Bas C. van Fraassen

TO SAVE THE PHENOMENA*

After the demise of logical positivism, scientific realism has once more returned as a major philosophical position. I shall not try here to criticize that position, but rather attempt to outline a comprehensive alternative.[1]

I

What exactly is scientific realism? Naively stated, it is the view that the picture science gives us of the world is true, and the entities postulated really exist. (Historically, it added that there are real necessities in nature; I shall ignore that aspect here.[2]) But that statement is too naive; it attributes to the scientific realist the belief that today's scientific theories are (essentially) right.

The correct statement, it seems to me, must indeed be in terms of epistemic attitudes, but not so directly. The aim of science is to give us *a literally true story of what the world is like;* and the proper form of acceptance of a theory is to believe that it is true. This is the statement of scientific realism: "To have good reason to accept a theory is to have good reason to believe that the entities it postulates are real," as Wilfrid Sellars has expressed it. Accordingly, an anti-realism is a position according to which the aims of science can well be served without giving such a literally true story, and acceptance of a theory may properly involve something less (or other) than belief that it is true.

The idea of a literally true account has two aspects: the language is to be literally construed; and, so construed, the account is true. This divides the anti-realists into two sorts. The first sort holds that science is or aims to be true, properly (but not literally) construed. The second holds that the language of science should be literally construed, but its theories need not be true to be good. The anti-realism I advocate belongs to the second sort.

II

When Newton wrote his *Mathematical Principles of Natural Philosophy* and *System of the World,* he carefully distinguished the phenomena to be saved from the reality he postulated. He distinguished the "absolute magnitudes" that appear in his axioms from their "sensible measures" which are determined experimentally. He discussed carefully the ways in which, and extent to which, "the true motions of particular bodies [may be determined] from the apparent," via the assertion that "the apparent motions . . . are the differences of true motions."[3]

The "apparent motions" form relational structures defined by measuring relative distances, time intervals, and angles of separation. For brevity, let us call these relational structures *appearances.* In the mathematical model provided by Newton's theory, bodies are located in Absolute Space, in which they have real or absolute motions. But within these models we can define structures that are meant to be exact reflections of those appearances and are, as Newton says, identifiable as differences between true motions. These structures, defined in terms of the relevant relations between absolute locations and absolute times, which are the appropriate parts of Newton's models, I shall call *motions,* borrowing Simon's term.[4]

When Newton claims empirical adequacy for his theory he is claiming that his theory has some model such that *all actual appearances are identifiable with* (*isomorphic to*) motions in that model.

Newton's theory goes a great deal further than this. It is part of his theory that there is such a thing as Absolute Space, that absolute motion is motion relative to Absolute Space, that absolute acceleration causes certain stresses and strains and thereby deformations in the appearances, and so on. He offered, in addition, the hypothesis (his term) that the center of gravity of the solar system is at rest in Absolute Space. But, as he himself noted, the appearances would be no different if that center were in any other state of constant absolute motion.

Let us call Newton's theory (mechanics and gravitation) *TN*, and *TN*(v) the theory *TN* plus the postulate that the center of gravity of the solar system has constant absolute velocity. By Newton's own account, he claims empirical adequacy for *TN*(0); and also claims that, if *TN*(0) is empirically adequate, then so are all the theories *TN*(v).

Recalling what it was to claim empirical adequacy, we see that all the theories *TN*(v) are empirically equivalent exactly *if all the motions in a model of TN*(v) *are isomorphic to motions in a model TN*($v + w$), for all constant velocities v and w. For now, let us agree that these theories are empirically equivalent, referring objections to a later section.

III

What exactly is the "empirical import" of *TN*(0)? Let us focus on a fictitious and anachronistic philosopher Leibniz*, whose only quarrel with Newton's theory is that he does not believe in the existence of Absolute Space. As a corollary, of course, he can attach no "physical significance" to statements about absolute motion. Leibniz* believes, like Newton, that *TN*(0) is empirically adequate; but not that it is true. For the sake of brevity, let us say that Leibniz* *accepts* the theory but that he does not *believe* it; when confusion threatens we may expand that idiom to say that he *accepts the theory as empirically adequate,* but does not *believe it to be true.* What does Leibniz* believe, then?

