3. For an attempt at a theory of confirmation appropriate to this realist notion of equivalence see C. Glymour, *Theory and Evidence* (Princeton: Princeton University Press), esp. chaps. 5 and 9.

4. For a discussion of that notion of the meaning of theoretical terms which allows meaning to accrue to them over and above the role they play in the theory see the author's "Semantic Analogy," *Philosophical Studies* 38 (1980), 217–34.

5. On realistic theories as explaining by unifying see M. Friedman, "Explanation and Scientific Understanding," *Journal of Philosophy* 71 (1974), 5–19.



# *When Explanation Leads to Inference*

NANCY CARTWRIGHT
*Stanford University*

## I. *Introduction*

When can we infer the best explanation? This question divides scientific realists on the one hand, from operationalists, instrumentalists, positivists, and constructive empiricists on the other. There obviously must be certain provisions to ensure that "the best" is good enough. But once these are understood, the realist's answer to the question is "always"; the anti–realists's, "never." The realist asks, "How *could* something explain if it were not true?" The anti–realist thinks this question exposes a mistaken view about what we do in explaining. Explanations (at least the high level explanations of theoretical science which are the practical focus of the debate) organize, briefly and efficiently, the unwieldly, and perhaps unlearnable, mass of highly detailed knowledge that we have of the phenomena. But organizing power has nothing to do with truth.

I am going to discuss two anti–realists, Bas van Fraassen and Pierre Duhem. Van Fraassen's book, *The Scientific Image*,[1] provides a powerful and elegant defense of a brand of anti–realism which he calls "constructive empiricism." Duhem's views are laid out in his classic work of 1906, *The Aim and Structure of Physical Theory*.[2] According to van Fraassen the constructive empiricist maintains:

> Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate. (Basically, a theory is empirically adequate "exactly if what it says about the observable things and events in this world is true.")[3]

Van Fraassen presents the difference between the realist and the constructive empiricist as one of attitude. Both may explain by showing how the phenomena at hand can be derived from certain fundamental principles. But the two kinds of philosophers have opposing attitudes to the principles. The realist believes that they are true and genuinely give rise to the phenomena; the constructive empiricist believes merely that the principles are sufficient to derive the phenomena.

The realist, says van Fraassen, is making a mistake. When a theory succeeds in saving the phenomena, the scientific realist is ready to infer that its laws are true (or near true, or true for the nonce) and that its entities exist. Van Fraassen holds that a theory's success at saving the phenomena gives reason to believe *just that*: that it saves the phenomena, and nothing more. That the theory is true is a gratuitous additional assumption.

This is the core of Duhem's view as well. Duhem has no quarrel with phenomenological laws, which can be confirmed by inductive methods. What he opposes are theoretical laws, whose only ground is their ability to explain. Like van Fraassen, Duhem rejects theoretical laws because he does not countenance inference to the best explanation. Neither van Fraassen nor Duhem is opposed to ampliative inference in general. They make a specific and concrete attack on a particular kind of inference which they see as invalid—inference to the best explanation—and thereby on the scientific realism to which it gives rise.

This is the real interest of their view. They have a specific objection to one mode of reasoning and one class of scientific conclusions. They do not argue from a sweeping sceptical position that starts with the weakness of our senses and the poverty of our capacities and concludes that no one can ever know anything. Nor do they argue from a theory of meaning that counts theoretical talk as devoid of truth value along with all of our claims to morality, causality, and religion. Finally, they are not transcendental idealists like Kant. Nor (to use Ian Hacking's apt label) are they transcendental nominalists like Hilary Putnam, who argues that since thought can never connect with reality, our knowledge can achieve at best an internal coherence. Duhem and van Fraassen make a distinction within the field of scientific knowledge, while scepticism, positivism and the transcendentalisms are global doctrines about the whole domain of science. Duhem and van Fraassen allow that many inferences are sound, but not inferences to pure theory that are justified only in terms of explanation.

Their arguments are persuasive. But I think that van Fraassen and Duhem eliminate more than they should. I share their anti–realism about theoretical laws. On the other hand, I believe in theoretical entities, and that is my main topic in this paper. Arguments against inference to the best explanation do not work against the explanations that theoretical entities provide. These are causal explanations, and inference from effect to cause is legitimate. I will have nothing new to say about the structure of these inferences. I



aim only to show that we can be realists about theoretical entities on van Fraassen and Duhem's very own grounds. Many of the themes I discuss here are developed further in *How the Laws of Physics Lie* (forthcoming, Oxford University Press), where this essay will also be republished.

