remain (Reasonably) Optimistic: Scientific Realism and the 'Luminiferous Ether'," and "Structural Realism: The Best of Both Worlds?"

[30] "Structural Realism," p. 117; "How to Remain (Reasonably) Optimistic," p. 340.

ogy,"[31] or its claims about the "nature" of entities[32]) and are therefore the legitimate objects of justified realist confidence. It is not at all clear that we can plausibly distinguish the merely "structural" claims of even current theories from their "content" or claims about the "nature" of entities, not to mention assess the record of past practitioners in making such discriminations successfully, but (to invoke the now familiar refrain) unless we can do both of these things we will not know what aspects of current theories are the ones we can safely regard as accurately describing features of the natural world. Appeals to vague intuitions simply will not do here: at best such an intuitive criterion renders the problem of reliable application especially acute, forcing the structural realist to qualify her beliefs in even the structural claims of a theory by her level of confidence that they are indeed structural claims. But even worse, a merely intuitive criterion of structure seems to run afoul of the historical record in even central cases: What prevents, for example, Weismann's insistence that germinal materials must be parceled out differently to heterogeneous parts of an organism's body from being a claim about the *structure* of inheritance and ontogeny?

Perhaps in light of such difficulties, Worrall sometimes suggests that the structural commitments of a theory consist simply of its equations or the abstract mathematical relationships it posits.[33] But there is something extremely misleading in saying even that the abstract mathematical relationships posited by past successful theories have described the "structure" of the natural world in ways that are still embraced by current theories. Francis Galton's ancestral law of inheritance, for instance, was the central mathematical formalism and the most predictively successful aspect (see below) of his "stirp" theory of inheritance: it claimed that the germinal materials received by each individual (and determining her characteristics) are composed of fractional contributions made directly from each ancestor, in the proportion

$$1/4p + 1/8pp + 1/16ppp\ldots$$

where p is the contribution from the parental generation (on each side), pp from the grandparental, and so on.[34] Of course it is true

[31] "How to Remain (Reasonably) Optimistic," pp. 336, 341.

[32] "Structural Realism," pp. 117–18; "How to Remain (Reasonably) Optimistic," p. 334.

[33] "Structural Realism," pp. 118–20; "How to Remain (Reasonably) Optimistic," p. 340.

[34] Galton's clearest formulation of the law of ancestral inheritance is found in his "The Average Contribution of Each Several Ancestor to the Total Heritage of the Offspring," *Proceedings of the Royal Society of London*, LXI (1897): 401–13.

enough to say that the fractional relationships described by Galton's ancestral law show up *somewhere* in the description of inheritance provided by contemporary genetics; consider the following passage from Robert Olby's seminal *Origins of Mendelism*:[35]

> Today Galton's ancestral law of inheritance still stands as a mathematical representation of the average distribution of continuously varying characters in a population of freely outbreeding individuals not subject to selection. It serves as a basis for predicting the average distribution of such characters in the population. It tells us that on an average a grandparental deviation will be diminished to one-eighth of its original magnitude in the grandchildren. The Mendelian theory, on the other hand, tells us that only one in eight grandchildren will have received his grandparent's genes for the said deviation. Expressed as averages for a population, however, both theories give the same prediction (*ibid.*, pp. 81–82).

But contemporary genetics does not recognize the fractional relationships expressed in Galton's law of ancestral heredity (that is, 1/4 from the parental generation (on each side), 1/8 from the grandparental, 1/16 from the great-grandparental, and so on, *ad infinitum*) as describing any fundamental or even especially significant aspect of the mathematical structure of inheritance.[36] By present lights, it would be extremely misleading, if not outright mistaken, to say that even the mathematical structure expressed by Galton's Ancestral Law is preserved in contemporary genetics. Thus, to evade the challenge of the historical record, the mathematized version of structural realism will have to retreat simply to the dogged insistence that such chunks of mathematical formalism as Avogadro's number, Fresnel's equations for the transmission of light, and Galton's Ancestral Law of Inheritance will be recoverable in *some* way, *some*where, *some*how from future science. This, of course, is a far cry from giving us *any* claim we can rely on as an accurate description of (even just the structure of) the natural world, and therefore invites the trust argument again with a vengeance. Like the others we have examined, then, Worrall's structuralist defense of a realist position seems to give the game away entirely: it either leaves us with no justifiable confidence in our ability to distinguish clearly those claims of current theories about the natural

[35] New York: Schocken, 1966.

[36] This seems especially clear if we keep in mind that Galton's law is expressed in terms of *generational* contributions to the stirp of the offspring from each side: it is, for instance, the grandparental *generation* on each side which contributes 1/8 to the makeup of the stirp as a whole, and thus each grandparent makes only 1/16 of the stirpal contribution.

world on which we may rely, or it forces us to draw such a distinction in a way that invites yet another renewed pessimistic induction over the historical record itself.

III. CONCLUSION

I have tried to identify a systematic pattern of argument that consistently reappears among recent serious and subtle efforts to recruit the historical record to the defense of scientific realism and to argue that this pattern of argument ultimately undermines the very cause it is intended to advance. Defenses of the reference of central terms in past theories like those offered by Hardin and Rosenberg and Kitcher simply give up the substantive tenets of realism they sought to defend, while Psillos's appeal to core causal descriptions, Leplin's appeal to partial truth, Kitcher's and Psillos's appeals to the selective preservation of those parts of our theories involved in their successes, and Worrall's appeal to historical continuity in structural commitments are all forced either to defend realist inferential entitlements that are so weak as to capitulate to the realist's opponent on the question of whether we can safely trust the accounts of nature given by current or future successful scientific theories, or to rely on some discriminatory ability with respect to features, aspects, or claims of current theories whose reliability is itself vulnerable to a compelling historical challenge. A convincing defense of realism from the specter of the historical record will have to do better.

P. KYLE STANFORD

University of California/Irvine