Leibniz* believes that *TN*(0) is empirically adequate, and hence, equivalently, that all the theories *TN*(v) are empirically adequate. Yet we cannot identify the theory that Leibniz* holds about the world—call it *TNE*—with the common part of all the theories *TN*(v). For each of the theories *TN*(v) has such consequences as that the earth has *some* absolute velocity, and that Absolute Space exists. In each model of each theory *TN*(v) there is to be found something other than motions, and there is the rub.

To believe a theory is to believe that one of its models correctly represents the world. A theory may have isomorphic models; that redundancy is easily removed. If it has been removed, then to believe the theory is to believe that exactly one of its models correctly represents the world. Therefore, if we believe of a family of theories that all are empirically adequate, but each goes beyond the phenomena, then we are still free to believe that each is false, and hence their common part is false. For that common part is phrasable as: one of the models of one of those theories correctly represents the world.

---

*To be presented in an APA symposium on Scientific Realism, December 28, 1976. Richard N. Boyd and Clark Glymour will comment; see this JOURNAL, this issue, pp. 633–635 and 635–637, respectively.

The research for this paper was supported by Canada Council Grant S74-0590. An earlier version was presented at the Western Division of the Canadian Philosophical Association (Calgary, October 1975).

I want to acknowledge my debt to Clark Glymour, Princeton University, for the challenge of his critiques of conventionalism in his dissertation and unpublished manuscripts.

Reprinted by permission from the author and *The Journal of Philosophy,* Vol. 73, No. 18, October 21, 1976.

It may be objected that theories will seem empirically equivalent only so long as we do not consider their possible extensions.[5] The equivalence may generally, or always, disappear when we consider their implications for some further domain of application. The usual example is Brownian motion; but this is imperfect, for it was known that phenomenological and statistical thermodynamics disagreed even on macroscopic phenomena over sufficiently long periods of time. But there is a good, *fictional* example: the combination of electromagnetism with mechanics, if we ignore the unexpected null results that led to the replacement of classical mechanics.

Maxwell's theory was not developed as part of mechanics, but it did have mechanical models. This follows from a result of Koenig, as detailed by Poincaré in the preface of his *Electricité et Optique* and elsewhere. But the theory had the strange new feature that velocity itself, not just its derivative, appears in the equations. A spate of thought experiments was designed to measure absolute velocity, the simplest perhaps that of Poincaré:

> Consider two electrified bodies; though they seem to us at rest, they are both carried along by the motion of the earth; . . . therefore, equivalent to two parallel currents of the same sense and these two currents should attract each other. In measuring this attraction, we shall measure the velocity of the earth; not its velocity in relation to the sun or the fixed stars, but its absolute velocity.[6]

The null outcome of all experiments of this sort led to the replacement of classical by relativistic mechanics. But let us imagine that values *were* found for the absolute velocities; specifically for that of the center of the solar system. Then, surely, one of the theories *TN(v) would be confirmed and the others falsified?*

This reasoning is spurious. Newton made apparent motions without presupposing more than that basic mechanics in which Maxwell's theories has models. Each motion in a model of $TN(v)$ is isomorphic to one in some model of $TN(v + w)$, for all constant velocities v and w. Could this assertion of empirical equivalence possibly be controverted by those nineteenth-century reflections? The answer is *no*. The thought experiment, we may imagine, confirmed the theory that added to TN the hypothesis:

HO. The center of gravity of the solar system is at absolute rest.
EO. Two electrified bodies moving with absolute velocity v attract each other with force $F(v)$.

This theory has a consequence strictly about appearances:

CON. Two electrified bodies, moving with velocity v relative to the center of gravity of the solar system attract each other with force $F(v)$.

However, the same consequence can be had by adding to TN the two alternative hypotheses:

Hw. The center of gravity of the solar system has absolute velocity w.
Ew. Two electrified bodies moving with absolute velocity $v + w$ attract each other with force $F(v)$.