## II. *Van Fraassen's Attack*

Van Fraassen asks, "Why should I believe in theoretical entities?" There is a canonical answer: there are no genuine regularities at the phenomenological level. It is only among theoretical entities that science finds true regularities. Once we have laid these out, we have a powerful explanatory scheme. The exceptionless laws that we postulate at the theoretical level can explain not only why phenomena are as regular as they are, but why we see the exceptions we do. Van Fraassen grants this. But, he asks, what has explanatory power to do with truth? What reason do we have for inferring from the fact that a bundle of principles save the phenomena to the fact that they are true? We need some reason, some good reason, though certainly not a conclusive reason. Many arguments wear their validity on their sleeve: "I think. Therefore, I exist." But not, "*P* explains *Q*. *Q* is true. Therefore *P* is true." Van Fraassen has raised a crucial question.

This argument, and Duhem's as well, assumes that truth is an external characteristic of explanation; i.e., that something could satisfy all the other criteria for being an explanation and yet fail to be true. This is the way we are often taught to think of Ptolemaic astronomy. It might well form a completely satisfactory explanatory scheme, yet that does not settle the question of its truth. This, for instance, is what the medieval Piccolomini, one of Duhem's heroes, says of Ptolemy and his successors:

> . . . for these astronomers it was amply sufficient that their constructs save the appearances, that they allow for the reckoning of the movements of the heavenly bodies, their arrangements, and their place. Whether or not things really are as they envisage them—that question they leave to the philosophers of nature.[4]

Duhem's own argument against inference to the best explanation is the argument from redundancy: for any given set of phenomena, in principle there will always be more than one equally satisfactory explanation, and some of these explanations will be incompatible. Since not all of them can be true, it is clear that truth is independent of satisfactoriness for explanation. Sometimes Duhem's argument is

read epistemologically instead. He is taken to make a point, not about what the criteria are, but rather about our ability to see if they obtain. On the epistemological reading Duhem maintains merely that there will always be different laws which appear equally true, *so far as we can ever tell*, and yet are incompatible.

This, I think, is a mistaken reading. For it is a general feature of our knowledge and does not show what is peculiar to the inference–to–best–explanation which Duhem attacks. Duhem is not, for example, opposed to phenomenological laws, which arise by inductive generalization. It is a familiar fact that it is possible to construct different inductive rules which give rise to different generalizations from the same evidence. Here too there will always be more than one incompatible law which appears equally true so far as we can tell. These kinds of problems with inductive inference were known to Duhem. But he did not dwell on them. His concern was not with epistemological issues of this sort, but rather with the relationship between truth and explanation.

I said that Duhem and van Fraassen take truth to be an external characteristic to explanation. Here is an analogy. I ask you to tell me an interesting story, and you do so. What you tell me satisfies all the criteria for being interesting, but that says nothing about its truth. I may add that the story should be true. But if I do so, that is a new, additional requirement like the one which Piccolomini's natural philosophers make. They will call something a *genuine* explanation only if it does all else it should and is *in addition* true.

Let me recapitulate. Van Fraassen and Duhem challenge us to tell what is special about the explanatory relation. Why does the truth of the second relatum guarantee the truth of the first? Two–placed relations do not in general have that characteristic. Consider another two-placed relation:—is the paper at the 1980 Western Washington Conference that immediately preceded—. In that year my paper immediately preceded Richard Wollheim's. Wollheim's paper may well have been true. But that does not make mine true.

### III. *The Case For Theoretical Entities*

Van Fraassen and Duhem argue that explanation has truth going along with it only as an extra ingredient. But causal explanations have truth built into them. When I infer from an effect to a cause, I am asking what made the effect occur, what brought it about. No explanation of that sort explains at all unless it does present a cause; and in accepting such an explanation, I am accepting not only that it explains in the sense of organizing and making plain, but also that it



presents me with a cause. My newly planted lemon tree is sick, the leaves yellowing and dropping off. I finally explain this by saying that water has accumulated in the base of the planter: the water is the cause of the disease. I drill a hole in the base of the oak barrel where the lemon tree lives, and foul water flows out. That was the cause. Before I had drilled the hole, I could still give the explanation and to give that explanation was to present the supposed cause—the water. There must *be* such water for the explanation to be correct. An explanation of an effect by a cause has an existential component, not just an optional extra ingredient.

