Data and Phenomena

Author(s): Jim Woodward

Source: *Synthese*, Vol. 79, No. 3 (Jun., 1989), pp. 393-472
Published by: Springer
Stable URL: http://www.jstor.org/stable/20116653

JIM WOODWARD

DATA AND PHENOMENA*

1.

My aim in this paper is to draw attention to a distinction and to a related set of procedures which are important to understanding science but which, until very recently, have been neglected or misdescribed by philosophers of science. The distinction in question is the distinction between data and phenomena and the procedures have to do with inferring claims about phenomena from claims about data. I shall begin with a general characterization of what I mean by data and phenomena. I shall then explore some of the implications of this characterization for a number of traditional issues concerning explanation, testing and theory-structure. Much of my discussion will draw on empirical case studies conducted by sociologists and historians of recent science. I shall try to suggest that many of the observations found in this literature about the openness, complexity, and imperfectly understood character of experimental systems, about the craft nature of experimental work, and about the different skills and preoccupations of experimentalists and theoreticians reflect real structural facts about testing, explanation, and theory-structure that deserve more attention from philosophers of science than they have hitherto received. In doing this, I shall also try to illustrate how these observations can be separated from some of the more grandiose conclusions about relativism and social construction with which they are often associated in the sociological literature.

Phenomena, as I shall use the term, are relatively stable and general features of the world which are potential objects of explanation and prediction by general theory. Examples of real or putative phenomena, some of which will be discussed in more detail below, include weak neutral currents, gravitational radiation, Brownian motion, proton decay, capacity limitations and recency effects in short term memory, and the proportionately higher rate of technical innovation among middle-sized firms in moderately concentrated industries. Data, by contrast, play the role of evidence for claims about

Synthese **79**: 393–472, 1989.

phenomena. As a rough approximation, data are what registers on a measurement or recording device in a form which is accessible to the human perceptual system, and to public inspection. As we shall see, data are typically not viewed as potential objects of explanation by or derivation from general theory; indeed, they typically are of no theoretical interest except insofar as they constitute evidence for the existence of phenomena. Examples of data which might play the role of evidence for some of the phenomena described immediately above include bubble chamber photographs (evidence for the existence of neutral currents), patterns of discharge in electronic particle detectors (proton decay), facts about reaction times and error rates in psychological experiments (effects associated with short-term memory), and statistics about patent and research and development expenditures in various industries (relative rates of technical innovation).[1]

One way of characterizing the contrast between data and phenomena is in terms of the notions of error applicable to each. In the case of data the notion of error, where realistically applicable at all, typically will involve perceptual or recording mistakes – misreading a dial or transposing digits when a number is entered into a laboratory notebook – or the outright manufacture of data, as in fraud. By contrast, in the case of claims about phenomena, much more complex and subtle kinds of error are possible – for example, failure to adequately control for various background and confounding factors or mistakes in statistical analysis or in procedures for data reduction. It is when one is concerned with claims about phenomena, rather than claims about data, that worries about whether one is detecting a real effect, rather than an artifact produced by peculiarities of one's instruments or detection procedures, become paramount. While arguments about the reality of phenomena sometimes turn on what might naturally be regarded as the possibility of perceptual error or the reliability of perceptual processes, they more commonly do not. Although I shall not press the point here, I think that one consequence of this is that discussions of the influence of theoretical preconceptions on perception of the sort one finds in writers like Kuhn (1970) and Hanson (1958) are of limited usefulness in understanding how one moves from claims about data to claims about phenomena and in understanding the characteristic kinds of mistakes which can affect such inferences.[2]

There is a second, not unrelated, contrast between data and phenomena which will be important for our subsequent discussion. Typically, a variety of different sorts of causal factors, many of them idiosyncratic to the details of a particular experimental arrangement or detection device, will play a role in the production of a given bit of data or in the assessment of its evidential significance. For example, the characteristics of the bubble chamber photographs used to detect neutral currents in an experiment at CERN described below (Section 3) will depend not just on the characteristics of neutral currents themselves but on the general causal principles underlying the operation of the chamber (the passage of a charged particle through the superheated liquid creates an ionized track along which boiling occurs), on the characteristics of the accelerator beam, on the dimensions of the chamber, on the liquid it contains, on the presence of various background factors, on the timing and details of the processes by which photographs are taken, and on many other factors as well. The quite different data produced in electronic particle detectors, which was also interpreted in a different experiment at NAL described below (Section 3) as evidence for neutral currents, will depend upon different but similarly idiosyncratic features of the apparatus and method of detection employed.

By contrast, phenomena are more widespread and less idiosyncratic, less closely tied to the details of a particular detection device or procedure. To describe something as a phenomenon is to suggest that it has stable recurrent features which can be produced regularly by some manageably small set of factors, which can themselves in principle occur in a wide variety of different situations. Thus, for example, neutral currents will occur (via the exchange of a Z^0 particle in reactions in which a neutrino interacts with a nucleon or in which a neutrino interacts with an electron) in a wide variety of different neutrino beam experiments and in many other situations as well. Because of this, one expects that a genuine phenomenon should be replicable or reproducible – it should be such that it recurs or can be made to recur in different situations or contexts.

For closely related reasons, one also expects that the same phenomenon should be detectable in a variety of apparent ways – via procedures and instruments that may produce quite a different data – and that its characteristics should not fluctuate greatly depending upon the particular detection technique employed. As the examples discussed below will illustrate, to claim that a certain measurement or experimental technique detects the presence of a phenomenon is to suggest the various extraneous causal influences which are unrelated to this phenomenon and which otherwise might be reflected in the data in a way which undermines the reliability of the technique have been eliminated, controlled, or corrected for. Indeed, if the technique successfully detects a phenomenon, the characteristics of that phenomenon ought to become more evident as we improve our method of detection – the phenomenon ought to be more readily identifiable and replicable as we control more sensitively for extraneous causal influences. If, as in the case of the alleged phenomena of parapsychology, a phenomenon becomes less readily detectable or reproducible as the sensitivity of one's measurement techniques is increased, and as novel candidates for error are controlled for, this will create the suspicion that the phenomenon in question does not exist.

It is often noted that scientists show little interest in the exact mechanical copying of an earlier experiment or measurement procedure.[3] Instead, scientifically interesting replications typically involve the use of more sensitive instruments and better experimental designs which may produce quite different data from that produced in the experiment being replicated. This reflects both the distinction between data and phenomena and that it is phenomena rather than data which are of intrinsic scientific interest and importance. It makes sense to think of a later, improved experiment as an attempt to replicate or reproduce the results of the earlier experiment, even though it may involve different techniques and produce different data, because it represents an attempt to detect or measure the same phenomenon. Phenomena thus give different experimental and detection techniques and different data a common interpretation.

There is another way of thinking about the distinction between data and phenomena which many scientists seem to find natural and suggestive. As many of the examples described below will illustrate, scientific investigation is typically carried on in a noisy environment; an environment in which the data we confront reflect the operation of many different causal factors, a number of which are due to the local, idiosyncratic features of the instruments we employ (including our senses) or the particular background situation in which we find ourselves. The problem of detecting a phenomenon is the problem of detecting a signal in this sea of noise, of identifying a relatively stable and invariant pattern of some simplicity and generality with recurrent features – a pattern which is not just an artifact of the particular detection techniques we employ or the local environment in which we operate. Problems of experimental design, of controlling for bias or error, of selecting appropriate techniques for measurement and of data analysis are, in effect, problems of tuning, of learning how to separate signal and noise in a reliable way.[4]

I suggested above that claims about phenomena are candidates for explanation by (or derivation from) general systematic theory, while claims about data typically are not. Thus, for example, the existence and some of the characteristics of neutral currents are predicted and explained by the Weinberg-Salam theory, which unifies the weak and electromagnetic forces, but the details of the bubble chamber photographs which constitute evidence for the existence of neutral currents are not explained, at least in anything like the same sense, by that theory or, as I shall argue below, by the conjunction of that theory with any so-called background theory. Similarly, the existence of gravitational radiation is predicted and explained by general relativity, but the data (which consisted of patterns of vibration in metal bars) produced by Joseph Weber's attempt to detect gravitational radiation, also described below, would not have been potential objects of theoretical explanation, even if Weber's attempt had been successful.

Underlying the distinction between data and phenomenon is the idea that the sophisticated investigator does not proceed by attempting to explain his data, which typically will reflect the presence of a great deal of noise. Rather, the sophisticated investigator first subjects his data to a great deal of analysis and processing, or alters his experimental design or detection technique, all in an effort to separate out the phenomenon of interest from extraneous background factors. It is this extracted signal rather than the data itself which is then regarded as a potential object of explanation by general theory. Thus, contrary to what many accounts of scientific explanation might lead one to expect, by no means everything that happens is a potential object of theoretical explanation. Figuring out what one should even try to explain – what the phenomena are in a given domain of inquiry – and what is mere noise is, as we shall see, an important aspect of scientific investigation, especially in relatively immature areas of inquiry like the social sciences.[5]

What matters in connection with the relation between data and phenomena is not that one be able to produce derivations or detailed causal explanations of the data but that the data should be *reliable evidence* for the phenomena in question. One of my central claims is that one can justifiably believe that data provide reliable evidence for some phenomenon without being in a position to explain or derive facts about the data and without understanding in detail the causal mechanisms by which the data are produced. As we shall see, the sorts of concerns which are relevant to reliability – control of potential confounding factors, elimination of background noise, and procedures for statistical analysis and data reduction, are typically not dealt with by the construction of detailed derivations or explanations. I thus deny at least one version of the common suggestion that inductive inference is always inference to the best explanation – that grounds for believing claim T on the basis of evidence E must always take the form of a demonstration that T figures in the best explanation of E.[6]

Showing all of this will require, among other things, a more detailed account of explanation and reliability. I shall devote the next section to a general characterization of scientific explanation and to an argument for the claim that facts about data are not, in the sense characterized, explained by theory. Section 3 will describe a particular case of phenomena-detection – the detection of weak neutral currents – in some detail. Section 4 will then discuss a number of considerations which are relevant to assessments of reliability and will show how a concern with reliability differs from a concern with explanation. As we shall see, most traditional accounts of theory-structure and theory-testing seems to cast little light on the considerations which are relevant to assessing whether data constitute reliable evidence for a claim about phenomena. Section 5 will consider a case (Joseph Weber's spurious detection of gravitational radiation) in which an experiment produced data which were not evidence for any phenomena of interest. Here, as we shall see, the claim that scientific theories are expected to explain claims about phenomena rather than data has considerable methodological bite. Sections 6 and 7 will take up some additional issues having to do with theory-structure and theory-testing. Section 8 will then connect my discussion with various claims in the sociology of science literature regarding the craft nature of experimental work and the openness of typical experimental systems.

2.

Good theoretical explanations in science have at least two features which could seldom be attained if the explananda of such explanations were facts about the data. The first feature has to do with the exhibition of patterns of dependency; the second with the role of unification and systematization in scientific explanation.

To begin with the first feature: as I shall use the notion of theoretical explanation, it is not a serious theoretical explanation of some outcome merely to assert that there is some unspecified causal mechanism which produces the outcome, or to assert that it is somehow produced via the interaction of various processes or entities, where the principles governing such interactions are left unspecified. For example, it is not a satisfactory theoretical explanation of the ideal gas law (although it is true) to say that it obtains because of facts about the interaction of the constituent molecules of gases. A satisfactory theoretical explanation instead requires a detailed exhibition of how features of the explanandum-phenomenon systematically depend upon factors invoked in the explanans. Often, but by no means always, this is accomplished by the derivation of the explanandum or some approximation to it from premises which include laws or generalizations of considerable scope. In highly mathematical sciences like physics, such derivations typically involve the solution of a system of differential equations, given certain initial and boundary conditions. By providing such a derivation one shows how the facts about the phenomenon to be explained actually do depend in a regular and systematic way on the factors appealed to in the explanans. For example, to provide the beginnings of a theoretical explanation of the ideal gas law one must do something like what Maxwell and Boltzmann did – one must show how, given certain general assumptions about the initial conditions governing the constituent molecules of the gas (e.g., assumptions about the distribution of molecular velocities), the character of the laws governing molecular interaction (e.g., that the laws in question are those of Newtonian mechanics), one can derive (e.g., by solving the Boltzman transport equation) some approximation to the ideal gas law.[7]

In other areas of scientific investigation such as molecular biology, the generalizations and mathematical structure required for explicit derivations may be lacking. But even here, good theoretical explanations are required to supply detailed accounts of the causal mechanisms responsible for the facts to be explained. The purpose of some accounts is similar to that of the derivations described above: the detailed and systematic exhibition of dependency relations. For example, the mechanisms linking the structure of the normal hemoglobin molecule to various facts about the oxygen-carrying capacity of blood and the different mechanisms linking sickle-cell hemoglobin to resistance to malaria and to a reduction in the oxygen-carrying capacity of the blood are now well-understood.[8] An account of the operation of such mechanisms will not take the form of a system of differential equations, or of explicit derivations from laws of nature, as with many theories in physics, but it will show in a detailed way how various sorts of changes in the structure of the hemoglobin molecule will produce systematic changes in the oxygen-carrying capacity of the blood. By exhibiting such patterns of dependency, such an account provides explanations of the oxygen carrying capacity of both normal and sickle-cell blood.

A second feature of good theoretical explanations has to do with the way in which such explanations systematize and unify. As a number of philosophers have noted, one mark of a good scientific explanation is that it makes reference to factors, generalizations, or mechanisms which can figure in the explanation of a range of different phenomena.[9] At least in many cases, an important component of scientific understanding involves seeing how a number of apparently independent and unrelated phenomena result from the repeated operation of a small set of factors or processes. As Michael Friedman emphasizes in an influential paper (Friedman 1974), such a demonstration has explanatory import because it exhibits connections and because it reduces arbitrariness and contingency by reducing the number of independent assumptions on which a theory must rely.[10] While it seems apparent that the extent to which such unification can be expected depends on the particular scientific domain in which one is working (it is a more salient feature of, say, contemporary high energy physics than contemporary neurophysiology), it also seems uncontroversial that the drive toward unification is both a readily recognizable feature of a great deal of contemporary scientific theorizing and that its presence, at least to some degree and in some respects, is an important desideratum in scientific explanation.

Both of these features of good theoretical explanation data lead to the conclusion that scientific theories typically to not provide theoretical explanations of claims about data. Let us begin with the desideratum of generality. As we have already noted, data are the result not just of some small number of recurring general processes which might be of interest to theorists, but are also the result of the interplay of a great many other factors which are idiosyncratic to the particular situation under investigation (e.g., to the particular measuring instrument or experimental design employed). If our interest is in trying to explain data, then we need to realize that as we vary these specific local features of the apparatus, the method of detection, and so forth, the explanation we give of the data generated also must change, often in rather fundamental ways. For example, the explanation of the data generated by the attempts described below to detect the weak neutral current at NAL, which took the form of patterns of discharge in electronic particle detectors, will differ in many important ways from the explanation of the data deriving from the experiments conducted at CERN, which took the form of bubble chamber photographs. Because of this, even if one fully understands the causal processes that produce a range of different items of data and can carry out the needed calculations, it commonly will be impossible to produce a unified, systematic account of those items. Explanations of the data, to the extent they can be given at all, will lack generality and will be closely tailored to individual cases; they often will be enormously complex, and will rely heavily on ad hoc or after the fact assumptions.

I emphasized above that phenomenena are characterized by certain recurrent, invariant features which data often lack. To say that a certain experiment or measuring procedure has detected a phenomenon is to suggest that the data produced do not just reflect causal interactions which are idiosyncratic to the particular experimental procedures employed, and that the phenomenon in question exhibits stable characteristics which are detectable via several different procedures. We can see that this feature of phenomena is closely tied to the fact that they typically are potential candidates for unified theoretical explanation in a way in which data are not. In undertaking to explain phenomena rather than data we avoid having to tell an enormous number of independent localized, idiosyncratic causal stories which depend on the details of the experimental techniques employed – we can focus instead on what is constant and stable across different experimental and measurement techniques. This opens up the possibility of explaining a range of cases in terms of a few factors or general principles.

Second, we may note that attempting to explain phenomena, which are the results of the operation of some small number of repeatable factors of theoretical interest, facilitates derivability, and the systematic exhibition of dependency relations. Many philosophers of science still do not sufficiently appreciate how difficult it is to construct detailed derivations of the behavior of complex systems, or to trace in detail the interactions giving rise to such behavior, even given an understanding in principle of the behavior of their parts, taken in isolation.[11] In both classical and quantum mechanics, for example, the behavior of only a very small number of systems is describable by means of equations that are susceptible of exact, analytic solution and even available approximation techniques quickly run up against severe limits. When – as in the case of much data – an outcome is the result of the interaction of many different causal processes which must be described by quite different theories, difficulties deriving from computational intractabilities and from ignorance of the appropriate way to represent complex interactions will be very common, as we shall see below. Even if one possesses – as is not always the case – a general theoretical understanding of the various causal processes underlying the production of data, it is a mistake to suppose that one can always just conjoin the disparate theories describing these processes into a sort of super theory which automatically allows for the derivation of determinate conclusions about the data.[12] Undertaking to explain phenomena rather than data helps to reduce problems of mathematical representation and of computation to manageable proportions.

I conclude this section with a remark by way of qualification and elaboration. In arguing that theories typically provide theoretical explanations of phenomena rather than data, I do not mean to deny that information about the causal processes underlying the production of data or the operation of instruments is relevant – as one consideration among many – to the assessment of reliability, or that sufficiently large mistakes about the character of such processes can undermine claims about the reliability of data. When someone uses an optical microscope or a bubble chamber to obtain reliable evidence, it must be the case that the specimen being examined plays some causal role in the measurement result obtained and that there is a complex causal chain leading from the particle interaction in the bubble chamber to the production of the image of the photographic plate. If such causal connections did not obtain, then the evidence in question would not be reliable. Some philosophers may wish to say that this shows that we do, after all, have in these cases 'explanations' of data – that the microscopic image is 'explained' by the presence of the specimen, the photographic image by the presence of the particle, and so forth. I have no fundamental objection to this usage, as long as two caveats are kept in mind. The first is that there is an enormous difference between the sort of 'explanation' envisioned immediately above,, in which a fact about the data is 'explained' merely by the claim that the phenomenon is involved in some (largely unknown) causal processes leading to the data, and the sort of explicit and systematic explanation of phenomena provided by a good scientific theory which satisfies criteria like those described above. We need some vocabulary to mark this difference. If we use an undifferentiated notion of explanation to describe in general both the relationship between theory and phenomena and the relationship between phenomena and data, we run the risk of obscuring this difference.