More generally, for each theory $TN(v)$ there is an electromagnetic theory $E(v)$ such that $E(0)$ is Maxwell's and all the combined theories $TN(v)$ plus $E(v)$ are empirically equivalent.

There is no originality in this observation, of which Poincaré discusses the equivalent immediately after the passage I cited above. Only familiar examples, but rightly stated, are needed, it seems, to show the feasibility of concepts of empirical adequacy

paper I shall try to generalize these considerations, while showing that the attempts to explicate those concepts *syntactically* had to reduce them to absurdity.

V

The idea that theories may have hidden virtues by allowing successful extensions to new kinds of phenomena, is too pretty to be left. Nor is it a very new idea. In the first lecture of his *Cours de philosophie positive*, Comte referred to Fourier's theory of heat as showing the emptiness of the debate between partisans of calorific matter and kinetic theory. The illustrations of empirical equivalence have that regrettable tendency to date; calorifics lost. Federico Enriques seemed to place his finger on the exact reason when he wrote: "The hypotheses which are indifferent in the limited sphere of the actual theories acquire significance from the point of view of their possible extension."[7] To evaluate this suggestion, we must ask what exactly is an extension of a theory.

Suppose that experiments really had confirmed the combined theory $TN(0)$ plus $E(0)$. In that case mechanics would have won a *victory*. The claim that $TN(0)$ was empirically adequate would have been borne out by the facts. But such victorious extensions could never count for a theory as against one of its empirical equivalents.

Therefore, if Enriques' idea is to be correct, there must be another sort of extension, which is really a defeat—but qualified. For a theory T may have an easy or obvious modification which is empirically adequate, while another theory empirically equivalent to T does not. One example may be the superiority of Newton's celestial mechanics over the variant produced by Brian Ellis; Ellis himself seems to be of this opinion.[8] This is a *pragmatic* superiority and cannot suggest that theories, empirically equivalent in the sense explained, can nevertheless have different empirical import.

VI

We still need a general account of empirical adequacy and equivalence. It is here that the syntactic approach has most conspicuously failed. A theory was conceived as identifiable with the set of its theorems in a specified language. This language has a vocabulary, divided into two classes—the observational and theoretical terms. Let the first class be E; then the empirical import of theory T was said to be its subtheory T/E—those theorems expressible in that subvocabulary. T and T' were declared empirically equivalent if T/E was the same as T'/E.

Obvious questions were raised and settled. Craig showed that, under suitable conditions, T/E is axiomatizable in the vocabulary E. Logicians attached importance to questions about restricted vocabularies, and this was apparently enough to make philosophers think them important too. The distinction between observational and theoretical terms was more debatable, and some changed the division into "old" and "newly introduced" terms.[9] But all this is mistaken. Empirical import cannot be isolated in this syntactic fashion. If that could be done, then T/E would say exactly what T says about what is observable, and nothing else. But consider: the quantum theory, Copenhagen version, says that there are things which sometimes have a position in space and sometimes do not. This consequence I have just stated without using theoretical terms. Newton's theory TN implies that there is something (to wit, Absolute Space) which neither has a position nor occupies volume. As long as unobservable entities differ systematically from observable entities with respect to observable characteristics, T/E will say that there are such things if T does.

The reduced theory T/E is not a description of the observable part of the world of T; rather, it is a hobbled and hamstrung version of T's description of everything. Empirical equivalence fares as badly. In section II, $TN(0)$ and TNE *must* be empirically equivalent, but the above remark about TN shows that $TN(0)/E$ is not TNE/E. To eliminate such embarrassments, extensions of theories were considered in attempts to redefine empirical equivalence.[10] But these have similar absurd consequences.