Likewise when I explain the change in rate of fall of a light droplet in an electric field, by asserting that there are positrons or electrons on the ball, I am inferring from effect to cause, and the explanation has no sense at all without the direct implication that there are electrons or positrons on the ball. Here there is no drilling a hole to let the electrons gush out before our eyes. But there is the generation of other effects: if the ball is negatively charged, I spray it with a positron emitter and thereby change the rate of fall of the ball: positrons from the emitter wipe out the electrons on the ball. What I invoke in completing such an explanation are not even laws of nature, but rather properties of electrons and positrons. I infer to the best explanation, but only in a derivative way: I infer to the most probable cause, and that cause is a specific item, what we call a theoretical entity. But note that the electron is not an entity of any particular theory. In a related context van Fraassen asks if it is the Bohr electron, the Rutherford electron, the Lorenz electron or what. The answer is, it is the electron, about which we have a large number of incomplete and sometimes conflicting theories.

Indeed I should use an example of van Fraassen's here to show how we differ. In a cloud chamber we see certain tracks which he says have roughly the same physical explanation as the vapor trail from a jet in the sky. In each case I may explain the trail by stating some laws. But what about the entities? I say that the most probable cause of the track in the chamber is a particle, and as I find out more, I can even tell you with some specificity what kind of particle. That, argues van Fraassen, is still quite different from the jet in the sky. For, there, van Fraassen says, look at the speck just ahead of the trail, or here, use this powerful glass to spy it out. There is no such spying out when we get to the cloud chamber. I agree to that premise but not to the conclusion. In explaining the track by the particle, I am saying that the particle causes the track, and that explanation, or inference to the most probable cause, has no sense

unless one is asserting that the particle in motion brings about, causes, makes, produces, that very track.

The particle in the cloud chamber is just one example. In general our belief in theoretical entities is founded on inferences from concrete effects to concrete causes. We look at the detailed structure of the effects and figure out, as best we can, exactly what could have brought them about given our knowledge of the circumstances of production. Here there is an answer to the van Fraassen-Duhem question. What is special about explanation by theoretical entity is that it is causal explanation, and existence is an internal characteristic of causal claims. There is nothing similar for theoretical laws.

Van Fraassen does not believe in causes. He takes the whole causal rubric to be a fiction. That is irrelevant here. Since van Fraassen does not believe in causes, he will not want to give causal explanations. We may have doubts about some particular causal claims, or, like van Fraassen, about the whole enterprise of giving causal explanations. These doubts bear only on how satisfactory we will count a causal explanation. They do not bear on what kind of inferences we can make from the success of that explanation.

We can see this point by contrasting causal explanation with the explanation of one law by another, or with the "preceding paper" relation I mentioned above. We need to sort the special van Fraassen–Duhem challenge we have been discussing from more general epistemological worries that make us question (as perhaps we always should) whether we have really got a good explanation. So let us introduce a fiction. God may tell you that Wollheim's paper is after mine, and that his paper is true. You have no doubts about either of those propositions. But this signifies nothing about the truth of my paper. Similarly, God tells you that Schroedinger's equation provides a completely satisfactory derivation of the phenomenological law of radioactive decay. You have no doubt that the derivation is correct. But you still have no reason to believe in Schroedinger's equation. On the other hand, if God tells you that the rotting of the roots is a completely satisfactory (causal) explanation of the yellowing of the leaves, or that the ionization produced by the negative charge explains the track in the cloud chamber, then you do have reason, conclusive reason, to believe that there is water in the tub and that there is an electron in the chamber.

## IV. *An Objection*

I argue that inferences to the most likely cause have a different



logical force than inferences to the best explanation. Larry Laudan raises a serious objection: "It seems to me that your distinctions are plausible only because you insist (apparently arbitrarily) on countenancing a pragmatic view of theoretical laws and a non–pragmatic view of causal talk."[5] In order to explain why I think this distinction is not arbitary, I will lay out two very familiar views of explanation, one that underlies the deductive–nomological model,[6] and the second, the view of Duhem. Van Fraassen challenges the realist to give an account of explanation that shows *why* the success of the explanation, coupled with the truth of the explanans, argues for the truth of the explanandum. I said there was an answer to this question in the case of causal inference. Similarly, I think there is an answer in the D.N. account. If the D.N. model is a correct account of what explanation is like, I agree that my distinction is arbitrary; but this is not so if Duhem is right.