Second, we need to take care that the suggestion that there is a watered-down sense of explanation in which data are explained in the above examples does not misleadingly direct attention away from what is really important in the relation between data and phenomena, which has to do with reliability rather than explanation. As we shall see in more detail below, in order for data to be reliable evidence for the existence of some phenomenon (and in order for an investigator to be justified in his assumptions about reliability), it is neither necessary nor sufficient that one possess a detailed explanation of the data in terms of causal processes leading to it from the phenomenon. It is not sufficient because, plainly, even when an instrument like a microscope or pan balance is defective, producing a distorted or artifactual image or a biased measurement, it can still be the case that there is an explanation of this misleading data assigning a causal role to the phenomenon one is trying to detect. Again, it is not sufficient because even where a phenomenon plays a causal role in the production of data, there may be so much (or so little) data or the data may be so infected with background noise, that given available techniques of data analysis and reduction extracting reliable information about the phenomenon of interest may be impossible. For example, although bubble chambers were used extensively in high energy physics experiments in the 1960s and 1970s, they were not regarded as useful and reliable detection devices in experiments designed to detect the W and Z particles conducted at CERN in 1983 because, although the passage of these particles through the chamber would play a causal role in the production of bubble chamber photographs which could serve as data, the production of the W and Z is extremely rare and the bubble chamber is unselective in what it records. Finding evidence for the existence of the particles would be literally a matter of locating one photograph among many millions – an impractical task. For this reason the experimentalists used electronic particle detectors instead.[13] Here again, we see a difference between a concern with causal explanation and a concern with reliability. It is very common to understand in principle how a phenomenon plays a causal role in the production of a certain body of data, without being in a position to extract reliable information from that data regarding the phenomenon in question.

Moreover, data can provide one with reliable information about some phenomenon (and one can be justified in believing this) even if one is ignorant of or quite mistaken about the character of the causal processes leading from the phenomenon to the data (and thus about the explanation of the data). The bubble chamber itself is an illustration of this. The chamber was invented in 1952–1953 by Donald Glaser and was used for particle detection from the mid-1950s onwards. Glaser and others who first worked on the chamber originally thought that the mechanism involved in bubble formation was electrostatic repulsion; it was several years later (1958) before a detailed correct theoretical explanation of the behavior of the chamber was given.[14] Bubble formation is instead due to heat deposited by the penetrating ionizing particle. Of course, this does not mean that no one was justified in making inferences from bubble-chamber data to conclusions about particle interactions prior to this time. Several additional cases are described below (Section 4) in which data provides reliable information about phenomena in the absence of any detailed knowledge of the causal processes underlying the production of the data.

3.

My discussion so far has been rather abstract. Before proceeding, it will be useful to have before our minds a concrete example of an experiment in which the reliability of an inference from data to phenomenon was at issue, and which illustrates a number of general claims made above. I have chosen an example which has been the subject of several recent detailed studies – the detection of weak neutral currents.[15] According to the dominant model of weak interactions in the 1960s, the Vector-Axial (V-A) model, the weak interactions are mediated solely by charged W particles (charged currents), the hypothetical carrier of the weak force. The Weinberg-Salam model, developed in the late 1960s, which unifies the electromagnetic and weak interactions, by contrast predicts the existence of 'neutral-currents' – weak interactions that are mediated by the neutral Z particle. When the latter model was shown to be renormalizable, the existence of the weak neutral current became a matter of intense scientific interest. The phenomenon was first detected in two independent experiments, carried out by a European group at CERN and an American group at NAL.

The experiment at CERN basially consisted in firing a neutrino beam into a huge bubble chamber called Gargamelle and then examining photographs of the result. The data from this experiment consisted of approximately 290,000 bubble chamber photographs, of which it was claimed approximately 100 were candidates for genuine neutral current events. For our purposes, one of the most important problems that arose in connection with the CERN experiment had to do with the neutron background. When an incoming neutrino strikes a nucleon in the bubble chamber, a charged current interaction will involve the production of charged particles like muons. By contrast, neutral currents involve the production of a neutrino when an incoming neutrino strikes a nucleon. While charged particles like muons leave tracks in bubble chambers, electrically neutral particles like neutrinos do not. Thus for both charged and neutral current interactions, a bubble chamber photograph will not show the incoming neutrino. Instead it will show a short shower of tracks left by the strongly interacting particles produced when a neutrino strikes a nucleon. However, a charged current process will exhibit, in addition to the shower, the long straight track of a high energy muon, while in a neutral current process the outgoing neutrino will leave no such track. The presence or absence of a muon is thus crucial to distinguishing neutral from charged current events.

The central interpretive difficulty raised by this feature of the experiment was the following: when neutrinos from the incoming beam strike the chamber and the surrounding apparatus, they produce a large but unknown number of neutrons. If one of these hits a neutron or proton in the bubble chamber, the resulting shower of hadrons will mimic a genuine neutral current event. No muon will be produced, and because it is chargeless, the neutron will leave no track.

To show that they had genuinely detected neutral currents, the CERN experimenters had to show that this 'neutron background' was not by itself large enough to account for all the apparent neutral current events. The magnitude of this background could not be derived with confidence from fundamental theoretical principles for several different reasons. First, the magnitude of the background depended on characteristics of the incoming neutrino beam and the surrounding equipment which were neither directly measurable nor fully understood. Second, such a derivation would have required a theory of the passage through matter of strongly interacting particles which was not available at the time of the experiment.

The difficult and controversial nature of the background problem is suggested by the frequent disagreements and changes of mind among participants in the experiment about the best way of dealing with the problem, and by different participants having found different kinds of evidence most persuasive in dealing with the problem.[16] One general line of attack was to attempt to estimate the size of the background using various Monte Carlo simulations and thermodynamic arguments. Here the idea was to establish, via the convergence of different assumptions and estimating procedures, an upper bound on the size of the background. Doing this helped to persuade many (but not all) of the experimentalists that the neutron background could not be causing all the candidates for neutral current events they were detecting. A second kind of consideration focused instead on the location of putative neutral current events within the chamber. For theoretical reasons, it was expected that neutron-induced events would occur more frequently near the walls of the chamber. When plausible candidates for neutral current events began turning up in the center of the chamber (and indeed relatively uniformly through most of its volume) this was an additional consideration in favor of the claim that the experimenters were detecting genuine neutral current events.

While this experiment was underway at CERN, a second attempt to detect neutral currents was being carried out at NAL by a group of investigators from Harvard, Pennsylvania, and Wisconsin – hereafter the HPW group. This experiment was also a neutrino beam experiment involving hadronic interactions, although the apparatus employed and the characteristic problems to which it gave rise differed from those at CERN. Here too, the experimenters faced an interpretive problem, and their failure to deal with it adequately caused them to change their minds twice about whether they had discovered neutral currents. The apparatus employed by the HPW investigators consisted of a front section (a calorimeter) in which hadron showers produced in neutrino interactions were detected and a rear section (a muon spectrometer) designed to detect associated muons. Unfortunately, because of the geometry of the apparatus it seemed quite possible that muons produced in the forward section in charged current interactions could escape at wide angles before reaching the rear detector, with the result that the charged current interactions would appear to be muonless or neutral current interactions.

The HPW group initially concluded, in part on the basis of estimates of the number of wide angle muons deriving from Monte Carlo simulations, that not all of the muonless events they were seeing were due to undetected muons – that is, that they had detected neutral currents. In reaching this conclusion they were influenced by their knowledge that the group at CERN was about to announce similar results. But several members of the group distrusted such simulations, and introduced a fateful modification in their apparatus in an attempt to improve its ability to detect possible escaping muons. In the original apparatus, the front and rear parts of the detector were separated by four feet of iron shielding. This was to prevent 'hadron punchthrough', for if the hadrons produced in the front portion penetrated to the rear detector, they would be registered as muons, with the result that neutral current events would appear as charged currents. Just as with the neutron background, the magnitude of this punchthrough could not be derived from fundamental theory, because this would have required a theory of the strong interactions.[17]

In an effort to detect more wide-angle muons, the HPW group moved the rear detector closer to the front detector, substituting a 13-inch steel shield for the previous four-foot shield, assuming that this would still be sufficient to prevent hadron punchthrough. When they did this, the neutral currents they previously thought they were detecting seemed to go away. Many members of the experiment concluded that neutral currents did not exist – a result which now caused great anxiety at CERN, which had already announced the discovery of neutral currents. In fact, unbeknownst to the experimenters, this change in the apparatus had the result that large numbers of hadrons were penetrating to the rear detector, causing neutral current events to go unregistered. It was some time, however, before a more careful and rigorous analysis of the punchthrough problem, and other considerations as well, convinced the experimenters of their mistake and led them to conclude, in agreement with the group at CERN, that neutral currents did after all exist. Peter Galison's summary nicely captures the complex considerations that eventually persuaded both groups of the reality of neutral currents.

In a certain limited sense, the neutral currents were 'there' from the start: both FNAL and CERN had photographs they would eventually present as evidence for weak neutral currents. The real work of the experiments, however, was for the collaborators to convince themselves that the photographs were significant and not an artifact induced by the apparatus or environment. What followed was almost a year and a half of a seemingly endless list of internal debates over the tracks and sparks, the acceptance, the efficiency, the neutron background, the muon spectrum, the neutrino flux, the beam purity, the through muons, the fiducial volumes, the cosmic rays, the neutral kaons, and the statistical significance of the results. (Galison 1983, p. 505)

This example illustrates a number of important features of data and phenomena. Note first the contrast between the electroweak gauge theory, from which the existence of neutral currents and indeed the ratio of neutral to charged currents, follows as the result of an explicit calculation, and the considerably less explicit and less well understood assumptions regarding the operation of the Gargamelle bubble chamber, the electronic spark chambers, and the surrounding apparatus. While the electroweak theory explains the existence of the weak neutral current (the phenomenon of interest), it is certainly false that the electroweak theory, even in conjunction with assumed background information about the operation of the bubble chamber, provides a serious, similarly detailed explanation of the extent of the neutron background, the characteristics of the bubble chamber photographs, the hadron punch, and so forth.[18] The absence of a theory of the strong interactions, the disagreements and misunderstandings among both groups about how their equipment was functioning, and the fact that different members of each group found different considerations most decisive in convincing them of the reality of neutral currents should make it apparent, if nothing else does, that the process of reaching this conclusion was very far from being a matter of quasi-algorithmic deduction from a generally accepted background theory.

To the extent that an inference involving data figures in the above examples, the direction of inference is not 'downward' from theory or phenomena to data, but rather 'upward' from data to phenomena – from, e.g., facts about bubble chamber photographs to claims about the trajectories, energy, momentum, change, and so forth of particles traversing the bubble-chamber. Moreover, as is perhaps already apparent from the above description, and as I shall argue in more detail below, it is doubtful that it is going to be very illuminating to think of this inference from data to phenomena as simply the carrying out of a derivation. Much of what is involved is more helpfully viewed as data reduction (how does one find the relevant information in 290,000 bubble chamber photographs?), as curve-fitting (fitting various possible particle trajectories to tracks in photographs), as statistical analysis, and as the development of various procedures for the control of possible confounding factors (e.g., the neutron background) and various other sources of error (e.g., possible misidentification of photographs by human observers.)

The reliability of these procedures certainly depends upon complex empirical considerations, but there are many aspects, both of the procedures themselves and the assessment of their reliability, that do not seem to involve the construction of detailed explanatory derivations. Achieving control over possible confounding factors, for example, is often a matter of physical manipulation of equipment. Thus the NAL experimenters attempted to deal with the hadron punchthrough problem by means of the introduction of shielding, exactly because detailed calculations from fundamental theory of the magnitude of the punchthrough could not be carried out. And even where calculation and derivation are involved, the relevant premises need not be supplied by a detailed causal theory of the sort that could be used to furnish an explanation. Thus, in the case of the neutron background, although calculations were performed, what was of interest was not theoretical explanation – the investigators were in no position to supply such an explanation even if they wanted to, because they did not fully understand the processes involved in the production of the neutrons and their passage through matter. Rather, what was of interest was simply whether one could exclude the serious possibility that this background might be large enough to account for all of the candidates for weak neutral current events.

Andrew Pickering's description of the attitude of the scientists involved in the CERN experiments toward the problem of the neutron background captures this point very clearly.

> ... as far as theorists were concerned, calculations of such quantities as the neutron background were of no interest. The details of, say, the interactions between neutrons and nuclear matter were a 'dead' area of particle theory – they engaged no active stream of theoretical practice. As long as the inputs to the experimenter's Monte Carlo's were 'reasonable' theorists had no reason to question them – if, that is, they paid them any attention at all.
>
> However, with regard to the weak neutral current the situation was quite different: this phenomenon engaged with the practice of the growing band of gauge theorists very directly .... (Pickering 1984, p. 110)

Here again we see the idea that what is of scientific interest is the extraction of a certain signal from very noisy data which are of no intrinsic interest in themselves (or rather, are worthy of attention only insofar as they are evidence for the signal).

4.

My discussion so far has centered on a general intuitive contrast between a concern with reliability of evidence and a concern with explanation. In this section I want to describe in somewhat more detail some characteristic considerations that are relevant to reliability: control of possible confounding effects and systematic error, replicability, problems of data reduction and statistical analysis, and calibration and empirical investigation of equipment. The considerations on this list are certainly not meant to be exhaustive: my intent is, rather, to illustrate the difference between establishing that data are reliable evidence for some phenomenon and establishing explanations of the data.

4.1. *Control of Possible Confounding Factors*

One characteristic concern in experimental design is eliminating or adjusting for possible confounding factors – factors which might produce data which are similar to those which would be produced by the phenomenon of interest and which are thus spurious candidates for that phenomenon. This category also includes factors which may so interact or interfere with data which is possible evidence for the phenomenon of interest that detection of the phenomenon against the background of such noise becomes impossible. The problems associated with the neutron background or hadron punchthrough in the weak neutral current experiments are one example of this. There are many different ways of dealing with the problem of controlling for confounding factors. One obvious device is physical isolation. For example, in experiments to detect the possible decay of the proton (LoSecco, Reines and Sinclair, 1985), events having some of the characteristics of proton decay can be produced by cosmic rays. Because of this, such experiments generally have been conducted in abandoned mining shafts, thousands of meters below the surface of the earth which provide shielding from many such rays.[19]

Another device for dealing with confounding factors, which provides considerable opportunities for the experimenter's ingenuity, is to find processes of production and detection which are affected in a regular, characteristic way by the phenomenon of interest but not by confounders – processes which lead to the production of a characteristic 'signature' in the data. Here the concern is not with whether the existence of some phenomenon or some general theoretical claim is, or is part of, a sufficient condition for the obtaining of some pattern in the data (the consideration which would presumably be of interest if we wish to derive claims about the data from theory), but rather with whether it is (something like) a necessary condition. For example, in cosmic ray experiments designed to detect magnetic monopoles, one important problem was to distinguish heavy charged particles like possible monopoles from light nuclei. Ordinary photographic emulsions are sensitive to both kinds of particles, but detectors consisting of sheets of Lexan – a commercial plastic – are sensitive only to the former particles (Pickering 1981a). Another important concern is to distinguish genuine monopoles from other heavy charged particles such as heavy nuclei, which can mimic many of their effects. One relevant consideration in doing this is that genuine monopoles should produce ionization at a rate which is independent of their velocities, while ordinary charged particles should produce more ionization as the particles slow down at the end of the track (Kragh 1981).

Consider another example. Suppose that one wishes to distinguish K^+ mesons with a momentum of 1 Gev/c from protons and Π^+ mesons at the same momentum. One way of doing so makes use of a phenomenon known as Cerenkov radiation: if a particle passes through a medium at a velocity which is greater than that of light in the medium, it will emit electromagnetic radiation, which can then be detected by photomultiplier tubes. By using different counters filled with different materials, one can distinguish particles of different velocities and hence identify different particles. For example, in the above case, one could arrange two Cerenkov counters, one filled with water and the other with the carbon dioxide at an appropriate temperature and pressure in such a way that the passage of a proton will not cause Cerenkov radiation to be emitted in either counter, the kaons will trigger the water counter, and the pions will trigger both counters. Thus, each particle can be made to produce a signature which distinguishes it from the others.[20] Although this procedure certainly makes use of theoretical knowledge, this knowledge figures in the design of the experimental arrangement itself, rather than in the construction of detailed explanations of data for the experiment. Reliability in the discrimination of different phenomena is not accomplished by deriving facts about the data but rather by exploring convenient physical features of the experimental apparatus – it is built in to the apparatus rather than achieved by calculation.[21]

Another procedure which can sometimes be used to deal with background or confounding factors, even when one does not have an accurate estimation (let alone a detailed theory) of the extent of the operation of such factors makes use of information one may have about whether such factors are is likely to operate uniformly over some spatio-temporal region of interest or instead are likely to be subject to unsystematic local variations. If it is plausible to assume (or to design one's experiment so that) a background factor operates uniformly, one can then look for similarities or differences produced by the presence or absence of the phenomenon of interest, even if one doesn't know how to calculate or measure the background in question.

Randomized experimental designs represent one characteristic application of this strategy. This strategy was also employed, for example, in an experiment performed in 1928 by Fox, McIlwraith, and Kurrelmeyer which may have produced evidence for parity nonconservation. This experiment, which has been extensively discussed by Alan Franklin (1986), involved looking for directional asymmetries in the double scattering of electrons from a radium source. Photoelectrons ejected from the apparatus by x-rays emitted by the radium constituted a major background problem. The authors write that while this background could not be entirely eliminated by the experimental design, "there is no reason to expect that [the number of photoelectrons ejected] would vary between the two settings [i.e., directional measurements] at which counts were made".[22] Rather than correcting for the background by directly calculating its magnitude, the experimenters attempt to correct for its influence by exploiting the likelihood of its operating uniformly at all settings.[23]

Similarly, if it is plausible to assume that background or confounding factors operate non-uniformly, one can often exploit this in detecting their influence, even in the absence of a detailed understanding of their operation. This sort of strategy was employed by Joseph Weber in an attempt to detect gravitational radiation, which will be discussed in more detail below. In an effort to separate out spurious events due to background noise of various kinds, Weber employed two detectors separated by a large spatial distance – the idea being that genuine gravitational radiation, which would be cosmological in origin, should register simultaneously on both detectors while other sorts of background events which were more local in origin were at least unlikely to do this.[24] Obviously, the virtue of this sort of strategy is precisely that it doesn't require that one identify the causes of particular background disturbances and correct for each individually – rather, the elimination of large numbers of disturbances which are local in origin is accomplished by the physical design of the apparatus.

Finally, yet another procedure for control is to actually estimate the size of the background, and to determine whether the effect of interest seems to remain after a correction for the background has been made. This, of course, was one of the procedures employed in the weak neutral current experiments. It is important to realize that even doing this need not involve producing a theoretical explanation or derivation from fundamental theory of the size of the background. For one thing it is often sufficient, as it was in the neutral current experiments, to set a rough upper limit on the size of the background – to show that not all data of a certain kind can be produced by the background, rather than to produce an exact calculation of its magnitude. As the neutral current experiments suggest, confidence in such an upper limit can be reasonably produced by convergence of various estimating procedures, all of which fall well short of a derivation from fundamental theory.

We can put all of this in a more general way: when an investigator is concerned with the control of possible confounding factors, he typically has the conception (which may be quite vague and open ended) of a number of possible causes which may produce data like that which would be associated with the phenomenon of interest or which may produce data which is so complex and noisy that it is no longer useful as evidence. The problem the investigator faces is not one (or at least primarily one) of developing a detailed explanatory theory of the operation of these possible causes or of explaining the patterns in the data they produce. It is rather one of figuring out how to *control* for their influence: the investigator needs to eliminate or at least render unlikely the possibility that the data is being produced by any of these alternative causes. While possession of an explanatory theory of one's apparatus can of course play a role in this, it is neither necessary nor sufficient for successful control. Even when such control involves reliance on complex calculations, the calculations will often involve simulations or statistical techniques, rather than explanatory theory. And often techniques of control do not involve calculation at all – for example, they are instead a matter of physically isolating an apparatus, or of discovering physical processes which are sensitive to the phenomenon of interest, but not to confounders.