The worst consequence of the syntactic approach was surely the way it focused philosophical attention on irrelevant technical questions. The expressions 'theoretical object' and 'observational predicate' mark category mistakes. Terms may be theoretical, but 'observable' classifies putative entities. Hence there cannot be a "theoretical/observable distinction." It is true surely that elimination of all theory-laden terms would leave no usable language; also that 'observable' is as vague as 'bald'. But these facts imply not at all that 'observable' marks an unreal distinction. It refers quite clearly to our limitations, the limits of observation, which are not incapacitating, but also not negligible.

VII

The phenomena are saved when they are exhibited as fragments of a larger unity. For that very reason it would be strange if scientific theories described the phenomena, the observable part, in different terms from the rest of the world they describe. And so an attempt to draw the conceptual line between phenomena and the transphenomenal by means of a distinction of vocabulary, must always have looked too simple to be good.

Not all philosophers who discussed unobservables, by any means, did so in terms of vocabulary. But there was a common assumption: that the distinction marked is philosophical. Hence it must be drawn, if at all, by philosophical analysis and, if attacked, by philosophical arguments. This attitude needs a Grand Reversal. If there are limits to observation, these are empirical, and must be described by empirical science. The classification marked by "observable" must be of entities in the world of science. And science, in giving content to the distinction, will reveal how much we believe when we accept it as empirically adequate.

A future Unified Science may detail the limits of observation exactly; meanwhile, extant theories are not silent on them. We saw Newton's delineation; for relativity theory, we have two revealing studies by Clark Glymour. The first shows that local (hence, I should think, measurable) quantities do not uniquely determine global features of space-time.[11] The second shows that these features also are not uniquely determined by structures each lying wholly within some absolute past cone—hence, I should think, by observable structures. It is the theory of relativity itself, after all, that places *an absolute* limit on the information we can gather, through the limiting function of the speed of light.

In the foundations of quantum mechanics much more attention has been given to measurement. Much of the discussion is about necessary limitations: the role of noise in amplification, the distinction between macro- and micro-observables.[12] Yet we have no such clarity as Glymour gave us for relativity theory, concerning the extent to which macro-structure determines micro-structure. The debate over scientific realism may at least have the virtue of directing attention to such questions.

Science itself distinguishes the observable that it postulates from the whole it postulates. The distinction, being in part a function of the limits science discloses on human observation, is anthropocentric. But, since science places human observers among the physical systems it means to describe, it also gives itself the task of describing anthropocentric distinctions. It is in this way that even the scientific realist must observe a distinction between the phenomena and the transphenomenal in the scientific world picture.

VIII

I have laid some philosophical misfortunes at the door of a mistaken orientation toward syntax. The alternative is to say that theories are presented directly by describing their models. But does this really introduce a new

element? When you give the theorems of T, you give the set of models of T—namely, all those structures which satisfy the theorems. And, if you give the models, you give at least the set of theorems of T—namely, all those sentences which are satisfied in all the models. Does it not follow that we can as advantageously identify T with its theorems as with its models?

But there is an ellipsis in the argument. It is being assumed that there is a specific language L which is the one language that belongs to T. And indeed, the theorems of T in L determine, and are determined by, the set of model structures of L (that is, structures in which L is interpreted) in which those theorems are satisfied. However, the assumption that there is a language L which plays this role for T places important restrictions on what the set of models of T can be like.

A theory provides, among other things, a specification (more or less complete) of the parts of its models that are to be direct images of the structures described in measurement reports. In the case of Newton's mechanics, I called those parts *motions;* in general, let us call them *empirical substructures.* The structures described in measurement reports we may continue to call *appearances.* A theory is *empirically adequate* exactly if all appearances are isomorphic to empirical substructures in at least one of its models. Theory T is *empirically no stronger* than theory T' exactly if, for each model M of T, there is a model M' of T' such that all empirical substructures of M are isomorphic to empirical substructures of M'. Theories T and T' are *empirically equivalent* exactly if neither is empirically stronger than the other. In that case, as an easy corollary, each is empirically adequate if and only if the other is.