If we could imagine that our explanatory laws *made* the phenomenological laws true, that would meet van Fraassen's challenge. But there is another, more plausible account that would do just as well. Adolf Grünbaum gives a brief sketch of the view:

> It is crucial to realize that while (a more comprehensive law) *G* entails (a less comprehensive law) *L* logically, thereby providing an explanation of *L*, *G* is *not* the "cause" of *L*. More specifically, laws are explained *not* by showing the regularities they affirm to be products of the operation of *causes* but rather by recognizing their truth to be special cases of more comprehensive truths (Ital. original)[7].

For any specific situation the fundamental laws are supposed to make the same claims as the more concrete phenomenological laws which they explain. This is borne out by the fact that the phenomenological laws can be deduced from the fundamental laws, once a description of the situation is supplied. If the phenomenological laws have got it right, then so too do the fundamental, at least in that situation. There is still an inductive problem: are the fundamental laws making the right generalization across situations? But at least we see why the success of the explanation requires the truth of the explanans. To explain a phenomenological law is to restate it, but in a sufficiently abstract and general way that states a variety of other phenomenological laws as well. Explanatory laws are true statements of what happens; but, unlike phenomenological laws, they are economical ways to say a lot.

This may seem straightforward. What else could explanation be?

But contrast Duhem. Duhem believes that phenomena in nature fall roughly into natural kinds. The realist looks for something that unifies the members of the natural kind, something they all have in common; but Duhem denies that there is anything. There is nothing more than the rough facts of nature that sometimes some things behave like others, and what happens to one is a clue to what the others will do. Explanations provide a scheme that allows us to make use of these clues. Light and electricity behave in similar ways, but the procedure for drawing the analogies is intricate and difficult. It is easier for us to postulate the electromagnetic field and Maxwell's four laws, to see both light and electricity as a manifestation of one single underlying feature. There is no such feature, but if we are careful, we are better off to work with these fictional unifiers than to try to comprehend the vast array of analogies and disanalogies directly. The explanatory schemes we posit work as well as they do, even to producing novel predictions, because phenomena do roughly fall into natural kinds. But in fact the phenomena are genuinely different. They only resemble each other some times in some ways, and the D.N. attempt to produce one true description for all the members of the same class must inevitably fail. We can not expect to find an explanatory law that will describe two phenomena which are in fact different, and yet will be true of both. What we can require of explanation is a scheme that allows us to exploit what similarities there are.

These are very cursory descriptions of the two views. But it is enough to see that the two embody quite different conceptions of explanation. Nor is it just a matter of choosing which to pursue, since they are joined to distinct metaphysical pictures. In practice the two conceptions meet; for in real-life explanations, failure of deductivity is the norm. Duhem predicts this. But proponents of the D.N. model can account for the practical facts as well. They attribute the failure of deductivity not to the lack of unity in nature, but to the failings of each particular theory we have to hand.

The difference between the two conceptions with respect to van Fraassen's challenge may be obscured by this practical convergence. We sometimes mistakenly assume that individual explanations, on either account, will look the same. Van Fraassen himself seems to suppose this; for he requires that the empirical substructures provided by a theory should be isomorphic to the true structures of the phenomena. But Duhem says that there can be at best a rough match. If Duhem is right, there will be no wealth of truly deductive explanations no matter how well developed a scientific discipline



we look to. Duhem sides with the thinkers who say "A physical theory is an abstract system whose aim is to *summarize* and to classify *logically* a group of experimental laws without aiming to explain these laws," where "to explain (explicate, *explicare*) is to strip reality of the appearances covering it like a veil, in order to see the bare reality itself."[8] In an effort to remain metaphysically neutral, we might take an account of explanation which is more general than either Duhem's or the D.N. story: to explain a collection of phenomenological laws is to give a physical theory of them, a physical theory in Duhem's sense, one that summarizes the laws and logically classifies them; only now we remain neutral as to whether we are also called upon to explain in the deeper sense of stripping away appearances. This is the general kind of account I have been supposing throughout this paper.

There is no doubt that we can explain in this sense. Physical theories abound, and we do not have to look to the future completion of science to argue that they are fairly successful at summarizing and organizing; that is what they patently do now. But this minimal, and non–question–begging, sense of explanation does not meet van Fraassen's challenge. There is nothing about successful organization that requires truth. The stripped down characterization will not do. We need the full paraphernalia of the D.N. account to get the necessary connection between truth and explanation. But going beyond the stripped down view to the full metaphysics involved in a D.N. account is just the issue in question.