Both this point and the sheer difficulty, emphasized above, of constructing adequate explanations in the case of many of the complex interactions involved in the production of data are illustrated by some remarks made by one of the younger experimenters participating in the neutral current experiment at NAL described above. He remarks, regarding the hadron punchthrough problem, that

> It is very difficult, even today to understand the propagation of hadronic showers through matter. The thing is very complicated with lots of pions and stuff that you don't really know how to model. . . .
>
> But smart people don't put themselves into a situation where they have to understand something which is not understandable. The reason the original [shielding] was four one meter thick pieces of iron was just to avoid this problem from day one and never have to calculate what happens in the middle of the iron, because you never looked in the middle. (quoted in Crease and Mann, 1986, p. 356)

4.2. *Replication*

Another consideration which is important in establishing reliability is replication of repetition. As we have already noted, a genuine phenomenon ought to be detectable in different circumstances and by means of different physical processes – this helps to insure that the data which one takes to be evidence for that phenomenon are not just the result of the operation of causal factors which are idiosyncratic to a particular apparatus or to local circumstances and that one is seeing a "real effect" and not an artifact. In the weak neutral current experiments, two different groups, using different apparatuses, methods of detection, and procedures for data analysis, obtained (apparent) evidence for the existence of the same phenomenon played an important, although tangled, role in convincing both groups that they had detected a real phenomenon, and that the procedures they employed were defensible. Similarly, in the well-known case of the discovery of the J/Psi particle, Samuel Ting at Brookhaven accumulated, over a period of months, a considerable amount of data which strongly suggested the existence of a particle at 3.1 Gev,[25] but being a very cautious man wished to do additional tests and data analysis to eliminate the possibility of artifacts before making a public announcement. But when a group at SLAC, in an independent experiment using a different kind of accelerator, similarly found a 'bump' at 3.1 Gev, both groups announced immediately and simultaneously. While concerns about priority certainly played a role in the decision to announce, both groups apparently concluded that these two independent verifications eliminated any serious possibility that the phenomenon was spurious.[26]

Conversely, even if an experiment and data analysis claiming to detect the existence of some phenomenon is carried out in a prima facie plausible and convincing way, there will be considerable grounds for skepticism if other experiments which ought to have detected the same phenomenon fail to do so. For example, as we have already noted, the change of mind which led the investigators at NAL to mistakenly conclude (for a short period) that they were not detecting neutral currents, created considerable doubts regarding claims for their discovery at CERN.[27] Similarly, in the case of an experiment which Price and others claimed to have detected a magnetic monopole, one immediate ground for suspicion was that several other searches for the monopole – searches which ought to have been successful if monopoles having the claimed characteristics were detectable by means of Price's procedures – were negative. This prompted a re-analysis of Price's data and an alternative interpretation – that the event detected was probably a fragmenting heavy nucleus – came to be generally accepted.[28] In other cases it may be less clear exactly what has gone wrong or exactly what is producing the spurious data, but if there is systematic failure of replicability or repeatability, it will generally be assumed that some unknown confounding factor (from a long list of possibilities) is operative and that the data are not reliable evidence. Here too, there is an obvious difference between establishing an explanation and establishing reliability – facts about repeatability can support or undermine a claim about the reliability of data even if one is far from having a satisfactory explanation of how the data in question have been produced.[29]

A closely related point is that the pattern of results within a body of data produced by a single experiment or investigation can itself contain important information relevant to reliability. Thus, if one's data exhibits a great deal of variability or spread, or if it exhibits time trends or other patterns that seem to have no theoretical basis (if, for example, one's measured values for the mass of a new particle first drift upward and then downward and then upward again), this will immediately raise the suspicion that unknown sources of error are operative. Conversely, if one's data are consistent and exhibit little variability, this itself can be evidence (although far from decisive evidence) that the data are reliable evidence for some phenomenon. For example, as Allan Franklin points out (Franklin 1986, pp. 169, 220), one reason why Robert Millikan's measurement of the charge e of the election was so convincing is that Millikan claimed (as it now appears, not entirely truthfully) that on measurements of hundreds of drops of different substances, all of the measured charges and changes of charge were integral multiples of his postulated value for e. As Millikan contended, it is virtually unimaginable that this kind of consistency could be due to an experimental artifact.[30]

4.3. *Calibration*

Another strategy which is widely used in assessing reliability involves what Allan Franklin calls calibration (Franklin 1986, pp. 175ff). (One can also think of this as falling under the general heading of empirical investigation of equipment, discussed below, but it is sufficiently important and distinctive to deserve mention in its own right.) Often an apparatus or experimental technique successfully reproducing a phenomenon with known characteristics is itself an important piece of evidence that the apparatus and technique are functioning reliably and that various potential background and confounding effects have been adequately controlled. One can then take this fact to support the claim that new data produced by the apparatus or technique provides reliable evidence. This sort of strategy was employed, for example, in a well-known experiment designed by Ray Davis to detect the solar neutrino flux. The method of detection relied on interactions between the solar neutrinos and chlorine atoms. These interactions produced Ar^{37} atoms which were then collected and used to determine the magnitude of the neutrino flux. One worry that was raised in connection with the experiment concerned the reliability of the recovery procedure – perhaps some Ar^{37} atoms were not being recovered, with the result that the estimate of the neutrino flux was mistaken. One of the ways Davis attempted to deal with this worry was by releasing a known number of Ar^{37} atoms (roughly the same number that was expected to be formed from interactions due to solar neutrinos) into the apparatus, and then attempting to recover them. When Davis was able to recover these atoms with the expected efficiency, this was an important piece of evidence that his recovery procedure was functioning reliably. This in turn supported the claim that the apparatus was producing reliable data when it was used to detect solar neutrinos.[31]

A similar strategy is frequently employed in high-energy physics. As the neutral current experiments described above illustrate, experiments in high energy physics often involve extremely complex and delicate equipment which is susceptible to many sorts of malfunctions and confounding effects which may be difficult to predict or directly detect. In such circumstances, the fact that an experiment successfully produces and detects phenomena with known characteristics, such as an already discovered particle, is itself an important piece of evidence that the experiment is producing reliable data. For example, when experimenters at CERN who were searching for the Z boson started up their experiment after having shut it down for the Christmas holidays in 1983, they were faced with the problem of whether their equipment was functioning reliably. One important factor which helped to convince them that this was indeed the case was that they were able to use their apparatus to successfully detect W bosons; particles which they had discovered the previous year and that their data indicated that these particles had just the characteristics they were previously discovered to have (e.g., the correct mass, and so forth). When their experiment then also began turning up new data, which the experimenters took to be evidence for the Z boson for which they were searching, the successful performance of their equipment in connection with the familiar W bosons was an important piece of evidence that this new data was also reliable.[32]

4.4. *Statistical Analysis*

Another set of considerations which bear on reliability but are not naturally viewed as having to do with construction of explanatory theory, concern the choice of statistical procedures. Such issues arise in a variety of connections: what sort of significance level or standard error should be required and how should this be calculated?[33] How should data be combined or aggregated? Can certain data legitimately be discarded? One interesting and characteristic sort of problem has to do, roughly, with a choice between quality and quantity in data. For example, in the weak neutral current experiments, many theorists originally favored a test involving scattering a muon neutrino off an electron. Such a test would have the advantage that it is "extremely clean of background effects because no strong interactions were involved" (Galison 1983, p. 483). Unfortunately, such events would also have a quite low probability of occurring – only a few would be produced. Data from weak neutral current reactions involving hadrons, although much noisier, and much more subject to background effects would be much more plentiful and thus could be used to support more reliable conclusions and error estimates. It was in part for this reason that experimentalists at CERN and NAL elected to do experiments of the latter kind. Here a concern with reliability and a concern with explanation pull in opposite directions – an interest in reliability (getting enough data) leads to a choice of experimental design which involves processes the details of which are less well understood than a possible alternative design.

4.5. *Data Reduction*

Data reduction is another consideration which is quite crucial in data analysis and the establishment of reliability, but has very little to do with the construction of derivations involving explanatory theory.[34] Bubble chambers, for example, produce an enormous amount of 'visual' data in the form of photographs; they register everything that takes place within them. Typically only a very small portion of this data will be relevant to the existence of the phenomenon one is trying to detect. (Recall that out of 290,000 photographs, roughly 100 were considered by the CERN group as candidates for neutral current events.) If bubble chambers were to be useful and reliable sources of evidence, fast and effective procedures for systematically searching through very large amounts of data for events of potential theoretical interest had to be developed. In the absence of such procedures, the bubble chamber would be, in the words of one prominent experimentalist, "nothing but an expensive toy" (Galison 1985, p. 304 quoting Luis Alvarez).

Two widely used procedures for dealing with this 'data bottleneck' are the use of relatively untrained personnel in the early stages of data analysis and (increasingly) heavy reliance on automated data processing and curve fitting procedures. Commonly, bubble chamber photographs are first scanned by workers who have no background in physics, but have been taught to recognize certain potentially interesting patterns. Interesting photographs are then subjected to further analysis by relatively junior physicists or by computer programs, which automatically fit trajectories to the tracks in photographs, calculate possible masses and velocities, and make tentative particle identifications. Photographs which continue to appear interesting are then brought to the attention of other investigators. Obviously the characteristics of both personnel and procedures and the errors to which they are subject (that is, whether they are likely to overlook relevant photographs or misidentify as relevant photographs which are not) are quite important in assessing whether an experiment has produced reliable evidence. Attempts are thus made to investigate these characteristics empirically by, for example, calculating personal error rates and to organize the whole procedure so to maximize reliability in light of this information.

When electronic particle detectors are used instead of bubble chambers, one also faces a problem of data reduction, but here the problem is solved in a quite different way, which results in the production of data having quite apparent characteristics. While bubble chambers record information indiscriminately, regardless of whether or not it is of interest, modern electronic particle detectors are highly selective. They come equipped with very sophisticated programs which specify that only events having certain charactersitics will be recorded, and that other events will be discarded. For example, in experiments at CERN in 1983 in which W bosons were produced by means of proton – anti-proton collisions, it was recognized that it would be impossible to record and effectively analyse relevant information about all of the collision products. Because W boson were expected to have a characteristic decay signature in which two electrons appeared back to back with high transverse momentum, the detector contained a trigger which was designed to only record events with high transverse momentum. Here, in contrast to the bubble chamber, a significant portion of the procedures for data reduction are literally built into the detector itself and quite different sorts of checks are required to ascertain whether these procedures are working reliably. Both in the case of both visual and electronic particle detectors, although the procedures employed for data reduction are crucial to establishing that the data produced is reliable evidence, these procedures seem to have little to do with the construction of theoretical explanations, and indeed, seem not to be captured at all in traditional accounts of theory structure.[35]

4.6. *Empirical Investigation of Equipment*

It should also be noted that the reliability of an instrument can often be investigated empirically, even when a detailed explanatory theory of the operation of the instrument is unavailable, by noting whether variations in some phenomenon (detected independently) are correlated systematically with changes in the information provided by a supposed detecting instrument. This point has been emphasized by Ian Hacking in connection with microscopes (Hacking 1981) and is perhaps most strikingly illustrated by empirical investigations into the reliability of ordinary visual perception. The details of the operation of the human visual system and of how it detects information about the external world are just beginning to be understood (cf. Marr 1982).

We still do not and certainly did not in the past possess a full explanatory theory of the operation of the visual system. Yet it would be absurd to conclude that vision is never a reliable source of information about the external world. Indeed, without possessing a detailed explanatory theory of the operation of the visual system one can investigate empirically the circumstances under which vision is or is not a reliable source of information – a point of considerable importance in some experimental contexts.

A good example of this is provided by the controversy between Rutherford and Chadwick of Cambridge and Petterson of Vienna in the 1920s, as described in a recent paper by Roger Stuewer (Stuewer 1985). The Vienna group claimed, on the basis of experimental evidence they had obtained, that carbon atoms can be caused to disintegrate under bombardment by alpha particles and to emit protons, which then travel short distances in the air (so-called 'short range protons'). The Cambridge group, on the basis of evidence they obtained, denied this and suggested that particles the Vienna group identified as protons from disintegration were in fact just alpha particles from the source they were using.

Both groups relied on observations (using a low powered microscope) of scintillations – tiny flashes of light produced when a charged particle strikes a scintillation screen. Issues about the circumstances under which trained observers could reliably distinguish and count scintillations of varying brightness turned out to be quite crucial. By using different observers in tandem with sources of known characteristics, the Cambridge group was able both to estimate the efficiency of different observers in detecting scintillations (that is, personal error rates) and the circumstances under which such observations were accurate – e.g., counts above 80 or below 10 per minute tended to be inaccurate. This work convinced the Cambridge group that it was unlikely that human observers could reliably distinguish protons and alpha particles by the brightness of the flashes they produced – one of the central methods of differentiation on which the Vienna group was relying.[36]

That this was the case and that the counters used (three young women, not the primary experimentalists) were not reliable observers of scintillations under the experimental circumstances obtaining in Vienna was confirmed by Chadwick during a visit to Petterson's laboratory. Chadwick showed that the counters claimed to observe scintillations even when (unbeknowst to them) Chadwick removed the particle source or introduced absorbing material exceeding the calculated range of the source (Stuewer 1985, pp. 286–7). Investigations of this sort thus produced evidence bearing on the reliability of a certain kind of detecting device (the human eye), quite independently of detailed knowledge of how the human visual system worked.[37]

5.

The significance and implications of my claim that what is demanded of an explanatory theory is that it account for the phenomena rather than the data become clearer when we examine a case in which it was mistakenly claimed that certain data were evidence for the existence of a phenomenon. Here, as we shall see, the distinction between data and phenomena has considerable methodological bite. In experiments conducted in the late 1960s, Joseph Weber, an experimentalist at the University of Maryland, claimed to have successfully detected the phenomenon of gravitational radiation.[38] The production of gravity waves by massive moving bodies is predicted (and explained) by General Relativity, a theory in which most theorists have considerable confidence. Nonetheless, gravitational radiation is so weakly coupled to matter that detection of such radiation by us is extremely difficult.

Weber's apparatus initially consisted of a simple large metal bar which was designed to vibrate at the characteristic frequency of gravitational radiation emitted by relatively large scale cosmological events. The central problem of experimental design was that to detect gravitational radiation one had to be able to control or correct for other potential sources of disturbance – electromagnetic, thermal, acoustic, and so forth. In part, this was attempted by physical insulation of the bar, but this could not eliminate all possible sources of disturbance: for example, as long as the bar is above absolute zero, thermal motion of the atoms in the bar will induce random vibrations in it. As mentioned above, one of the ways Weber attempted to deal with this difficulty was through the use of a second detector which was spatially separated from his original detector. Since some coincident disturbances will occur in the two detectors just by chance, however, various complex statistical arguments and other kinds of checks were required to show that it was unlikely that all of the coincident disturbances could arise in this way.

While Weber claimed to have detected the existence of gravitational radiation from 1969 on, his claims are now almost universally doubted. In part, this is for theoretical reasons – given the sensitivity of his apparatus, Weber found far more radiation than is consistent with current cosmological theories. But this skepticism is also based in large measure on considerations having to do with deficiencies in Weber's experimental design and techniques of data analysis; deficiencies which in themselves led the majority of experimentalists to regard Weber's results as spurious.[39]

I shall briefly describe a few of these considerations, which nicely illustrate a number of the remarks about reliability made in the previous section. First, the statistical techniques used by Weber to analyze his data were thought to be problematic. As one anonymous scientist put it:

> We felt that the way he was doing his statistical analysis was open to great misinterpretation. By massaging data again and again, knowing what you want for an answer, you can increase the apparent statistical significance of any bump... I'm pretty sure he could get these out of pure noise. (Collins 1981, p. 40)

These suspicions were heightened when Weber claimed to detect evidence for gravitational radiation in data provided by another group which, because of a misunderstanding on Weber's part about synchronization, should have been reported as containing pure noise.

> Another important consideration, cited by a number of scientists, was Weber's failure to improve his signal to noise-ratio despite several years' intensive modification of his apparatus. The signal obstinately stayed just above the threshold (and had deteriorated).... (Collins 1981, p. 42)

As a number of writers have noted, this sort of result is often characteristic of 'pathological' or 'pseudo-science'.[40] A third, related difficulty was the failure of other experimentalists, using similar or more sensitive apparatus, to replicate Weber's results. As one scientist puts it:

> Once, say, two other groups have repeated the experiment, with greater sensitivity, and found nothing, you have to say, either 'all these people are incompetent to repeat this experiment', or 'the first person has made a mistake'. And it's a fairly easy choice. According to the rules of science, Weber has been disproved, *even though you can't necessarily find what it is or how he went wrong.* (Collins 1981, p. 42, my emphasis)

The general pattern illustrated by this example should by now be familiar. Note first that what is of lasting theoretical interest is not

Weber's data, which even if his experiment were successful, would reflect the influence of a great deal of noise or extraneous causal influences, but whether a phenomenon has been detected for which his data may be regarded as evidence. Evidence that Weber has successfully done this is *not* provided by constructing an explicit deduction which shows how the observed data result from General Relativity and a detailed theory of how Weber's instruments work – the possible sources of error are probably too numerous and complex for that, as the quotation immediately above indicates – but rather by the sorts of considerations embodied in the criticisms described above: control of various possible sources of error, adequacy of statistical analysis, replicability by others, and improvement of the signal with more sensitive apparatus. It is certainly true that Weber makes a causal claim which is mistaken (or at least for which there is no good evidence) – he thinks that some of the vibrations he observes are due to gravitational radiation when (as best we can tell) there is no reason to think that any are. But it is at best misleading to think of his failure as mainly a failure in the construction of an explanatory theory. His failure is rather a failure in connection with the considerations bearing on reliability described above.

The point that it is not Weber's data, but the gravitational radiation for which those data are (mistakenly) taken to be evidence which is the potential object of theoretical explanation is reflected in the interesting asymmetry exhibited in the last quotation above. If Weber's experiment had detected gravitational radiation, it would be a requirement on any theory of gravitation that it account for this phenomenon. However, there is no corresponding theoretical interest or urgency in accounting for Weber's *data*. It may very well be, as the remark quoted above suggests, that for some of Weber's data, the details of the processes that produced it will never be completely understood, but this does not matter, since we have other grounds (other than detailed knowledge of these processes) for believing that the experiment was unreliable.[41] If the point of doing science was to explain the data or to explain what was actually observed, this asymmetry would be difficult to understand. It would instead seem perfectly reasonable to demand that subsequent theories account for Weber's data, just as they must account for the data from any *reliable* experiments involving gravitational radiation. Distinguishing between data and phenomena and imposing the requirement that only phenomena must be theoretically explained allows us to avoid this methodologically pernicious result. We can say instead that until experiments like Weber's produce reliable evidence for gravitational radiation, there is nothing in the results of such experiments which stands in need of systematic theoretical explanation.[42]

6.

In the philosophical literature, issues having to do with the distinction between data and phenomena often arise in connection with discussions of theory-structure and correspondence rules. Some critical remarks on two standard views of these matters will help to orient the reader to what is distinctive about the position developed above. On the so-called received view of theory-structure defended, for example, by Ernest Nagel in *The Structure of Science*, (Nagel 1961) a theory consists of a set of claims stated in a 'theoretical' vocabulary, a set of claims stated in an 'observational' vocabulary and a set of correspondence rules which establish connections between these two sets of terms. No clear distinction is made within this framework between data and phenomena – both are lumped together as what lies at the observational end of correspondence rules. One immediate consequence of this omission is that two quite different potential uses of correspondence rules – that of (a) establishing connections between claims about phenomena and the claims of explanatory theory and that of (b) specifying procedures, experimental and otherwise, by which one moves from claims about data to claims about the phenomena – are systematically conflated.