In section V, I distinguished two kinds of extensions, the first a sort of victory and the second a sort of defeat. Let us call the first a *proper extension:* this simply narrows the class of models. We may call a theory *empirically minimal* if it is not empirically equivalent to any of its proper extensions. Glymour has convincingly argued, in the work cited above, that General Relativity is not empirically minimal. The reason is, in my present terms, that only local properties of space-time enter the descriptions of the appearances, but models may differ in global properties. This is a further nontrivial example of empirical equivalence.

The second sort of extension I shall not try to define precisely. The idea is that models of the theory may differ in structure other than that of the empirical substructures. In that case the theory is not empirically minimal, but this may put it in the advantageous position of offering modeling possibilities when radically new phenomena come to light. An example may yet be offered by hidden-variable theories in quantum mechanics.[13]

In terms of the concepts now at our disposal, and the examples given, we can conclude that there are indeed nontrivial cases of empirical equivalence, non-uniqueness, and extendability, both proper and improper. Such cases are now seen to be quite possible *even if the formulation of the theory has not a single term that cannot be called observational, in some way.* And now it should be possible to state the issue of scientific realism, which concerns our epistemic attitude toward theories rather than their internal structure.

All the results of measurements are not in; they are never all in. Therefore we cannot know what all the appearances are. We can say that a theory is empirically adequate, that all the appearances will fit (the empirical substructures of) its models. Though we cannot know this with certainty, we can reasonably believe it. All this is the case not only for empirical adequacy but for truth as well. Yet there are two distinct epistemic attitudes that can be taken: we can *accept* a theory (accept it as empirically adequate) or *believe* the theory (believe it to be true). We can take it to be the aim of science to produce a literally true story about the world, or simply to produce accounts that are empirically adequate. This is the issue of scientific realism versus its (divided) opposition. The intrascientific distinction between the observable and the unobservable is an anthropocentric distinction; but it is reasonable that the distinction should be drawn in terms of *us*, when it is a question of *our* attitudes toward theories.

NOTES

1. For some criticisms, see my "Theoretical Entities: The Five Ways," *Philosophia* 4 (1974): 95–109, and "Wilfrid Sellars on Scientific Realism," *Dialogue*, XIV, 4 (December 1975): 606–616.

2. Cf. my "The Only Necessity Is Verbal Necessity," forthcoming in this JOURNAL, LXXIV, 2 (February 1977).

3. F. Cajori, ed., *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World* (Berkeley: University of California Press, 1960), p. 12.

4. Herbert A. Simon, "The Axiomatization of Classical Mechanics," *Philosophy of Science*, XXI, 4 (October 1954): 340–343.

5. See, for example, Richard N. Boyd, "Realism, Undetermination, and a Causal Theory of Evidence," *Noûs*, VII, 1 (March 1973): 1–12.

6. Henri Poincaré, *The Value of Science*, B. Halsted, tr. (New York: Dover, 1958), p. 98.

7. *Historical Development of Logic*, J. Rosenthal, tr. (New York: Holt, 1929), p. 230.

8. "The Origins and Nature of Newton's Laws of Motion," in R. Colodny, ed., *Beyond the Edge of Certainty* (Englewood Cliffs, N.J.: Prentice-Hall, 1965), pp. 29–68.

9. For example, David Lewis, "How to Define Theoretical Terms," this JOURNAL, LXVII, 13 (July 9, 1970): 427–446. This paper is not subject to my criticisms here; on the contrary, it provides independent reasons to conclude that the empirical import of a theory cannot be syntactically isolated.

10. See fn 5 above. We could say that Boyd's paper, like Lewis's, provides independent evidence that empirical import cannot be syntactically isolated. But Boyd concludes also that there can be no distinction between truth and empirical adequacy for scientific theories.

11. "Cosmology, Convention, and the Closed Universe," *Synthese*, XXIV, ½ (July/August 1972): 195–218; discussed in my "Earman on the Causal Theory of Time," *op. cit.*, pp. 87–95 (referred to therein by an earlier title).

12. See, for example, N. D. Cartwright, "Superposition and Macroscopic Observation," *Synthese*, XXIX (December 1974): 229–242, and references therein.