There is still more to Laudan's criticism. Laudan himself has written a beautiful piece against inference to the best explanation.[9] The crux of his argument is this: it is a poor form of inference that repeatedly generates false conclusions. He remarks on case after case in the history of science where we now know our best explanations were false. Laudan argues that this problem plagues theoretical laws and theoretical entities equally. "What I want to know," he says of my view, "is what *epistemic* difference there is between the evidence we can have for a theoretical law (which you admit to be non–robust) and the evidence we can have for a theoretical entity—such that we are warranted in concluding that, say, electrons and protons exist, but that we are not entitled to conclude that theoretical laws are probably true. It seems to me that the two are probably on an equal footing epistemically." Laudan's favorite example is the electromagnetic aether, which "had all sorts of independent sources of support for it collected over a century and a half." He asks, "Did the enviable successes of one– and two–fluid

theories of electricity show that there really was an electrical fluid?"[10]

I have two remarks, the first very brief. Although the electromagnetic aether is one striking example, I think these cases are a lot rarer than Laudan does. So we have an historical dispute. The second remark bears on the first. I have been arguing that we must be committed to the existence of the cause if we are to accept a given causal account. The same is not true for counting a theoretical explanation good. The two claims get intertwined when we address the nontrivial and difficult question, when do we have reasonable grounds for counting a causal account acceptable? The fact that the causal hypotheses are part of a generally satisfactory explanatory theory is not enough, since success at organizing, predicting, and classifying is never an argument for truth. Here the idea of direct experimental testing is crucial. Consider the laser company, Spectra Physics. Engineers at Spectra Physics construct lasers with the aid of the quantum theory of radiation, non–linear optics, and the like; and they calculate their performance characteristics. But that will not satisfy their customers. To guarantee that they will get the effects they claim, they use up a quarter million dollars worth of lasers every few months in test runs.

I think there is no general theory, other than Mill's methods, for what we are doing in experimental testing; we manipulate the cause and look to see if the effects change in the appropriate manner. For specific causal claims there are different detailed methodologies. Ian Hacking, in this same volume ("Experimentation and Scientific Realism") gives a long example of the use of Stanford's PEGGY II to test for parity violations in weak neutral currents. There he makes a striking claim: "The experimentalist does not believe in electrons because, in the words retrieved from medieval science by Duhem, they 'save the phenomena.' On the contrary, we believe in them because we use them to *create* new phenomena, such as the phenomenon of parity violation in weak neutral current interactions."[11] I agree with Hacking that when we can manipulate our theoretical entities in fine and detailed ways to intervene in other processes, then we have the best evidence possible for our claims about what they can and cannot do; and theoretical entities that have been warranted by well tested causal claims like that are seldom discarded in the progress of science.

## V. *Conclusion*

I believe in theoretical entities. But not in theoretical laws. Often when I have tried to explain my views on theoretical laws, I have met with a standard realist response: "How *could* a law explain if it weren't true?" Van Fraassen and Duhem teach us to retort, "How could it explain if it *were* true?" What is it about explanation that guarantees truth? I think there is no plausible answer to this question when one law explains another. But when we reason about theoretical entities the situation is different. The reasoning is causal, and to accept the explanation is to admit the cause. There is water in the barrel of my lemon tree, or I have no explanation for its ailment, and if there are no electrons in the cloud chamber, I do not know why the tracks are there.



NOTES

1. Bas C. van Fraassen, *The Scientific Image* (Oxford: Clarendon Press, 1980).
2. See Pierre Duhem, *The Aim and Structure of Physical Theory*, trans. Philip P. Wiener (New York: Antheneum, 1962).
3. Van Fraassen, op. cit., p. 12.
4. Pierre Duhem, *To Save the Phenomena*, trans. Edmund Doland and Chanenah Maschler (Chicago: University of Chicago Press, 1969), p. 82.
5. In correspondence, dated 15 September 1981.
6. For a description of the deductive-nomological model of explanation, see C. G. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, N. J.: Prentice–Hall, Inc., 1966).
7. Adolf Grünbaum, "Science and Ideology," *The Scientific Monthly*, (July 1954), pp. 13–19.
8. Duhem, *The Aim and Structure of Physical Theory*, op. cit., p. 7.
9. Larry Laudan, "A Confutation of Convergent Realism," *Philosophy of Science*, 48 (March 1981), pp. 19–49.
10. In correspondence referred to in Note 5.
11. See Ian Hacking, "Experimentation and Scientific Realism," *Philosophical Topics*, Vol. 13, No. 1 (1982).