This conflation is illustrated by Nagel's own discussion. Nagel's most extended example of a correspondence rule involves Bohr's early 'solar system' model of the atom which among other things explained certain facts about the spectrum of hydrogen. In connection with this example, Nagel suggests at one point that correspondence rules play the role of establishing a connection between the more theoretical features of Bohr's account (which have to do, e.g., with electron transitions between various orbits – so-called 'electron jumps') and various facts about line spectra which Bohr's theory explains. Nagel writes:

> The Bohr theory associates the wave length of a light ray emitted by an atom with the jump of an electron from one of its permissible orbits to another such orbit. In

consequence, the *theoretical* notion of an electron jump is linked to the *experimental* notion of a spectral line. Once this and other similar correspondences are introduced, the experimental laws concerning the series of lines occurring in the spectrum of an element can be deduced from the theoretical assumptions about the transitions of electrons from their permissible orbits. (Nagel 1961, p. 95)

Here the role of a correspondence rule is, in my terms, to connect theoretical claims with claims about phenomena. Moreover, as Nagel contends, the connection which is established is a deductive and explanatory connection: claims about spectral lines can be deduced from (and are systematically explained by) theoretical claims embodied in Bohr's theory.

However, at a number of other points in his discussion, Nagel seems (without realizing this) to understand the role of correspondence rules quite differently: they are in effect taken to prescribe procedures by which one might move from what I would call data to phenomena. Thus, Nagel suggests that the function of correspondence rules is similar to that ascribed to "operational definitions" (p. 43), or to Reichenbach's "coordinative definitions", or that such rules identify "observational procedures" (p. 94) or "laboratory procedures" (p. 95) which specify when "theoretical notions" are applicable, or that such rules describe how theory can be "brought into relation with what is observed in the laboratory" (p. 61). Nagel seems to have this sort of role for correspondence rules in mind when he tells us that sometimes correspondence rules "supply only a necessary condition for use of theoretical terms" and then gives the following example:

... under the experimental conditions obtaining in a Wilson cloud chamber, the condensation of water vapor in fine lines appears to be a necessary condition for describing this effect in terms of the theoretical notion of the passage of alpha particles. (Nagel 1961, p. 101)

The correspondence rule described in this example seems to have the role of specifying experimental evidence (data) or procedures which are relevant to detecting the passage of an alpha particle or in telling us a bit about how an instrument which detects such particles works. By contrast, the earlier rule connecting electron jumps and spectral lines told us nothing about how we should go about trying to measure or observe spectral lines or what instruments to use. While it is plausible that this earlier rule functions so as to permit the deduction of phenomenal claims about the position of spectral lines from Bohr's theory, no similar relationship seems to hold between the notion of a spectral line and the data or procedures used to measure such lines. The rules describing the data, instruments, or procedures used to measure the position of spectral lines do not establish deductive or explanatory connections. The conflation of these quite different notions of (conceptions of the function of) correspondence rules is a natural consequence of the failure to distinguish data and phenomena.

A second more fundamental difficulty with the traditional Nagelian account of theory-structure and correspondence rules is this: when a scientific theory is used to predict and explain various claims about phenomena and data is gathered and analyzed to see if it supports these claims, it is often inappropriate to think of either the data or the procedures for data-gathering and data-analysis as an integral part of, or as specified by, the explanatory theory under investigation. This point shows up in several different ways. First, while theorists will of course be interested in determining what claims about phenomena follow from their theories, in many areas of both the natural and social sciences, theorists will *not* specify the data, the experimental procedures, or the techniques of data analysis which can be used to test these claims about phenomena. In a number of disciplines, this is reflected in a sociological division of theorists and experimentalists into two distinct groups. For example, one finds such a division between theorists and experimentalists in many areas of physics. It is rare for a physicist to do both theoretical and experimental work – an indication of the different skills and sensitivities, as well as specific knowledge required for work in both areas.[43] When a theorist elaborates a new theory (electroweak gauge theory, general relativity) which predicts the existence of some new natural phenomenon (the weak neutral current, gravitational radiation), the question of how best to design an experiment to test this prediction and how to analyze the data produced by the experiment will typically not be prescribed, at least in detail, by the theorist, but rather will be left to the experimentalist. It will be the job of the experimentalist to determine what instruments or measurement techniques to employ, how to eliminate or control for various possible sources of error, and how to deal with the other characteristic concerns relevant to reliability described above. If theories consisted of correspondence rules connecting general theoretical claims with claims about data or if the ultimate object of the theorist was to explain facts about the data, one would expect the theorist to prescribe the relevant considerations having to do with measurement and experimental design in considerable detail – the theory would not be fully articulated until one did this. Instead, this is precisely what one does not find.

Again, if explanatory theories consisted of explicit correspondence rules connecting theoretical claims to claims about the data and if prescriptions of appropriate experimental procedures were part of such theories, one would not expect experimentalists to disagree about how it is appropriate to test such theories, or one would suppose that to the extent that there is such disagreement, this would amount to testing different theories. Here again, this is precisely what one does not find. As we have seen, given a physical theory which claims that certain natural phenomena such as gravitational waves or weak neutral currents exist, there may be considerable controversy about what procedures may appropriately be employed to detect those phenomena. Participants in such controversy do not think of themselves as testing different theories, but rather as disagreeing about what is required to sensitively or reliably detect the existence of weak neutral currents or gravitational waves. They disagree, that is to say, about such matters as experimental design, or about how to control for different kinds of systematic errors, or appropriate procedures for data-analysis. It is because such matters are not settled by the general theories that predict the existence of the above phenomena that this controversy is possible and there is a distinct set of skills and preoccupations associated with the experimentalist.

A similar remark holds with respect to mathematically sophisticated theorizing in the social sciences.[44] Here too, one often finds a division of labor between, for example, the microeconomic theorist who elaborates a model purporting to explain the behavior of a certain market and the econometrician, statistician or (in some cases) experimentalist who elaborates procedures for testing these claims about market behavior. Indeed, in such cases there may be disagreement about as fundamental a matter as whether a claim about some economic phenomenon should be tested experimentally or non-experimentally, where different choices will produce quite different data, different potential sources of error and different problems of data analysis. Here too, such choices are typically not dictated by rules which are part of the explanatory theory under test, as the standard view would suggest.

The point that an explanatory theory of some phenomenon will commonly not prescribe (whether by correspondence rules or in some other way) the experimental procedures, techniques of data analysis, and so forth used to detect that phenomenon has a further implication which is important for philosophical discussions of testability. Although many philosophers of science have supposed otherwise, it is not at all unusual for a scientific theory (a respectable, non-metaphysical theory) to predict the existence of some phenomenon and yet for no one to have any very clear idea how to design an experiment or to gather data which would reliably establish the existence or non-existence of that phenomenon.[45] Indeed, it seems to be a serious possibility that human beings may never figure out how to do this for some phenomena-claims. For example, in a paper on weak neutral currents, Cline, Mann and Rubbia write that (as of 1974):

> The weak interaction of an electron and a positron to form a neutrino and an antineutrino, both of the electron type, surely exists, but it cannot be detected by any known means because there are no final products other than neutrinos. (Cline, Mann and Dubbia 1974, p. 118)

Similarly, a class of theories (so-called grand unified theories) which have been proposed recently to unify the strong, and electroweak interactions all seem to predict the decay of the proton. The simplest and most natural theory in this family predicts a half-life for the proton of about 10^{30} years. This is close to the present limit of experimental detectability, but the prediction now seems to be disconfirmed. However, other more complex theories in this class predict considerably larger half-lives for the proton and many physicists seem to hold that there are theoretical reasons for thinking that some theory in this class is (at least approximately) true. At present it is not clear how to get reliable data on whether these predicted very slow decays occur. Indeed, it seems perfectly possible that the decay of the proton may be a genuine phenomenon, and yet be slow enough that it will never be detected, given facts about the sensory and intellectual capacities of human beings and the sorts of instruments they can build. Three experimenters who established one of the most recent lower limits on proton decay write:

> Does the proton decay? Will the question ever be answered? The present lower limits on the lifetime can doubtless be extended somewhat, but this process cannot continue indefinitely. Perhaps some day a detector 10 times as large as [the device we used] might

> be built, but a detector 100 or 1,000 times as large is not feasible. Cost is not the only constraint (although it is a formidable one). If the proton lifetime is much greater than 10^{33} years, the irreducible background of neutrino interactions would probably obscure many decay modes no matter how large the detector was. Adding to the mass would simply increase the number of background events in the same proportion as the decay events. (Loseco, Reines and Sinclair 1985, p. 62)

Given the standard (Nagelian) conception of theory structure and correspondence rules, the fact that a theory can predict the existence of phenomenon which no one is sure how to detect or regarding which there is great disagreement concerning appropriate methods of detection appears very puzzling. In our treatment, this is only to be expected. Whether one can obtain data which are reliable evidence for or against the existence of some phenomenon depends on a great many matters which are independent of the structure of the theory one may posit to explain that phenomenon. Getting data which are reliable as evidence depends on such matters as whether one can construct devices (e.g., large enough accelerators) which produce the phenomenon in question, or whether one can find natural processes (producing data observable by human beings) which are sensitive to the phenomenon of interest and yet not sensitive in the same way to too many mimicking or confounding factors or background noise, on whether various other techniques for the control of such factors are available, and on whether the phenomenon of interest can be made to occur sufficiently often to produce enough data to yield statistically reliable conclusions. Figuring out how (if at all) this can be done requires both considerable thought and ingenuity and a special conjunction of circumstances – it is often not at all obvious until long after a theory has been constructed. This in part explains why the full range of phenomena a theory purports to explain is typically much wider than the class of phenomena which can be used to seriously test the theory.[46]

One can put all of this in a slightly different way. Many traditional accounts of theory-structure tend to regard a theory proposed to explain some phenomenon and the techniques used to detect that phenomenon as parts of a single integral unit. (The idea that correspondence rules connect theoretical terms in explanatory theories directly to terms which characterize data is one expression of this assumption). But both historical and contemporary investigations of science reveal a surprising amount of decoupling or independence between these two aspects of scientific practice.[47] These investigations suggest that, in many cases, the early stages of development of an explanatory theory are not driven, at least to any very large extent, by attempts to apply it in a detailed way to experimental situations. Instead such development is largely driven by internal, theoretical considerations, which may be quite independent of detailed experimental constraints.[48] The grounds on which a new theory is first proposed will typically include, for example, the claim that the theory holds out the promise of providing a unified account of a number of phenomena previously thought to be unrelated[49] or that it represents a line of development of previous theory which is required or suggested by other generally accepted theoretical commitments[50] or that it will resolve some long-standing inconsistency or inadequacy in previous theory.[51] Much of the subsequent work done by theorists will then consist in exploring and elaborating the structure of the new theory and in clarifying various theoretical puzzles which it generates, in checking to see whether the theory is consistent with other generally accepted theories, in learing how to calculate with the theory (and perhaps in devising new techniques or mathematical apparatus for carrying out such calculations), in learning how to use the theory to model various situations of interest and in exploring the relationship of the theory to various alternative theories which might be obtained by weakening or generalizing some of its key assumptions. To a very large extent these activities can proceed independently of experimental practice, and independently of issues about how to move from data to phenomena.

A similar pattern holds in connection with techniques for phenomenon-detection. These techniques often exhibit patterns of development which are surprisingly independent of the development of explanatory theory, patterns which are driven instead by such factors as the development of new instruments and experimental procedures, or by new techniques of data analysis, or by the relatively theory-independent discovery of a new phenomenon which then becomes a focus of subsequent investigation. Typically, distinctive sets of techniques and practices bearing on reliability will be associated with different general types of instruments or experiments – techniques and practices having to do with such matters as the detection and control of various sources of experimental error or the analysis of data in experiments of that type. Thus, for example, as Peter Galison describes in a recent paper (Galison 1985), bubble chambers involve a distinctive technology, distinctive sources of experimental error, distinctive problems of data analysis and reduction (cf. Section 4), and are particularly well-suited for the investigation of certain kinds of phenomena. The techniques and practices designed to insure that bubble chambers produce reliable evidence developed largely independently of theories designed to explain the phenomena bubble chambers were used to detect and, in a number of respects, are independent of scientists' understanding of the causal principles governing the operation of the chamber itself. In a number of fundamental respects these techniques remained unchanged as bubble chambers were used to detect new phenomena and as the theories postulated to explain those phenomena changed. On the other hand, independent technological innovations like the development of high speed computers exerted an extremely important influence on the experimental practices associated with bubble chambers.

Because of this relative independence, explanatory theory, and techniques for phenomena-detection do not develop as a single unit, in automatic lockstep, but can exhibit a wide variety of different patterns of interaction. In some cases, most of the scientific activity associated with a theory will be of the internal, theory-driven kind described above, and there will be little relevant activity concerned with the detection of phenomena – whether because it is still unclear what distinctive predictions about phenomena the theory makes or because, while such predictions are known, no techniques of detection exist for checking them. The development of general relativity in the period from about 1930 to 1960 is an example of this sort of pattern.[52] Current work on superstring theory represents another example. Conversely, it is also possible for a class of phenomena to be detected and for reliable techniques for investigating their behavior to develop well before any serious and systematic explanatory theory of their behavior is proposed – the discovery of various forms of novel radiation in the period from about 1895 to 1910 furnishes a striking illustration of this.[53] By emphasizing the differences between the relationship between phenomena and explanatory theory, where the characteristic concern is with explanation and prediction, and the relationship between data and phenomena, where the characteristic concern is with reliability of evidence, the account offered in this paper makes it clear how these various patterns of independent development are possible.

In more recent, post-Nagelian work on theory structure and correspondence rules it is generally recognized that the rules and procedures for connecting data to phenomena are not, properly speaking, part of or specified by the theory that explains those phenomena. However, most commentators take these rules or procedures to be supplied by other explanatory theories – so-called 'background' or 'auxiliary' theories – which provide causal explanations of the operation of the instruments or measurement devices one employs. In effect, the basic picture of theory structure underlying the standard, Nagelian view (with its strong assumption about the possibility of deducing claims about data from general theory) is retained, but the number of different theories said to be involved in concrete testing situations is multiplied. A sequence or hierarchy of correspondence rules is assigned where the standard view would ascribe just one such rule. According to this approach, a general, high-level explanatory theory will involve rules or generalizations connecting theoretical claims with what we would call claims about phenomena and then various auxiliary or background theories (themselves causal theories) will supply rules of the same general sort connecting phenomena-claims to data-claims. In virtue of this sequence of rules, a general theory is understood as explaining or permitting the explicit deduction of relationships obtaining in the data.

A well-known statement of such a view can be found in Kenneth Schaffner's paper, 'Correspondence Rules' (Schaffner 1969). According to Shaffner, one function of correspondence rules is to establish 'causal sequences' connecting theory with data. Commenting on the sorts of correspondence rules supposedly associated with Bohr's theory of the atom, Schaffner writes

> What connects theories, especially micro theories of the above type [e.g., Bohr's Theory] with observations are *like* the C-rules [i.e., correspondence rules] linking electron jumps with spectral lines – but these connections are sufficiently different from simple linkages to warrant a distinctive analysis. These types of linkages are better understood as *causal sequences* commencing with an action of a theoretical entity and culminating in an 'observationally' accessible event or situation. These latter phenomena may be as diverse as the 'first throw reading of a ballistic galvanometer' or a 'line with specific intensity and curvature in a bubble chamber photograph'. What

warrants the particular parts and interconnections of the causal sequence are assumed or borrowed scientific theories. (p. 286, emphasis in original)

According to Schaffner, in this example, the relevant "causal sequence analysis" might go something like this:

There is an electron transition (between energy levels $E_2 - E_1$) → a photon of energy $E_2 - E_1 = hv$ is emitted by the electron → photon moves with velocity c in a straight line to a glass prism → it is refracted through the prism (this itself might well be expanded) → it passes next to a cross hair and through the eye's lens to impinge on the observer's retina → it (or perhaps a number of photons) causes a neural excitation → the observer is aware of a specific color and shape. (p. 287)

Here the idea is quite explicit that theory stands to phenomena in something very like the same relation that phenomena stand to data, and that just as one wants a theory which explains facts about phenomena, so also one needs an explanatory theory establishing connections between phenomena and data – a theory which specifies the causal mechanisms by which data is produced. By putting these kinds of claims together, one produces an account of how the claims of theory explains facts about raw data.[54] Schaffner's theory thus adopts many of the assumptions about theory structure – in particular the assumption that the appropriate, sought-after relationship between phenomena and data is an explanatory one, in which the direction of inference is downward from phenomena to data – that I have been most concerned to criticize.

The difficulties with this sort of view should be familiar by now. As we have already suggested, in inferring phenomena-claims from data-claims one's interest is often not in explanation or in uncovering the details of the causal mechanisms by which data is produced, as Schaffner claims, but rather in establishing that the data-claims are reliable evidence for the phenomena-claims. While establishing this certainly involves empirical investigation and theory in the sense of assumptions that go beyond the data, it often need not involve the construction of an explanatory theory of data-claims. Many of the crucial assumptions and procedures involved or assumed in the analysis of data are not remotely the sort of assumptions that can figure as premises in a causal explanation of facts about the data and are not assumptions of a sort that would be supplied by a causal theory of the operation of one's instruments – they are rather, as we have noted, statistical assumptions and procedures of various sorts, assumptions about the character, distribution, and control of various kinds of error, assumptions about when it is appropriate to discard data, and assumptions about how to reduce large masses of data. Such assumptions bear on issues having to do with the reliability of evidence or the possibilities of extracting certain kinds of information from the data, not on questions of explanation. Moreover, as we have also argued, there are other very good reasons having to do with the need for unification and systematization in scientific explanation, for not taking what is ultimately explained by a physical theory like Bohr's to be facts about particular instrument readings or perceptual judgments by particular observers. While there is certainly a minimal sense in which it must be true that if one is to do science one must establish a "connection of a theoretical process with laboratory experience" (Schaffner 1969, p. 287) the connection one seeks to establish need not be an explanatory connection.

7.

One of the central themes of my discussion has been that the relationship between data-claims and phenomena-claims is importantly different from the relationship between phenomena-claims and the claims of explanatory theory. In this section, I want to comment very briefly on the implications of this idea for some standard accounts of confirmation and testing. The first point to be made is that standard accounts like the hypothetico-deductive model or Popperian falsificationism are, at best, accounts of the relationship which ought to obtain between phenomena-claims and the claims of explanatory theory. Such accounts are not plausible reconstructions of the relationship which ought to obtain between data-claims and phenomena-claims. That is to say, these standard accounts ought to be understood as attempts to capture the relationship which obtains between, for example, claims about the existence of neutral currents and the Weinberg-Salam model, and not as attempts to capture the relationship betleen neutral currents and the bubble chamber photographs which constitute evidence for neutral currents.