13. See Stanley Gudder, "Hidden Variables in Quantum Mechanics Reconsidered," *Review of Modern Physics*, XL (1968): 229–231; and section III of my "Semantic Analysis of Quantum Logic," in C. A. Hooker, ed., *Contemporary Research in the Foundations and Philosophy of Quantum Theory* (Dordrecht: Reidel, 1973), pp. 80–113.

Nancy Cartwright

THE REALITY OF CAUSES IN A WORLD OF INSTRUMENTAL LAWS

INTRODUCTION

Empiricists are notoriously suspicious of causes. They have not been equally wary of laws. Hume set the tradition when he replaced causal facts with facts about generalizations. Modern empiricists do the same. But nowadays Hume's generalizations are the laws and equations of high-level scientific theories. On current accounts, there may be some question about where the laws of our fundamental theories get their necessity; but it is no question that these laws are the core of modern science. Bertrand Russell is well known for this view:

> The law of gravitation will illustrate what occurs in any exact science . . . Certain differential equations can be found, which hold at every instant for every particle of the system . . . But there is nothing that could be properly called 'cause' and nothing that could be properly called 'effect' in such a system.[1]

For Russell, causes 'though useful to daily life and in the infancy of a science, tend to be displaced by quite different laws as soon as a science is successful.'

It is convenient that Russell talks about physics, and that the laws he praises are its fundamental equations—Hamilton's equations or Schroedinger's, or the equations of general relativity. That is what I want to discuss too. But I hold just the reverse of Russell's view. I am in favour of causes and opposed to laws. I think that, given the way modern theories of mathematical physics work, it makes sense only to believe their causal claims and not their explanatory laws.

1. EXPLAINING BY CAUSES

Following Bromberger, Scriven, and others, we know that there are various things one can be doing in explaining. Two are of importance here: in explaining a phenomenon one can cite the causes of that phenomenon; or one can set the phenomenon in a general theoretical framework. The framework of modern physics is mathematical, and good explanations will generally allow us to make quite precise calculations about the phenomena we explain. Rene Thom remarks the difference between these two kinds of explanations, though he thinks that only the causes really explain: 'Descartes with his vortices, his hooked atoms, and the like explained everything and calculated nothing; Newton, with the inverse square of gravitation, calculated everything and explained nothing.'[2]

Unlike Thom, I am happy to call both explanation, so long as we do not illicitly attribute to theoretical explanation features that apply only to causal explanation. There is a tradition, since the time of Aristotle, of deliberately conflating the two. But I shall argue that they function quite differently in modern physics. If we accept Descartes's causal story as adequate, we must count his claims about hooked atoms and vortices true. But we do not use Newton's inverse square law as if it were either true or false.

One powerful argument speaks against my claim and for the truth of explanatory laws—the *argument from coincidence*. Those who take laws seriously tend to subscribe to what Gilbert Harman has called inference to the best explanation. They assume that the fact that a law *explains* provides evidence that the law is true. The more diverse the phenomena that it explains, the more likely it is to be true. It would be an absurd coincidence if a wide variety of different kinds of phenomena were all explained by a particular law, and yet were not in reality consequent from the law. Thus the argument from coincidence supports a good many of the inferences we make to best explanations.

The method of inference to the best explanation is subject to an important constraint, however—the requirement of nonredundancy. We can infer the truth of an explanation only if there are no alternatives that account in an equally satisfactory way for the phenomena. In physics nowadays, I shall argue, an acceptable causal story is supposed to satisfy this requirement. But exactly the opposite is the case with the specific equations and models that make up our theoretical explanations. There is redundancy of theoretical treatment, but not of causal account.

There is, I think, a simple reason for this: causes make their effects happen. We begin with a phenomenon which, relative to our other general beliefs, we think would not occur unless something peculiar brought it about. In physics we often mark this belief by labelling the phenomena as effects—the Sorbet effect, the Zeeman effect, the Hall effect. An effect needs something to bring it about, and the peculiar features of the

Reprinted by permission from the author and from *PSA 1980*, Vol. 2, P. Asquith and R. Giere, eds., 1981, pp. 38–48.