One reason for this is that the traditional accounts generally require deductive relationships between hypothesis and evidence of a sort that typically do not exist between phenomena-claims and data-claims. Both the hypothetico-deductive account and Popperian falsificationism require that evidential claims be deducible (at least in the case of deterministic theories) from the hypothesis which they are used to support or test. But if my argument above is correct, claims about data frequently are not deducible from claims about phenomena even in conjunction with background theory. More importantly, those who argue for the existence of phenomena on the basis of a body of data do not do so primarily (if at all) by appealing to the existence of any such deductive relationship. For example, the arguments concerning the existence of neutral currents described above did not turn on whether facts about bubble chamber photographs could be deduced from claims about the existence of neutral currents and other theoretical assumptions.

A second reason for holding that many of the traditional accounts of testing and support are not plausible reconstructions of the evidential relationship which ought to obtain between data and phenomena is simply the absence in these accounts of any explicit treatment of many of the considerations which we found to be most central in connection with making inferences from data to phenomena: control of experimental error, replicability, procedures for data reduction, calibration of instruments, and so forth. Traditional accounts of theory confirmation in effect begin when worries about experimental error, replicability, control of background factors, and related matters have been adequately dealt with. This is just another way of saying that the traditional accounts presuppose that one has established the existence of phenomena and address themselves to questions having to do with the kind of support claims about phenomena provide for general explanatory theories; they do not address themselves to questions about the support a body of data provides for claims about phenomena.

If the relationship between data and phenomena, on the one hand, and phenomena and explanatory theory, on the other, differs in the ways I have claimed, this naturally opens up the possibility that the kinds of inferential or inductive strategies which are appropriate in moving from claims about data to claims about phenomena may be different in important ways from the kinds of strategies which are appropriate in connection with moving from claims about phenomena to claims concerning the correctness of explanatory theories. In what follows, I want to suggest this may indeed be the case, although I should emphasize that my remarks are intended to be merely suggestive and illustrative of possibilities that deserve more detailed exploration elsewhere.

First, consider the contrast between what we may call falsificationist and confirmationist inductive strategies.[55] As a rough approximation, one follows a falsificationist strategy if one invents bold hypotheses of considerable generality, tests them in exactly those areas in which they seem most likely to break down, readily considers alternatives to the hypothesis one presently holds and searches for evidence which will discriminate between this hypothesis and alternatives, rejects a hypothesis when one seems to obtain falsifying evidence, eschews ad hoc or post hoc procedures of hypothesis modification and attaches a special value to hypotheses which make successful novel predictions. By contrast, one adopts a confirmationist strategy if one looks for evidence which will support one's favored hypothesis in those areas where one antecedently thinks it most likely that such evidence will turn up, fails to systematically consider alternatives to this hypothesis, attaches considerable weight to the successful discovery of confirming evidence and is relatively less interested in uncovering negative or falsifying evidence, and readily tolerates ad hoc or post hoc modifications of one's hypothesis as new evidence becomes available.

It is no doubt true (allowing for the element of caricature in the above formulations) that elements of both general strategies can be found at all levels of scientific inquiry. Nonetheless, I think that it is not an implausible conjecture that the relative weight and role of these two strategies will be different, depending on whether one is interested in assessing whether a body of data supports a claim about a phenomenon or in assessing whether a claim about a phenomenon supports a general explanatory theory. At least in many areas of science, it seems plausible that the testing of explanatory theory against claims about phenomena both does and ought to involve strategies with a large falsificationist component. In high energy physics, for example, one finds the characteristic strategy of testing accepted theories under novel and extreme conditions (e.g., higher and higher energies) where it seems most likely that falsifying evidence will be obtained, should it be obtainable at all. Considerable importance is attached to the systematic generation and exploration of theoretical alternatives, even if many of these appear quite speculative. When two theories make similar predictions with respect to established phenomena, investigators look for points at which the two theories make novel predictions which diverge and which are experimentally testable. Successful novel predictions are regarded as providing particularly good support for a theory, and failure to make novel predictions is taken to be a prima facie defect. In general, the underlying strategy is one of investigating the theory under conditions in which, if it is false, its falsity can be most readily detected.[56]

By contrast, in inferring claims about phenomena from data, scientists typically and unavoidably employ strategies with a large confirmationist component. In large measure this is dictated by the nature of the task at hand. Detecting a phenomenon is typically not a matter of exploring those cases in which a hypothesis seems most likely to turn out to be false or in which it makes predictions that diverge from some alternative hypothesis. Detecting a phenomenon is, rather, a matter of arranging an experimental set-up and controlling for various sources of error in a way that makes it as likely as possible that the phenomenon will produce unambiguous evidence for its existence. This involves looking for the phenomenon exactly where one thinks it is antecedently most likely one will find it, arranging one's detection device or experimental apparatus in such a way as to create the optimal conditions for successful detection, and adopting procedures of data analysis that maximize one's chances of being able to locate and extract information about the phenomenon. Since many phenomena are rare and difficult to detect, phenomenon detection also often involves patiently persisting in the search for a positive instance, even when previous attempts have been unsuccessful. Detecting a phenomenon is like looking for a needle in a haystack or, to return to a metaphor used above, like fiddling with a malfunctioning radio until one's favorite station finally comes through clearly. Of course, one wants such searches to be designed in such a way that they will turn up negative results if what one is looking for does not exist but, in contrast to the testing of explanatory theory, the idea of structuring the investigation around the search for falsifying evidence seems quite inappropriate.

Once this is recognized, the confirmationist character of many of the strategies and procedures scientists use for establishing the existence of phenomena appears less surprising and disturbing. Consider, for example, the well-documented tendency of many experimentalists to focus mainly on the search for 'positive instances' which support claims about phenomena they think may be true, and to discard or ignore at least some evidence which seems prima facie inconsistent with these claims. This sort of tendency is usually regarded as a paradigm of 'confirmation bias', and although it is clearly objectionable on a purely falsificationist methodology, it begins to look rather less problematic when one reflects that large portions of the data produced in most experiments will be subject to sources of error which are difficult to detect and imperfectly understood, and that other portions of the data may be simply irrelevant to the phenomena one is trying to detect. If, for example, one can detect a substantial number of cases in which the measured value of the charge on an oil droplet appears to be an integral multiple of the same fundamental unit, it may be entirely reasonable to discard a smaller amount of data which is prima facie inconsistent with the result (as Robert Millikan did), on the grounds that in these cases the experiment was probably infected with unknown sources of error. Once one distinguishes between the evidential relationship which ought to obtain between data and phenomena from the evidential relationship which ought to obtain between phenomena and general explanatory theory, the idea that confirmationist strategies of this kind may be warranted in connection with the former relationship, even when inappropriate in connection when the latter relationship, looks quite plausible.[57]

I turn now to a second way in which different inductive strategies may be appropriate in connection with the move from data to phenomena and the move from phenomena to general theory. As I have already noted, arguments which appeal in various ways to the improbability of certain kinds of coincidences play an important role in supporting inferences from data to phenomena. For example, that a phenomenon apparently has been detected in independent experiments which employ different experimental designs embodying different physical principles is often an important piece of evidence that one has detected a real effect, and not just an artifact. Here the idea is that while it is quite possible that an unknown confounding factor may be operating in the case of any one experiment, it often would be an improbable coincidence if confounding factors were to operate in different pieces of equipment in such a way as to produce evidence in all of them which pointed to the existence of the same phenomenon. Often it will be much more plausible to conclude instead that the different experiments have detected a common phenomenon.

Arguments which turn on the improbability of coincidences are of course sometimes also employed in connection with the support furnished by phenomena-claims for explanatory theories. Consider a theory which correctly predicts the existence and characteristics of a wide range of different phenomena. It is sometimes argued, in support of such a theory, that it would be an improbable coincidence for these various phenomena to be just as the theory predicts them to be and yet for the theory in question to be seriously false. According to some writers (e.g., Thagard 1978). Darwin appealed to just such an argument in support of his theory of natural selection. According to other writers (e.g., Harman 1965; Salmon 1984), we are justified in believing in the approximate correctness of atomic theory and kinetic theory and in the reality of atoms and molecules because many different phenomena (having to do with Brownian motion, the ideal gas laws, and so forth) are successfully predicted by these theories and because it would be an improbable coincidence if this were the case and yet the theories were not even approximately correct.

While appeals to coincidence of this sort are sometimes legitimate in connection with explanatory theories, I think it is clear that the empirical assumptions on which such arguments rest are more likely to be satisfied in the case of inferences from data to phenomena than in the case of inferences from phenomena to explanatory theories. In the case of a phenomenon which is apparently detected by several instruments, one is often in a position to know that the instruments operate in accord with distinct physical principles and one often knows enough about possible confounding factors to support the claim that it is unlikely that such factors will affect distinct instruments in exactly the same way. The situation is importantly different with respect to the judgment that, say, it would be an improbable coincidence if various phenomena were just as standard as kinetic theory predicts them to be and yet that theory turned out to be seriously mistaken. This judgment depends crucially on whether there are (or are available) alternative theories, significantly different from standard kinetic theory, which also predict that the phenomena in question will obtain. Judgments of this sort require an exploration of at least some relevant portion of the space of alternative theories to standard kinetic theory and of the implications of such theories for the phenomena which kinetic theory explains. It is easy to be mistaken about such judgments – one can easily fail to think of a possible alternative theory at all, or fail to realize that it can be developed in detail in such a way as to predict phenomena in the domain of interest.

The development of versions of the wave theory of light in the nineteenth century that successfully predicted many optical phenomena, such as stellar aberration, which were previously thought to be not susceptible of any wave-theoretical treatment furnishes an illustration of this point. The problematic character of the argument that the particle theory of light successfully predicts a range of optical phenomena and that this would be an improbable coincidence if the particle theory was seriously mistaken is, I take it, plain enough. Unlike arguments from coincidence for the reality of phenomena, arguments from coincidence for the approximate truth of general explanatory theories like kinetic theory or the particle theory of light require information which goes far beyond readily obtainable information about the operation of one's instruments and of potential confounding factors. Just because it is justifiable to argue that a phenomenon detected by a number of different experimental procedures is unlikely to be an artifact, it does not follow that a precisely parallel line of argument can be used, without special additional assumptions, to support the correctness of explanatory theories like kinetic theory. Distinguishing between data and phenomena and recognizing the differences between the kinds of evidential relations which ought to obtain between, on the one hand, claims about data and claims about phenomena, and on the other hand, between claims about phenomena and the claims made by general theories like the kinetic theory, helps to alert us to this point.[58]

8.

At the beginning of this essay, I remarked on the claims of sociologists of science regarding the craft nature of experimental work. In this section I want to explore some of these claims in more detail and to suggest that they are just what one would expect if my contentions about data, phenomena, and the strategies scientists use for assessing whether data are reliable evidence are correct.[59]

I begin with a few remarks by way of summary of the relevant sociological claims. A common theme running through much of the sociological literature and also through a number of commentaries by experimentalists themselves is that experimentation involves a large 'tacit' or 'craft' element.[60] It is claimed that conducting an experiment in a way that produces reliable results is not a matter of following (even unconsciously or implicitly) algorithmic rules which can be written down in a way which specifies exactly what is to be done at each step. Often the only way one can learn to perform a certain kind of experiment reliably is by doing a number of experiments of that kind oneself, under the guidance of an expert who already knows how to perform such experiments. Thus, according to Trevor Pinch, experimenters possess skills which "often enable the experimenter to get the apparatus to work without being able to formulate exactly or completely what has been done" (Pinch 1986, p. 112). It is also claimed that assessing whether another investigator has produced reliable results is not a matter of checking to see whether he has followed some specifiable set of rules, but rather requires expert knowledge of a sort that is typically possessed only by those who have actually worked with experimental systems of the kind in question.

These claims are supported by various sorts of empirical evidence, although much of it is unsystematic and anecdotal in character. For example, in a well-known study, Harry Collins (1975) investigated a number of experimental groups working in Britain to recreate a new kind of laser that had been successfully constructed elsewhere. Collins found that no group was able to reproduce a working laser simply on the basis of detailed written instructions. By far the most reliable method was to have someone from the original laboratory who had actually built a functioning laser go to the other laboratories and participate with the members of those laboratory in the construction. Rather than thinking of the transmission of the ability to reliably reproduce an experiment as involving the transmission of explicit rules, Collins suggests that

> [t]he model which seems most appropriate [for the transmission of the ability to reliably reproduce an experiment] is one which involves the transmission of a *culture* which legitimates and limits the parameters requiring control in the experimental situations, *without necessarily formulating, enumerating or understanding them*. (Collins 1975, p. 208, emphasis in original)[61]

Remarks on experimental work by working scientists themselves often express similar claims.

A closely related theme, which is developed in considerable detail in Trevor Pinch's recent book *Confronting Nature* (Pinch 1986), has to do with the role of the 'personal warrant' of the experimenter in assessing the reliability of his experimental results. According to Pinch, other scientists often will place at least as much weight on an experimentalist's general reputation for careful, painstaking work as on the technical details of his experiment in assessing whether his data constitute reliable evidence. Pinch plausibly connects this theme with claims about the craft nature of experimental work similar to those described above. Commenting on the reasons that the results of Ray Davis, the primary experimentalist in the solar neutrino experiment, were widely accepted, Pinch writes

> Many of the episodes described in earlier chapters [of *Confronting Nature*] have revealed a picture of science in which the personal warrant of scientists is perhaps surprisingly to the fore. We have seen that scientists have, on more than one occasion, placed personal warrant – that is knowing and trusting a particular scientist or trusting someone else's respected judgement – before the detailed technical merits of the work under discussion. For instance, several scientists based their faith in Davis' results upon personal warrant rather than upon a deep understanding of his experimental techniques. The reason for this, I suggest, is partly related to the complexity and nature of the tasks which these types of scientist carry out. There are many different sorts of activity involved which require a variety of competencies. Furthermore, many of the practices possess a tacit craft element which makes it difficult to formalize exactly what has been done. Similarly it is well known that an experimenter's ability often resides in being able to tweak the apparatus in the appropriate way. Such knowledge is acquired by a highly specialised and skilled practitioner after a long 'apprenticeship' and such knowledge is not widely shared. (Pinch 1986, p. 205–6)

Memoirs and reminiscences by scientists themselves often express similar ideas.[62]

These claims about experimental practice are both plausible and understandable in the light of my previous discussion. If assessing the reliability of an experiment were just a matter of constructing a derivation of one's data from general theory or of arriving at a theoretical understanding of how one's instruments normally work, then one would perhaps not expect craft considerations or personal warrant to play any very large role. The calculations and theory in question would be a matter of public record, and could be assessed by anyone with the right sort of mathematical and theoretical training. As I have repeatedly emphasized, however, establishing reliability is typically not merely, if at all, a matter of constructing a derivation or a causal theory of one's instruments. Instead the assessment off reliability involves the sorts of considerations described in Section 4 above, and many of these considerations naturally support the expectation that craft considerations will play a large role in experimental practice.

Consider, for example, the importance of controlling for confounding and background factors. Geneal theory may yield an understanding in principle of how an experimental apparatus (such as a cloud chamber) works in the absence of confounding factors and it may also tell one that certain factors (e.g., the presence of an undetected local magnetic field) are potential confounders. But general theory will not by itself tell one whether some confounding factor is actually present in a given experimental context, or even whether it is the sort of confounding factor which is likely to be present in typical laboratory situations. Instead, whether various sources of experimental error actually constitute serious problems in practice is often discoverable only through experience – by empirical investigation of the behavior of one's equipment. Similarly, general theory often will not suggest how to actually detect whether a confounding factor is present or what sort of physical change in an apparatus will most effectively control for its possible influence.[63] Here too, information about how to do this often is best obtained by experience with other similar experiments.

Moreover, as we have already noted, one important piece of evidence that an experiment is functioning reliably, and that there has been adequate control for various sources of error, is that the experiment can reproduce known phenomena with expected characteristics. It seems clear that getting an experimental apparatus to do this involves a substantial element of "learning by doing". This helps to make it intelligible that, as the sociologist G. D. L. Travis claims, it is often the case that "the only way for scientists (or anyone else) to know that they have the knowledge to produce a competent replication is actually to produce what everyone agrees to be a working experiment" (Travis 1981, p. 13).

When one reflects that in most experimental situations there will be not just one, but a long and open-ended list of possible confounding factors, that while one may possess a general theory specifying the operation of some of these, typically no theory will be available which specifies the operation of all of them individually or how they will interact, both the claims of sociologists about the openness, complexity, and imperfectly understood character of experimental systems and the idea that dealing effectively with these potential sources of error is not just a matter of following explicit rules look rather persuasive.

A number of the other considerations relevant to reliability described above support a similar conclusion. Consider, for example, decisions to discard data. As several writers have stressed,[64] subtle clues in the behavior of an apparatus or an experimental subject or in the pattern of data produced in an experiment can alert an experienced investigator to the possible presence of various sources of error or malfunction, and hence justify a decision to discard data. Here too, learning to recognize these clues has a substantial craft element, which seems to come only with considerable experience with the domain under investigation.

Similar considerations apply in connection with the role of personal warrant in science. Given the complexity of many experimental systems, the numerous sources of error that may be present, and the difficulty of detecting and correcting for these just by relying on general theoretical knowledge, the most sensible epistemic strategy for someone who has not had a good deal of experience with the kind of experiment in question may be to attach considerable weight to the general reputation of the experimenter and to the reaction of other experts in assessing the reliability of the results. It is often plausible that those who have done competent and reliable experimental work in the past possess abilities and character traits that make it likely that they will continue to do so in the future. Because of this, relying in part on an experimenter's past record and general reputation (provided of course that this reflects his real performance) can be quite reasonable. If the potential sources of error in an experiment are numerous and complex, and difficult to capture in a general theory or explicit set of rules, and hard to recognize if one has not had considerable first-hand experience with similar experiments, the strategy of relying on reputation and personal warrant may lead many scientists to a more reliable assessment than the strategy of attempting to assess the technical details of the experiment directly. It is thus not surprising that, by their own testimony, scientists frequently employ this strategy. Conversely, it is plausible to regard the frequency with which the strategy is employed as itself evidence for many of the claims made above regarding the imperfectly understood character of many experimental systems and the impossibility of establishing reliability merely by the construction of explicit derivations or the invocation of the sort of information supplied by explanatory general theory.[65]

9.

I suspect that a great many other issues in the philosophy of science and many features of scientific practice will look quite different once the distinction between data and phenomena is taken seriously. By way of conclusion, I shall briefly describe a few additional illustrations.

1. A central theme of my discussion has been that the techniques and assumptions which warrant taking certain data as evidence for some phenomena are typically not supplied by explanatory theories of those phenomena – indeed the reliability of those techniques and assumptions is to a surprising degree epistemologically independent of a detailed knowledge of the causal processes producing either the phenomena or the data. This suggests the possibility of a kind of investigation the value of which has been insufficiently appreciated in recent analyses of science: the careful investigation of the properties of some phenomenon, in the absence of a systematic explanatory theory of that phenomenon.[66] As Ian Hacking claims in his *Representing and Intervening* (Hacking 1983), the tendency in recent philosophy of science has been to assume that if empirical investigation is to yield useful results, it must amount to the explicit testing of explanatory theory; that an investigation which is not guided by the prior formulation of a set of explanatory hypotheses is likely to amount to "mindless empiricism" or "naive inductivism". This tendency is of course encouraged by the idea that empirical investigation always requires strong and extensive assumptions about the correctness of various explanatory theories anyway, and that one might as well explicitly formulate and test such assumptions.

If my discussion above is correct, this is a misleading picture of how a great deal of scientific investigation proceeds. The early history of many scientific disciplines is full of examples of investigations into the existence and characteristics of various kinds of phenomena and the invention of ingenious instruments and analytic techniques to facilitate such investigations, even though good explanatory theories regarding either the phenomena themselves or the operation of the instruments were lacking. Unambiguously establishing the existence and character of a certain range of phenomena has been in many historical cases an important preliminary to the construction of good explanatory theory; it constrains the range of possible candidates of explanatory theory enormously and makes systematic development and comparison of these candidates feasilble. One finds this pattern in connection with, say, early investigations in electricity and magnetism, or thermodynamics, or certain areas of chemistry.[67] Nor is this a mark of relatively immature or primitive science. The possible existence of many of the physical phenomena mentioned above – the weak neutral current, magnetic monoples, the decay of the proton – were investigated by experimentalists prior to the construction of explicit explanatory theories predicting these effects. In remarks made shortly after the discovery of neutral currents, Richard Feynman agreed that the detection of neutral currents provided support for the Weinberg-Salam theory. But he added

> I would like to follow the advice of [the experimentalist] Mann. Neutral currents should be studied in their own right. That means the experimentalists should say, all right, we have neutral currents, let's find out what their properties are. [Rather than just comparing them to the theory of Salam and Weinberg]. (Quoted in Galison 1983, p. 504)

The idea that it is possible and potentially valuable to investigate phenomena in their own right in this way (and not just as part of an explicit test of some explanatory theory) and that the results of such an investigation can yield important constraints on subsequent theorizing is closely associated with another idea which was briefly suggested above – that claims about the existence of phenomena and the appropriateness of certain instruments and techniques for their investigation represent important kinds of continuity in scientific change, kinds of continuity which are distinct from continuity at the level of explanatory theory. [Note that this is very different from the continuity at the level of observation traditionally stressed by positivist philosophers of science, which on my analysis comes out as (roughly) continuity at the level of data. This last sort of continuity, if it exists at all, is relatively uninteresting for understanding science.] I do not of course mean to deny that claims about phenomena or about procedures for their detection are sometimes mistaken at least in the short run – our discussion above provides several examples. But I do mean to suggest that there are many actual cases – especially when there has been replication, detection by various procedures, careful control of error – in which it is reasonable to believe this outcome is quite unlikely. Quite probably any subsequent theory of the weak interactions will need to account for the existence and detected characteristics of weak neutral currents, just as any theory of gravitation which replaces or modifies general relativity will need to account for such phenomenon as the gravitational red shift and the deflection of starlight by the sun. The language used by scientists regarding phenomena claims reflects this point. Such claims are often said to be 'demonstrated', 'established', or 'proved'. This language is less often used in connection with high level explanatory theory.

2. The distinction between data and phenomena also suggests one natural way of thinking about the relative maturity of a scientific discipline. In some areas of inquiry (e.g., portions of physics, chemistry, and biology) one finds both well-established claims about many phenomena and explanatory theories in which one can have considerable confidence. In other areas one finds many well-established claims about phenomena (claims which it is reasonable to think will probably be accounted for by subsequent explanatory theory), but often the present claims of explanatory theory deserve to be treated with skepticism. For example, the literature in cognitive psychology is full of ingenious experiments and techniques of data analysis, which establish the existence of many striking phenomena – chunking and recency effects in memory, various visual illusion effects, effects said to be associated with mental imagery such as the well-known results of Shepard regarding mental rotation. These effects exhibit the features which are characteristic of phenomena – they are robust and stable across different subjects and experimental settings, and readily reproducible. There is no reason to think that they are artifacts of the particular experimental techniques employed. It is a good guess that if comprehensive psychological or physiological explanatory theories of memory and perception are ever developed, they will need to account for these phenomena, among others. However, while present attempts at explanatory theorizing in psychology are often interesting and suggestive, in my view skepticism with regard to their (even approximate) correctness is often in order, largely because there are too many fundamentally different alternative theories among which it is at present impossible to discriminate experimentally at least on the basis of the sorts of non-physiological evidence on which psychologists typically focus. Here one often has believable claims about

phenomena, a not inconsiderable achievement, but not about explanation.

In still other areas of investigation, while there is plenty of data, it is much less clear what the phenomena are. Consider, for example, the use of surveys as a source of information about political attitudes in an electorate – particularly attitudes toward national political issues of some enduring significance and potential relevance to voting behavior such as civil rights, government expenditures, and foreign policy issues. While these surveys have produced an enormous amount of data, a number of political scientists have recently expressed skepticism about whether such data tell us anything about the existence of stable attitudinal phenomena. There is reason to think that answers to survey questions are often extremely sensitive to small changes in the wording of questions, to the order of questions, and to variations in the content of other questions in the survey. There is also reason to think, although this is a matter of controversy, that over significant periods of time (four or eight years) the response of many subjects to the same question will exhibit relatively small correlations – that is, that there is considerable instability of response over time. The political scientist Phillip Converse suggests in a well-known article (Converse 1970), that at least for some issues, the population will divide into two groups, the smaller of which will exhibit very high correlations over time and the much larger of which will exhibit near zero correlations (the responses of this group are nearly "random" (p. 173)). Converse suggests that only the first group has "real and stable attitudes" or "genuine attitudes" and that a natural interpretation of the second group is that it consists of those with "no real attitudes" or "non-attitudes", (p. 175) but who felt obliged to produce some response to the question nonetheless (hence the essentially random character of their response). Clearly, while we have plenty of data regarding the second group, it is not clear that this data tells us anything of significance about the attitudinal phenomena of interest. As another writer suggests

> The analysis of non-attitudes is the analysis of noise, not genuine information; as such, the substantive significance of such data is at best dubious. (Asher 1983, p. 17)

If much of the data produced by surveys is not reliable evidence for the existence of attitudinal phenomena, it may be premature to attempt to construct explanatory theories relating such phenomena to, say, voting behavior, as a number of political scientists have attempted to do.

This example illustrates again the non-trivial character of phenomenon-detection (in contrast to the mere gathering of data, which is often easy). If the point of doing science were to explain the data, it would always be clear enough what one ought to try to explain; one's concern just would be to produce an appropriate explanatory theory. But in some areas of investigation, it is not even clear what one should regard as possible candidates for explanation – it is not clear what is a 'real' effect, and what is an artifact produced by our methods of measurement, unknown confounding factors, and so forth. Learning to separate the real from the artifactual is an important kind of scientific progress, although it is not progress in the construction of explanatory theory. If we make a distinction between data and phenomena we can readily make sense of these features of scientific enterprise; they are far less intelligible on traditional analyses, which deny the discussion.

3. So far I have been commenting on some of the implications of my analysis for the understanding of scientific practice; I turn now, by way of conclusion to an issue which is more internal to philosophy of science: Bas van Fraassen's version of "constructive empiricism" (van Fraassen 1980). According to van Fraassen, accepting a scientific theory ought to involve the belief only that it is "empirically adequate", where the test of empirical adequacy is that "what [the theory] says about observable things and events is true" (p. 12). Elsewhere he tells us that the only non-pragmatic consideration relevant to acceptance is whether a theory "saves the phenomena", where the phenomena are given by "all possible observations of the actual world". It is clear that van Fraassen means by "phenomena" roughly what I mean by [all possible] data. Given van Fraassen's views about what can be observed – views which seem quite plausible – it is typically data rather than phenomena which are observed and which do not depend on assumptions which go beyond what is uncontroversially observable. van Fraassen's view thus becomes, in the terminology of this paper, that to accept a scientific theory is to believe that it saves the data.

Critics of *The Scientific Image*, have (quite correctly) insisted that acceptance must involve more than this. What I want to point out here is that if the discussion above is correct, the demand that a theory must save the data is not even a plausible *necessary* condition for reasonable acceptance. Good explanatory theories typically do not save the data, although they do save the phenomena in the sense of "phenomena" described above. Consider a very simple case. You want to measure the expansion of a metal bar upon heating – the phenomenon which interests you. You make a series of measurements of the increase in length and these yield of course a scatter of different results – your data. These are roughly normally distributed and making the usual optimistic assumptions about the absence of systematic error and so forth, you take the mean of this distribution as the true length after heating. Now on van Fraassen's criteria for observability it will be the actual measurement results themselves (the data) rather than this mean value which will be a plausible candidate for what is observed. Similarly, on van Fraassen's principles, while bubble chamber photographs or perhaps the tracks in the chamber will be plausible candidates for what is observed, the trajectories which are fitted to tracks by complicated curve-fitting procedures and the particles which produce the tracks will not be observed. Yet if my discussion above is correct, it is the true increase in length upon thermal expansion (as given by the means of the individual measurements) rather than the scattered results of individual measurements which must be saved by (or derivable from or explained by) the theory of thermal expansion. Similarly, it is facts about the energy, mass, charge, and other properties of various elementary particles and not the complex and idiosyncratic features of particular bubble chamber photographs which must be saved by any theory of the weak interactions.

What we observe in van Fraassen's sense will typically reflect not just features of the phenomena in which we are interested but various other background or confounding factors which we regard as noise or error and which show up in our data. The apparently plausible demand that science should save, predict, or explain what we observe is just the demand that science should save, predict, or explain the complex results of the presence of this background noise – the idiosyncratic scatter of individual length measurements in the thermal expansion example. If the discussion above is correct, many general scientific theories succeed exactly because they do not try to do this. Such theories succeed because investigators first try to filter out the noise, background, and confounding factors in their data, and only then attempt to extract the signal or phenomenon which they think will be a plausible candidate for explanation or prediction.

Might van Fraassen respond to this objection by extending his notion of observability to at least some phenomena? It is hard to see how he could do this consistently with his general epistemic principles. While I have argued that claims about phenomena often do not require for their warrant knowledge of the details of explanatory theory, I have also emphasized that claims about phenomena (and above the reliability of procedures used to detect them) typically do require for their warrant assumptions that go well beyond the data, assumptions about processes that are, in van Fraassen's sense, unobservable. For example, the use of the mean value of individual length measurements as the true value of the length involves assumptions about the character of various unknown small causes producing deviations from a true value (e.g., that these operate independently and additively) which do not seem to be capable of direct observational checks of the sort favored by van Fraassen. Similarly, correcting for the neutron background or hadron punchthrough in the neutral current experiments requires many assumptions about unobservable factors and processes.

More generally, whenever one makes inferences about the existence of phenomena on the basis of data, one must make assumptions about various possible sources of error and about their control and detection, assumptions about the reliability of procedures for gathering, reducing and analyzing data, assumptions about the sensitivity of various detection or testing procedures and many other kinds of assumptions as well. These assumptions will not be just about what is observable, in van Fraassen's sense. It is a mistake to suppose that a theory must "save" all that one has observed (this is why one has a notion of experimental error) and a mistake to think that agreement with the results of observation provides by itself good reason to believe any part (even the "observational" part) of what a theory claims. By itself such agreement counts for little in the absence of the use of appropriate procedures to insure reliability and good error characteristics in one's testing procedures, as the spurious detection of magnetic monopoles or gravitational radiation described above show. Here again, one sees the importance of distinguishing data and phenomena in understanding science.

NOTES

* Many of the ideas in this paper result from discussions with James Bogen. A collaborative paper, entitled 'Saving the Phenomena' (*Philosophical Review*, July 1988, pp. 303–52) focuses on the implications of the distinction between data and phenomena for traditional accounts of observation in science. The present paper develops a number of the ideas in that discussion in more detail, provides additional examples, and explores the implications of those ideas for traditional accounts of theory-structure and theory-testing.

Portions of this paper were presented as part of a workshop at the Philosophy of Science Association Meetings in Pittsburgh in October 1986 and as part of an invited symposium at the Pacific Division meetings of the American Philosophical Association in March 1987. Ronald Laymon and Trevor Pinch provided very useful comments at the Philosophy of Science Association meetings. For encouragement and helpful suggestions I am also indebted to Nancy Cartwright, Allan Franklin, Ron Giere, Peter Galison, Arthur Kuflik, Thomas Nickles, Lee Rowen, and to two anonymous referees for *Synthese*.

[1] Although most of the examples discussed below of phenomena detection and the characteristic problems to which they give rise are drawn from the physical sciences and involve experiments, it is important to understand that issues about whether a body of data constitutes reliable evidence for some phenomenon of interest arise in the social and behavioral sciences, and in non-experimental contexts as well. Indeed, issues about what the relvant phenomena are and which are the reliable techniques for their detection are particularly important and often particularly controversial in the latter contexts. One can think of many of the characteristic techniques of data analysis in the social and behavioral sciences – for example, regression analysis and multidimensional scaling techniques – as techniques for detecting phenomena in non-experimental contexts.

Consider, for example, the social scientific phenomenon mentioned in the text above – the relatively higher rate of technical innovation in moderately concentrated industries. Whether this is a genuine phenomenon has itself been a matter of considerable controversy. The authors of a recent study (Kamien and Schwartz 1982) devote approximately half of their book to arguing that the above relationship is indeed real – that it has the characteristics of a phenomenon and is not an artifact of various statistical and measurement assumptions they employ. They investigate the relationship by regressing various measures of technical innovation on various measures of firm size and market concentration, the underlying assumptions about functional form being supplied by economic theory. They show that the relationship is relatively robust under different assumptions about how to measure these quantities and that it is fairly constant and stable across different industries. It is only in the second half of their book that the authors turn their attention to what they call 'theoretical explanation' and, as the argument of this paper would lead one to expect, it is this phenomenon and other related results, and not the data from which they have been detected, which are regarded as appropriate objects of theoretical explanation.

[2] This claim is defended in more detail in Bogen and Woodward (1988).

[3] Empirical evidence for this claim is provided in Collins (1975), for example.

[4] The use of the model of signal and noise to describe problems of phenomenon-

detection is extremely widespread among both natural and social scientists – see for example, the discussion of Joseph Weber's unsuccessful attempts to detect gravitational radiation in Section 5 below or Herbert Asher's remarks about detecting voters' political attitudes in Section 9 below.

[5] See for example, the discussion of research on political attitudes in Section 9 below. The importance of this point for the construction of explanatory theories in the social sciences is emphasized by Roy D'Andrade in a recent paper.

> What sciences generally explain is a regular phenomenon of some sort – like the acceleration of falling bodies or the elliptical orbits of the moon and the planets.... It is easier to develop and apply an explanatory account when one knows what it is that one is explaining.
>
> Thus a more appropriate task for psychology and the social sciences is to account for the kinds of regular phenomena we can observe naturally or produce in the laboratory. It would be a better test of progress if psychologists and social scientists held themselves accountable for being able to explain or predict such phenomena as the relation between number of items presented to subjects and the accuracy of recall or the relation between prestige and wealth, not such idiosyncratic events as whether Mrs. Puffaway will quit smoking or whether Mr. Seemsodd will have a psychotic break or whether Someland will declare war on Someotherland. (D'Andrade 1986, p. 28–29)

[6] There are a number of different versions of the doctrine of inference to the best explanation, and only some of these are inconsistent with the claims I make below. One version of the doctrine merely claims that if E is evidence for the existence of phenomenon P, there must exist a causal connection of some appropriate sort between E and P. Thus it might be claimed that if tracks in a bubble chamber are to be evidence for the passage of a proton through the chamber, the proton must figure causally in the production of the tracks. Nothing I say below is inconsistent with this claim.

This first version of the doctrine of inference to the best explanation amounts to a contention about the truth conditions for evidential claims. It should be distinguished from a second version of the doctrine according to which is understood as a contention about legitimate grounds for belief or for hypothesis-choice. According to this second version, the fact that a hypothesis would, if true, provide a good explanation of some body of evidence is, by itself, grounds for belief in the hypothesis – indeed, all grounds for belief take, at bottom, this form. The idea is (roughly) that in determining whether evidence supports hypothesis H one needs to consider whether H, if true, would provide a better explanation of E than any competing hypothesis. If this is the case, and only if this is the case, then one is entitled to believe H on the basis of evidence E. Something like this version is defended by, for example, Gilbert Harman in his (1965).

Just what this second version amounts to in practice depends, of course, on what sort of criteria for explanatory goodness one adopts. I claim that on at least one natural account of explanatory goodness – the account given below in Section 2, which emphasizes the role of unification and the systematic exhibition of dependency relations in explanation – not all inductive inference conforms to the pattern described in the second version of the doctrine. For example, in the neutral current experiments described below, the considerations which convinced the experimenters that certain bubble chamber photographs constituted evidence for neutral currents did not take the form of the exhibition of a detailed and systematic potential explanation of these photographs in which neutral currents figured, and an argument that this explanation, if its premises were true, would provide a better explanation of the photographs than any alternative.

[7] For a more extended defense of these claims see Woodward (forthcoming).

[8] For an accessible summary by a philosopher of science, see Rosenberg (1985, pp. 73–83). For a more detailed account, see Dickerson and Geis (1983).

[9] For defenses of this idea, see for example, Friedman (1974), Kitcher (1982), Glymour (1984), and Woodward (forthcoming).

[10] Attempts to provide purely formal and domain independent characterizations of explanatory unification tend to be unilluminating. However, within specific domains of investigation it is sometimes possible to provide scientifically interesting formal characterizations of explanatory unification. For example, within physics, unification is typically achieved via the satisfaction of certain symmetry and invariance requirements. Thus, in the case of the weak and electromagnetic force, unification is achieved by means of the imposition of the requirement that the unifying theory satisfy a local gauge symmetry. See Woodward (forthcoming) for further discussion.

[11] The difficulties of constructing detailed derivations of the behavior of complex systems, even when one has a good understanding of many of their parts taken in isolation are emphasized in the work of Nancy Cartwright and Ronald Laymon – see for example, Cartwright (1983) and Laymon (1985). Both writers also emphasize the central role of idealizations and approximations in facilitating derivability and computational tractability.

[12] For an example of an account which appears to assume that derivations of claims about data from such a super-theory will always be possible, see the discussion of Schaffner's account of correspondence rules (Schaffner 1969) in Section 6.

[13] For a description of the experiments at CERN in which the W and Z particles were discovered and more details about why electronic particle detectors rather than bubble chambers were used, see Watkins (1986).

[14] For details of this episode, see Galison (1985).

[15] The description that follows relies heavily on Peter Galison's superb paper on the discovery of neutral currents (Galison 1983). Relevant additional material, on which I have also drawn, is contained in Pickering (1984) and Crease and Mann (1986).

[16] As Galison puts it:

> ... among the hadron group's participants different types of evidence were given different weights. For example, some participants were persuaded by the relative number and distribution of associated events; some other collaborators were persuaded by the thermodynamic analysis, yet others by Monte Carlo simulations. Still others remained skeptical until the problems of the neutron cascade and kaon regeneration were fully understood (Galison 1983, p. 506).

Nonetheless

> using a variety of approaches, techniques and approximations, the members of the collaboration persuaded themselves that they were looking at a real effect. (Galison 1983, p. 440)

[17] Galison writes

> The reason precise predictions could not be calculated for the hadron punchthrough is related to the reason the Gargamelle group was having such a hard time calculating the neutron interaction length: both problems involved the passage through matter of strongly interacting particles. Strong interactions presented a much more difficult problem than the well-understood electromagnetic interactions involved, for instance, in a muon's passage through matter. Compounding the problem was the absence of good data on the energy and momentum distribution of the hadrons being produced. This was the first observation of high-energy neutrino reactions; and the composition of the reaction products had not been studied at all. (Galison 1983, p. 499)

[18] The reader who thinks that such explanations must be "possible in principle" if the experimenters really had reliable evidence for the existence of neutral currents is referred to Note 17 and to Section 4 below.

[19] If the experiment were instead conducted at the surface of the earth and one attempted to deal with the cosmic ray background by identifying and rejecting events due to cosmic rays, a trillion cosmic ray events would have to be rejected in the course of a year before one could expect to find a single proton decay. This example provides an additional illustration of an observation made above: even given an understanding in principle of the mechanism by which confounding events are produced and the ability in principle to identify individual confounding events, actually correcting for all of them may be a practical impossibility.

[20] Details of this example are taken from (Hughes 1985).

[21] A similar contrast in social scientific research between on the one hand, attempting to achieve reliability by explicit calculation and theoretical modeling of confounding factors and, on the other hand, by building control of confounders into the design of one's investigative procedures is discussed at length in Cook and Campbell (1979). They make it clear that the second strategy is preferable in many contexts.

[22] Cox, McIlwaith and Kurrelmeyer (1928), quoted in Franklin (1986, p. 210).

[23] The experimenters appeal to a similar argument at a number of other points in their discussion. For example, in arguing against the possibility that the asymmetry in scattering they observed was due to an 'accidental asymmetry' in their apparatus, they write

> The radium and the point in the counter were doubtless not exactly centered. But they were removed and replaced repeatedly in the course of the observations, and it seems unlikely that their accidental dislocations could be so preponderantly in one direction as are the observations. Any effect due to this cause could be offset by turning the counter and the rod that carries the radium through 180°. (quoted in Franklin 1986, p. 210)

Here too the point is that while there may have been asymmetries in the apparatus on any particular occasion of use, the repeated changes in the source and detector made it plausible that these errors were distributed symmetrically and unlikely that any repeatedly observed asymmetry in scattering was simply an artifact. Here also there is an obvious difference between this sort of strategy and a strategy which attempts to correct for possible errors due to instrumental asymmetry by means of measurement and derivation.

[24] For a description see for example, Collins (1975).

[25] For an account of the events involved in this discovery, see (Crease and Mann 1986). Ting's description of some of these tests nicely illustrates the way in which physical manipulation of an apparatus can be used to check for artifacts.

> "There were many, many tests, okay?" Ting told us. "The first is to change the thickness of the target, let's say by a factor of two, and see whether the counting comes off by a factor of two. If it's a scattering from the side of the magnet or something, it won't. Or change the magnet current by ten percent. A true peak better show up in the same place, and not move. Or change both magnets to positive polarity, and see what happens. Plug up the magnet to a smaller aperture, and see whether the rate changes. All kinds of tests." (quoted in Crease and Mann 1986, pp. 374–75)

[26] Both philosophers and historians of science often write as though 'external', explanations which appeal to sociological factors and 'internal' explanations which appeal to the role of evidence, and methodology in influencing scientist's behavior yield competing, mutually exclusive accounts. But in the case described above, as at a number of other points in my discussion, we see how a 'sociological' factor (having to do with competition in science, and the desire to be first in making a discovery) and methodological considerations do not compete but rather illuminate and reinforce one another. Of course both Ting and Richter at SLAC were heavily influenced by the "sociological" factor of not wanting to be second in announcing an important discovery. But once each knew of the other's results, each was probably justified, on 'internal', methodological grounds, in claiming the existence of a new particle.

[27] For example, an internal memorandum circulated within the CERN group describes the NAL result as "an apparent lack of neutral current type events" and then adds that "in the near future, we can expect to be heavily questioned about the reliability of our experiment" (quoted in Galison 1983, p. 499).

[28] For details of this, see Pickering (1981). One of the principal advocates of the alternative heavy nucleus interpretation was Luis Alvarez. In his recent autobiography (1987), he makes it clear that the fact that he had previously conducted a number of unsuccessful searches for monopoles led him to be sceptical of Price's claims from the outset.

[29] Compare also Heinz Pagels' remark in his recent book *Perfect Symmetry* (Pagels 1985) regarding the possible detection of a magnetic monopole by means of a super conducting ring set up at Stanford. Although "the current in the ring jumped from zero to precisely the value expected if a monopole had gone through the ring",

> [t]he event was never repeated. Perhaps the current *was* caused by a monopole and monopoles may be so rare that this was a lucky catch. However, physics is based on reproducible results. Until the experiment can be repeated, most physicists will think of the event as a fluke. (Pagels 1985, p. 289)

[30] As Franklin puts it

> The consistency of the data argued for their validity and against their interpretation

as artifacts. No remotely plausible malfunction of the apparatus could produce such a consistent result (Franklin 1986, p. 169).

[31] Details are provided in Pinch (1986).

[32] For details see Watkins (1986).

[33] An illustration: a paper describing relatively early results from the NAL weak neutral current experiment sent to *Physical Review Letters* as criticized by referees on the grounds that the statistical techniques employed were, as Galison puts it, "insufficiently conservative" (p. 496). In effect, the NAL group calculated the probability that the detected ratio, R, of charged to neutral current events would have been obtained, given the assumptions of the pre-Weinberg-Salam theory of weak interactions (which predicted the non-existence of neutral currents) – i.e., the probability that the detected ratio R was due to a chance fluctuation and that the older theory was true. The referees wanted the authors to base their conclusions on the uncertainty in the value of R they obtained instead (Galison 1983, p. 496).

Another illustration is provided by Allan Franklin (1986). Experimental confirmation of parity non-conservation was reported in three different papers published in 1957. Two of these papers obtained at least a thirteen standard deviation effect and were published immediately; the third was initially rejected because the results obtained involved only a two deviation effect ("An effect of less than three standard deviations is quite insufficient in such an important and subtle effect experiment", the editor comments) and was only accepted when the investigation obtained additional, more compelling evidence.

[34] For interesting discussions of problems of data-reduction in high energy physics, which emphasize the quite different problems associated with the use of electronic particle directions and visual detectors like bubble chambers, see Galison (1985), Pickering (1984), and Watkins (1986).

[35] The philosophy of science literature is full of imagined examples in which a physicist sees certain tracks in a cloud chamber or bubble chamber and then immediately identifies them as due to this or that particle. Such examples are sometimes said to involve 'seeing' the particle in question and are used to motivate very general claims about the theory-ladeness of observation and the difficulties of distinguishing sharply between 'theoretical' and 'observational' claims. Whatever plausibility this may have as a characterization of simple cloud chamber observations, it represents a quite misleading picture of how information regarding elementary particles is obtained from sophisticated detectors like bubble-chambers. In bubble-chamber experiments, theoretically interesting events are often quite rare: to have human observers watch for them is hardly a good strategy for detecting them. Often not only the events themselves but the tracks they produce take place much too quickly or over spatial dimensions which are much too small to be detected by the unaided human visual system. Moreover, trajectories of electrically neutral particles like neutrons produce no tracks – no visual record – at all; instead their presence is inferred from other tracks and theoretical assumptions like conservation laws. Once a permanent record of the data is obtained in the form of photographs, procedures for data analysis are extremely time-consuming and complicated, involving complex measurement techniques, and various statistical and curve fitting procedures. Often these procedures are carried out by different people or in part by machine. The individual results of such analyses often will be aggregated,

and then subjected to various background corrections and further statistical analysis. They may then be compared with results obtained under different methods of detection or under varying initial conditions. Thus, even in the case of detectors like bubble-chambers which produce data in the form of visual images, perception by a single agent (even highly theory mediated perception) is not a very useful picture for what is involved. And of course other kinds of detectors – for example, electronic particle detectors – do not even produce data which one might naturally regard as visual recordings or images of particles or trajectories. Here the data take the form of patterns of discharge in the detectors or perhaps computer printouts which summarize these, and it seems even less helpful to regard obtaining information about a particle by means of these records as tantamount to (theory laden) visual perception of the particle.

In general, the extent to which the processes by which a phenomenon is detected resemble paradigmatic cases of direct visual perception (or direct perception by any other sense) is not a good guide to whether those processes are likely to yield reliable information about the phenomenon. The use of detectors that produce visual records which do not at all resemble the phenomena detected, or which require lengthy complex and indirect procedures of information extraction, can often increase the reliability of the detection procedure even as they reduce its similarity to paradigmatic cases of vision. In trying to understand when and why it is legitimate to claim that one has detected a phenomenon on the basis of a body of data it is much more illuminating to focus on the notion of reliability directly, rather than on whether the detection process can be assimilated to ordinary visual perception.

[36] Thus Rutherford wrote to Neils Bohr:

> The idea that you can discriminate between slow alpha particles and H particles by the intensity of the scintillation is probably the cause of their [the Vienna group] going wrong. Under the normal conditions of experiment such discrimination by eye is terribly dangerous. (quoted in Steuwer 1985, pp. 268–69)

[37] Another case: as Irving Klotz describes in his 'The N-Ray Affair' (1980), the French physicist Rene Blondlot claimed to detect a new kind of radiation, the N-ray; his evidence was a supposed characteristic increase in brightness in the spark produced by an electric discharge device when the rays were incident on it. When other physicists failed to detect this effect, suspicions were aroused. Klotz describes how on a visit to Blondlot's laboratory, R. W. Wood interposed his hand (which would supposedly block N-rays) between the N-ray source and the detector at irregular intervals in a darkened room (so that the movements were not observable by Blondlot and his colleagues). The result of this was that there was no correlation at all between the variations in brightness Blondlot and his co-workers claimed to detect and the movements of Wood's hand. ("Visual judgements of light intensity are notoriously unreliable", Klotz comments). Various other manipulations of the source by Wood also failed to produce corresponding variations in the observed brightness of the detection device. Here too one sees how empirical investigations of the reliability of human visual perception under certain conditions can be carried out in the absence of an explanatory theory of the operation of the human visual system.

[38] Weber's experiment and the reactions of other scientists to his work are described in

Collins (1975) and (1981). Other accessible discussions of Weber's experiment on which I have relied include Davis (1980, especially pp. 102–117), and Will (1986).

[39] An additional, more 'theoretical' consideration which at first seemed to support Weber's claim to have detected gravitational radiation should also be mentioned here, both for the sake of completeness and for its intrinsic interest in illustrating the complex interaction between theory and experiment. The detectors used by Weber were most sensitive to gravitational radiation when the direction of propagation was perpendicular to their axes. Thus if the waves were coming from a fixed direction in space (as would be plausible if they were due to some astronomical event), they should vary regularly in intensity with the period of revolution of the earth. Moreover, any periodic variations due to human activity should exhibit the regular twenty-four hour variation of the solar day. By contrast, the pattern of change due to an astronomical source would be expected to be in accordance with the sidereal day which reflects the revolution of the earth around the sun, as well as its rotation about its axis, and is slightly shorter than the solar day. When Weber initially appeared to find a significant correlation with sidereal, but not with solar time in the vibrations he was detecting, this was taken by many other scientists to be important evidence that the source of the vibrations was not local or terrestrial, but instead due to some astronomical event. Conversely, when this 'sidereal correlation' disappeared as Weber continued his experiment and did further analysis of his data, this undermined the confidence of many that he had detected gravitational radiation. As one researcher, quoted by Collins, put it

> The major thing that dissuaded everybody though – that eventually killed it off – was that no more talk or mention had been made of the sidereal connection. (Collins 1981, p. 41)

Here we see how the fact that a putative phenomenon is detected as having just those characteristics which, on theoretical grounds, one would expect it to have, can be important evidence that the one's detection device is working reliably, that possible sources of error have been adequately controlled for, and thus that the phenomenon is genuine and not an artifact. Conversely, failure to detect a phenomenon with the expected characteristics can be evidence that one's experiment is infected with some unknown source of error and that one's data are unreliable. Additional examples of this strategy are described in both Galison (1982) and Franklin (1986). As both Galison and Franklin remark, while there need not be anything perniciously circular about the use of this strategy (it does not guarantee that one will detect a phenomenon with the expected characteristics, regardless of the way the world is), it nonetheless is probably an important source of experimental error in a number of historically significant cases. If one holds a mistaken theory about the phenomenon one is trying to detect and systematic error is present which produces apparent evidence for a phenomenon with those characteristics, one may mistakenly conclude that no such error is present and that one's experiment is reliable. Nonetheless, the general utility of using the fact that an experiment produces a phenomenon with expected characteristics as itself evidence for reliability is obvious, given the great difficulties of directly detecting and eliminating all possible sources of error in many experiments.

[40] See for example, Feynman (1985, especially pp. 338–46).

[41] Compare the remark of Heinz Pagels quoted in Note 29.

[42] This methodological point has interesting implications for a number of areas of scientific investigation. For example, as James Bogen and I note (Bogen and Woodward 1988), although experiments in parapsychology have produced a great deal of data, it is controversial whether they have produced any reliable evidence for parapsychological phenomena. Critics have challenged the reliability of parapsychological experiments on the grounds that the results of such experiments are not replicable, that there has been inadequate control for confounding factors like unconscious cuing or fraud, and that defective statistical techniques have been used. Until there is reason to think that there is reliable evidence for parapsychological phenomena, it is premature to attempt to construct systematic explanatory theories in parapsychology. For an additional application of this idea, see Section 9 below.

[4] As Andrew Pickering remarks:

> One important distinction is that between theorists and experimenters. Theory and experiment constitute distinct professional roles within [High Energy Physics]. Each form of practice is highly technical, drawing upon quite different forms of expertise, and it is rare to find an individual who successfully engages in both. (Pickering 1984, p. 22)

[44] As a further illustration, consider the contents of a typical recent collection of methodological essays in the social sciences entitled *Theory Building and Data Analysis in the Social Sciences* (Asher, Weisberg, Kessel and Shively 1984). The first part, on theory building, contains material and is animated by concerns which most philosophers of science will find familiar. There is, for example, the obligatory chapter on philosophy of science ('From Paradigms to Research Programs') and papers on formal models in political science and social choice theory. Here the concern is with the construction of explanatory theories and with criteria appropriate to the assessment of such theories. The second part, on data analysis, has a quite different focus of concern, which will be much less familiar to most philosophers of science. (No chapter on relevant work in philosophy of science here.) Here the characteristic concern is, in effect, with procedures for extracting information from data or with finding stable patterns in data which might be relevant to the existence of phenomena. Here there are discussions of the following sort: (1) What (if any) are the appropriate measures of central tendency and dispersion when one has data measured on a nominal or ordinal scale? What are the different possible measures of association in connection with such data and what are the advantages of each? (2) What are the properties of general linear regression models? What are the procedures for estimating the coefficients in such models and what conditions must be satisfied to be reliable? (3) Given data about roll-call votes in Congress under what conditions can one take such data to provide information about the underlying dimensionality of the 'issue space' which is reflected in these votes. For example, what sort of patterns of votes would justify one in concluding that these votes indicated position along a single left-right continuum?

In each of these discussions, the aim is neither the reporting of raw data nor (for the most part) the construction of explanatory theory. Rather, the characteristic concern is with the development of general techniques for the extraction of claims about phenomena from data, claims which will then become objects of explanation by the theories or models described in the first part of the collection.

[45] The traditional discussions of testability of course often acknowledge the point that there may be (as they are called) 'practical' or 'technological' limitations on our ability to test certain claims. However, the usual idea is that (as the standard notion of correspondence rules suggests) even in such cases a respectable theory ought to describe in some detail the data and experimental procedures that would be relevant to testing its claims – it is just that the procedures may not be such that they can be or will be carried out by us. This notion of 'merely practical' difficulties in testing is itself full of obscurities, but even if it is allowed, my discussion above is intended to suggest an important range of cases which this notion does not cover: cases in which it is simply unclear, at the time a scientific theory is first constructed, which if any procedures or data are likely to yield reliable evidence concerning its claims about phenomena.

[46] Our discussion above emphasizes difficulties in reliably assessing phenomena-claims. It is worth noting the existence of another characteristic difficulty in the testing of explanatory theory which is ignored in analyses of theory structure like Nagel's. An appreciation of this feature further underscores how wrong-headed it is to think that a theory automatically contains a set of rules connecting theoretical claims with data claims. In the case of a number of theories, there are great difficulties in deducing many determinate phenomena-claims of a sort which are even candidates for empirical testing and which can be used to discriminate between the theory and competitors. This can be due to mathematical or computational intractabilities, or the absence of suitable approximation techniques, or to divergence difficulties (as in the case of many theories in particle physics prior to the discovery of appropriate renormalization techniques) or to uncertainty about how to apply the theory to new situations, or to conceptual and mathematical confusions of various kinds. Learning to understand what a theory implies about some phenomenon is, as remarked below, a characteristic preoccupation of theorists and is frequently a long and laborious process, full of missteps. (For an instructive and characteristic example of this, involving various attempted derivations of the gravitational red shift from general relativity, see Earman and Glymour (1980).)

So connecting theory and observable data can be difficult and problematic for two distinct reasons: (1) deducibility difficulties and (2) difficulties in finding reliable detection techniques of the sort described in the main body of this paper. Both of these possibilities (and the general point that the establishment of connections between theory and data is not an automatic matter, but depends on a conjunction of favorable, quite contingent circumstances) are nicely illustrated in a discussion of Stephen Weinberg's concerning theories of the four fundamental forces. Weinberg remarks that (as of 1974) the general problem with testing theories of the strong interaction has to do with calculational difficulties.

> For the strong interactions there was no lack of possible renormalizable theories; rather, the trouble was (and indeed still is) that strength of the interaction invalidates any single approximation scheme that might be used to draw consequences for any given field theory that might be checked by experiment. (1974, p. 53)

By contrast, the problem in the case of theories of gravitational interaction is (at least in part) one of experimental detection

> The problem here is the opposite of that for the strong interactions: gravitational effects [over very short distances, for which general relativity is assumed to require

modification] are so weak that one can get no help from experimental measurements, at the current level of precision, in finding correct theory. (1974, p. 54)

The weak interactions represent a case in which happily both these characteristic difficulties can be overcome and experimental testing is possible.

> The weak interactions present an intermediate case: they are strong enough so that good experimental data are available... and yet weak enough so that approximate calculations are practicable. (p. 54)

Both of the above difficulties are insufficiently emphasized in traditional discussions of theory structure and testing.

[47] For some recent historical and sociological studies, that suggest a substantial amount of independence of the development of explanatory theory and experiment, see Pais (1986), Pickering (1984) and Galison (1985).

[48] The considerations which follow are typically not regarded as reasons for belief in the truth of the theories in question. They are rather regarded as reasons for pursuit, reasons for taking theories of a certain general sort seriously and for exploring and developing them to a point at which they are susceptible of detailed experimental test.

[49] This sort of consideration was quite important in connection with the development of theories that unify the weak and electromagnetic forces and more recently in connection with theories that attempt to unify the electroweak and strong forces.

[50] This was one of the central motivating considerations when, after the invention of elementary quantum mechanics, theorists like Dirac turned their attention to the development of a consistent relativistic version of the theory.

[51] The long struggle to develop a renormalizable electroweak theory, culminating in the realization that gauge theories in which vector bosons are given masses by the Higgs mechanism are renormalizable is a case in point. The discovery that gauge theories are asymptomatically free and hence can furnish realistic theories of the strong interactions represents another recent example. The general point is that these developments mainly involved the exploration of the conceptual and mathematical structure of various general classes of theories with little eye toward detailed applications to experimental phenomena. It was only when these conceptual and mathematical issues were resolved – when, for example, theorists discovered how to construct unified electroweak theories which were renormalizable – that attention seriously turned to the question of which theory in a general class was most phenomenologically adequate.

[52] For discussion see Will (1986), especially chapter one.

[53] For relevant discussion see Pais (1986). Speaking of experimental investigations of various forms of radioactivity, Pais writes

> [these investigations] were principally the concern of a fairly modest-sized but elite club of experimental radioactivists. In those days, theoretical physicists did not play any role of consequence in the development of this subject, both because they were not particularly needed for its descriptive aspects and because the deeper questions were too difficult for the time. (p. 104)

Elsewhere, commenting on the period from 1895 to 1905, Pais emphasizes the relatively independent development of experiment and theory.

Another characteristic of the tumultuous decade is that experimental advances took place on one front, those in theory on an almost second front. Thus Planck's theory addressed blackbody radiation and (during those early years) nothing else, relativity gave an underpinning to Maxwell's electromagnetic theory, yet, in spite of its recognized universal nature, could not at once be brought to bear on other kinds of observations. On the other hand, radioactivity represented a class of phenomena in search of a theory. (p. 137)

[54] As Fred Suppe puts it, in the course of summarizing Schaffner's proposal

> ... the causal sequence correspondence rules specify causal links between states described by TC [the conjunction of theory and correspondence rules] and by observation reports, and so provide theoretical explanations of the behavior reported in the observation reports. In so doing the correspondence rules are scientific laws, generally being in the form of theories other than TC, used as auxiliary hypotheses in applying the theory TC to phenomena. (Suppe 1977, p. 105)

[55] For the use of this terminology, see, for example, Mynatt, Doherty and Tweney (1977) and Klayman and Ha (1987). The idea that both falsificationist and confirmationist inductive strategies may be appropriate in different situations, depending on the empirical details of the inductive problem one faces, is suggested by a number of writers, although I know of no attempt to relate this insight to the distinction between data and phenomena.

[56] For example, this was the general strategy followed in connection with the testing of the Weinberg-Salam theory. Once that theory was shown to be renormalizable, attention focused on a novel prediction made by that theory – the existence of neutral currents – which was widely believed to be implausible and which differed sharply from predictions made by previously accepted theories.

[57] As another illustration of the same general point, consider the fact about which there again seems to be gneral agreement, that data analysis in many areas of science often has a post hoc or ad hoc flavor and often involves 'massaging' (cf. p. 423 above) or 'mining' data for significant-looking results. While some kinds of data mining are clearly objectionable, my guess is that a certain amount of ad hoc massaging of data in phenomena detection in both salutary and unavoidable. Applying the expectations generated by a rigorously falsificationist methodology of science to all activities of data analysis would be misguided, even if those expectations are appropriate in connection with the assessment of explanatory theory.

It is worth adding that the differences in the inductive strategies thought to be appropriate in connection with the assessment of explanatory theory and the detection of phenomena also shows up in the quite different attitudes taken toward mistakes in these two areas. At least in physics, it is generally thought to reflect badly on an experimentalist if he claims to have detected a phenomenon which does not exist. On the other hand, it seems generally to be true that much less (if any) stigma attaches to the theorist who invents or puts forward a theory which turns out to be false. In the case of theorists, Popperian bold conjectures are at least somewhat encouraged; similar behavior among experimentalists in connection with claims about phenomenon-detection are strongly discouraged.

[58] My discussion here is indebted to Deborah Mayo's account (Mayo 1986) of the logic

of Jean Perrin's well-known experimental determination of Avogadro's number and his arguments for the correctness of kinetic theory (Perrin 1923). As Mayo and Nancy Cartwright (1983) point out, a number of philosophical commentators seem to have misunderstood Perrin's argument, mistakenly interpreting it as an argument from coincidence for the correctness of kinetic theory. Perrin does appeal to the fact that a number of different experimental procedures (involving, for example Brownian motion, electrolysis, and alpha decay) for determining Avogadro's number all yield very nearly the same result. Perrin also claims that it is very implausible that this agreement is a coincidence. That is, he takes this agreement to effectively rule out the possibility that each of the values of Avogadro's number obtained by these different procedures represents an experimental artifact, with results that just happen to agree with each other. In the language of this paper, he takes the agreement to show that the different experimental procedures have detected a common phenomenon. Perrin recognizes, however, in accordance with the argument given above, that this agreement among different experimental determinations of Avogadro's number is not by itself convincing evidence for the correctness of kinetic theory. Instead he also considers the predictions that follow from possible alternative theories, which are significantly different from classical kinetic theory. He contends that if any of these alternatives, (which would presumably claim that nature is continuous, and that atoms and molecules do not exist) were correct, there would be no reason to suppose that the formula derived by Einstein for the mean displacement of a particle in Brownian motion would be accurate; the experimental verification of the displacement formula is thus a test which discriminates sensitively between kinetic theory and these alternatives.

On this analysis, Perrin's argument is importantly different from Wesley Salmon's claim that

> If there were no such micro entities as atoms, molecules and ions, then these different experiments designed to ascertain Avogadro's number would be genuinely independent experiments, and the striking numerical agreement in their results would constitute an astonishing coincidence.
>
> ...[the] existence and characteristics [of these micro entities] – as specified by various theories – *explain* this 'remarkable agreement'. (1984, p. 220)

From the perspective of this paper, these remarks conflate two quite different inductive inferences – one from data to phenomena and one other from phenomena to general theory. The "remarkable agreement" and "astonishing coincidence" to which Salmon appeals do represent good reasons for taking the various experimental determinations of Avogadro's number to have detected a genuine phenomenon but do not by themselves represent good reasons for belief in the approximate correctness of atomic or kinetic theory.

[59] Among recent philosophers of science the craft character of scientific work is particularly emphasized in Robert Ackermann (1985). I now think that I too quickly dismissed this aspect of Ackermann's discussion in my (1986).

[60] Among those with first-hand scientific experience, the role of craft or tacit knowledge is emphasized by, for example, Michael Polyani (1958) and John Ziman (1978). For some illustrative remarks by a practicing scientist, consider the following anecdote, which is related by Robert Crease and Charles Mann (1986) in their semipopular history

of particle physics, *The Second Creation*. Crease and Mann conceived the idea of repeating Ernest Rutherford's classic experiments on alpha particle scattering which demonstrated the existence of the nucleus. They approached a well-known experimental physicist, Samuel Devons, for help in executing the project. Crease and Mann write

> We broached the idea one day to him and had the embarrassing experience of hearing a kind man attempting not to laugh in our faces; we had obviously made his day. He answered simply enough, but a guffaw kept creeping into the edges of his voice. 'In principle, the experiment is simple', he said. 'In practice, it would be nearly impossible. The main problem . . . is that experiment is a *craft*, like making an old violin. A violin isn't a very complicated-looking gadget. Suppose you went to a violin maker and said, "Could you kindly help me make a Stradivarius? I'm interested in violin-making and I'd like to see how it was done." He'd smile at you just like I did. Because craft is a knowledge you have in your fingertips, little tricks you learn from doing things, and they don't work and you do them again. You have little setbacks, and you think, how can I overcome them? And then you find a way. Every time your equipment changes you forget all the old techniques and have to learn new ones.' (p. 336, emphasis in original)

[61] See for example, Pais (1986 p. 478). Commenting on the use of high energy accelerations, Pais writes

> Even with all due regard for precision, the coaxing of a high energy beam through a donut remains part craft, part art. There does exist a vast body of theoretical analysis of stability against wiggles and wobbles. The problem is so delicate and complex, however, than often the most unforseen perturbations arise during the first test runs of a new machine. To subdue these, there lies the art. (p. 478)

[62] See for example, Feynman (1985). Commenting on the 'chaotic' experimental situation with respect to beta-decay shortly after the discovery of parity non-conservation, Feynman writes

> . . . I asked some of the experimenters what the situation was with beta decay. I remember three guys, Hans Jensen, Aaldert Wapstra, and Felix Boehm, sitting me down on a little stool, and starting to tell me all these facts: experimental results from other parts of the country, and their own experimental results. *Since I knew those guys, and how careful they were, I paid more attention to their results than to the others*. Their results, alone, were not so inconsistent. (p. 250, my emphasis)

Feynman's description (pp. 253–55) of experimental results obtained by Telegdi, which at first appeared to be inconsistent with the V-A Theory of the weak interactions devised by Feynman and Gell-Mann and which were then re-analysed by Telegdi himself and found to be consistent with the V-A theory reveal a similar theme: considerable importance is attached to the experimenter's general reputation ("I knew Telegdi was excellent") and both Feynman and Gell-Mann agree that if there is an error in the experiment, the experimentalist himself is much more likely to find it than they are. ("I was convinced by that time that something must be wrong with his experiment and that he would find it – he's much better at finding it than we would be. That's why I said that we shouldn't try to figure it out but just wait.")

[63] Philosophers who are inclined to doubt this claim should read through the description of the neutral current experiments in Section 3 above. Additional examples are provided in Holton's (1978) account of Millikan's oil-drop experiments and Pickering's (1986b) account of a series of quark search experiments conducted by Marpurgo in 1966–1970. The latter experiments afford a particularly striking illustration of the general point. The underlying idea of the experiments was similar to that of Millikan's oil-drop experiments. A sample (a grain of graphite) whose charge is to be measured is suspended in a magnetic field and a horizontal electric field generated by two parallel capacitor plates is applied. Quarks have fractional changes (±1/3e or ±2/3e) and evidence for fractional change on the sample would thus be evidence for free quarks.

As Pickering notes, a major source of error in this kind of experiment, "acknowledged by all who have attempted such measurements" has to do with possible inhomogeneities in the electric field, which can produce spurious evidence for fractional charges. Such inhomogeneities are due to "aparallelism of the plates, or to electric dipole layers on their surface". These "interact with permanent and induced electric dipole moments of the ball to produce forces that mimic those due to fractional charges" (Pickering 1981b, p. 220). Marpurgo and his colleagues used classical electromagnetic theory to attempt to estimate the electric field inhomogeneities in their apparatus, but remark that they cannot rely on the estimates, because "they are clearly affected by several uncertainties" (quoted in Pickering 1981a, p. 221). Instead, the experimenters adopted the procedure of increasing the distance between the plates on the assumption that this would reduce inhomogeneities in the electric field and thus any evidence for spurious charges. While they give a theoretical justification (with heavy reliance on idealizations and approximations) for this procedure, they also make it clear that they discovered the procedure itself by observing the behavior of their apparatus in the course of performing the experiment. They also make it clear that this observed behavior constitutes an important part of their justification for the procedure. In particular, that apparent fractional charges on various grains disappeared as the distance between the plates was increased was taken to show that increasing this distance would help to eliminate inhomogeneities and thus to distinguish real from spurious charges. As the experimenters remark, their original theoretical analysis had supported just the opposite conclusion:

> The question, of course, naturally arises of why we have not always operated from the start keeping the platelets [capacitor plates] at a very large distance so as to avoid the need to apply the prescription [concerning movable plates]. We plan to do this in future and all the more recent measurements are being done at a reasonably large distance; but in the set of measurements about which we are now reporting this was not done systematically (that is from the beginning) due to the following reason: we were led to think to the possibility of the Volta force effect [due to inhomogeneity of the field], while we were working in the static version, keeping the platelets at a small distance (1.4 mm). *Precisely it was seen that an apparent residual charge effect, present in some cases, did disappear when operating on the same grain at a large (double) distance*. It was felt necessary to explore how general this behavior was, both in the static and in the resonance version in order to learn more of the spurious charge effects.
>
> It should be added that our initial tendency was at work at a relatively small

distance between the platelets: this tendency was motivated by the argument that the gradient of the applied field should be smaller in this case; it was realized only later that this argument was not only false for small separations between the platelets due to the formation of irregular graphite depositions on the platelets, but also is irrelevant for the Volta force, the main part of which does not depend on the gradient of the applied field. (quoted in Pickering 1981b, p. 220, emphasis in original)

The general point is that both the extent of the error due to electric field inhomogeneity and the means for dealing with it were discovered by the experimenters in the course of interacting physically with their apparatus and by observation of the results (disappearance of apparent fractional charges) produced when such interactions occur. As Pickering puts it,

> ... theoretical considerations of their apparatus being insufficient alone to determine their [experimental] procedures in advance, these procedures were stabilized by production of credible results. (p. 222)

[64] Robert Millikan's reliance on such clues in the experiments in which he measured the charge of the electron is emphasized in Gerald Holton's well-known study (1978). As Holton points out, Millikan used such clues to assign different weights to his data, reflecting supposed differences in quality, and to justify discarding some data. In describing his early experimental results, Millikan explicitly graded the quality of his data. He wrote

> The observations marked with a triple star are those which were marked 'best' in my notebook and represent those which were taken under what appeared to be perfect conditions. This means that we could watch the drop long enough to be very certain that it was altogether stationary: that we could time its passages across the cross-hairs with perfect precision and that it showed no apparent retardation in falling through the two equal spaces. The double-starred observations were marked in my notebook 'very good'. Those marked with single stars were marked 'good' and others 'fair'. (Quoted in Holton 1978, p. 53)

In his later measurements, Millikan continued his practice of grading his data on the basis of various clues provided by the character of the run and his practice of discarding data produced by unsatisfactory runs, but did not acknowledge this in his published papers. As Holton comments:

> [Millikan] evaluated his data and assigned qualitative indications on their prospective use, guided by both a theory about the nature of the electric charge and a sense of the quality or weight of the particular run. This practice is familiar to anyone who has done basic experimental research: In the midst of a run, one does respond to small clues of the extent to which the numbers one is recording do in fact stem from the phenomena being observed. (1978, p. 70)

For a rather different account of Millikan's experiments, which attaches considerably less weight to the role of such 'craft' considerations, and which claims instead that Millikan's procedures can be largely justified in terms of familiar, explicit methodological precepts, see Franklin (1986).

[65] Several additional observations about the role of personal warrant in science are perhaps in order here. First, many sociologists draw relativist conclusions about scientific knowledge from the prominent role that personal warrant plays in many areas of scientific investigation. But it is hard to see how these conclusions follow. In using the strategy of relying on reputation and personal warrant, one in effect treats the experimenter himself as an instrument or black box, whose reliability can be empirically investigated in just the same way, in principle, as the reliability of any other instrument (cf. Section 4). Just as the fact that a telescope has yielded reliable evidence in the past regarding astrophysical phenomena is often grounds for thinking that it is presently yielding reliable evidence regarding some new phenomenon, which is detected under similar conditions, so also for the experimentalist. And just as the use of this strategy in the case of telescopes seems to lend no support for relativism, so also for reliance on personal warrant in the case of the experimentalist.

Secondly, it is worth reiterating that the legitimacy of appeals to personal warrant depends upon whether there are good reasons for believing that the experimenter in fact possesses general character traits and abilities that make it plausible that he has performed a reliable experiment on the occasion under investigation. Of course it is not always true that an experimentalist's general reputation for reliability reflects his actual performance or that prominent and influential scientists are always reliable in the sense indicated. For example, as Allan Franklin has pointed out to me, Rubbia had a long history of spectacular but mistaken experimental claims, beginning with the bungled HPW experiment concerning neutral currents before his correct identification of the W and Z particles in 1983. Despite his prominence and influence, Rubbia's actual performance does not seem to support heavy reliance on his personal warrant. For details, see (Taubes 1986).

Finally, it is also worth underscoring the obvious point that the role assigned to personal warrant above is necessarily derivative and parasitic on the existence of procedures and strategies for assessing the reliability of experiments which do not just involve appeal to personal warrant. In order to assess an experimenter's past performance, one must of course be able to determine whether or not the results of his previous experiments were correct; this requires, in my view, that some community of experts exist who do not just appeal to the warrant of the experimenter in assessing those results.

[66] It is also worth emphasizing in connection with this point that even when a general explanatory theory successfully predicts the existence of some novel phenomenon, the detection or discovery of that phenomenon is often regarded as an extremely important achievement in its own right, quite apart from the support it provides for the general theory. For example, the existence and characteristics of the W and Z bosons detected at CERN in experiments in 1983 referred to above were predicted by the standard model of the electroweak interactions. According to Watkins (1986), confidence in this model, deriving from other sources, was so high among most physicists by 1982 that few doubted that the W and Z existed even before they had been detected. The successful detection of the W and Z thus did not result in a large increase in the credibility of the standard electroweak theory, since this was already high, but it nonetheless won an immediate Nobel prize for the primary experimentalists, Rubbia and van der Meer. This illustrates again how the detection and investigation of a phenomenon is an important

activity in its own right, and that its significance is not just ancillary to the testing of explanatory theory.

[67] For the case of electricity and magnetism see, for example, Heilbron (1982).

BIBLIOGRAPHY

Ackermann, R.: 1985, *Data, Instruments, and Theory*, Princeton University Press, Princeton, New Jersey.

Anderson, J.: 1980, *Cognitive Psychology and Its Implications*, W. H. Freeman, San Franscisco.

Asher, H.: 1983, 'The Research Process', in H. Asher, H. Weisberg, J. Kessel and P. Shively (eds.), *Theory-Building and Data Analysis in the Social Sciences*, University of Tennessee Press, Knoxville, Tennessee.

Bogen, J. and J. Woodward: 'Saving the Phenomena', *The Philosophical Review* **97**, 303–52.

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

Cline, D., A. Mann and C. Rubbia: 1974, 'The Detection of Neutral Weak Currents', *Scientific American* **231**, 108–19.

Collins, H.: 1974, 'The Seven Sexes: A Study in the Sociology of Phenomenon, or the Replication of Experiments in Physics', *Sociology* **9**, 205–24.

Collins, H.: 1981, 'Son of Seven Sexes: The Social Destruction of a Physical Phenomenon', *Social Studies of Sciences* **11**, 33–62.

Converse, P.: 1970, 'Attitudes and Non-Attitudes: Continuation of a Dialogue', in E. Tufte (ed.), *The Qualtitative Analyses of Social Problems*, Addison-Wesley, Reading, Massachusetts.

Cook, T. and D. Campbell: 1979, *Quasi-Experimentation: Design and Analysis for Field Settings*, Houghton Mifflin, Boston.

Cox, R. T., C. G. McIlwain and B. Kurrelmeyer: 1928, 'Apparent Evidence of Polarization in Beams of B-Rays', *Proceedings of the National Academy of Sciences* (USA) **14**, 544–49.

Crease, R. and C. Mann: 1986, *The Second Creation*, Macmillan, New York.

D'Andrade, R.: 1986, 'Three Scientific World Views' in D. Fiske and R. Shroeder (eds.), *Methatheory in Social Science*, University of Chicago Press, Chicago.

Davies, P.: 1980, *The Search for Gravity Waves*, Cambridge University Press, Cambridge, England.

Dickerson, R. E. and I. Geis: 1983, *Hemoglobin: Structure, Function, Evolution and Pathology*, Benjamin Cummings, Menlo Park, California.

Earman, J. and C. Glymour: 1980, 'The Gravitational Red Shift as a Test of General Relativity', *Studies in History and Philosophy of Science* **11**, 175–214.

Feynman, R.: 1985, *Surely You're Joking Mr. Feynman*, W. W. Norton, New York.

Franklin, A.: 1986, *The Neglect of Experiment*, Cambridge University Press, Cambridge, England.

Friedman, M.: 1974, 'Explanation and Scientific Understanding', *Journal of Philosophy* **71**, 5–19.

Galison, P.: 1982, 'Theoretical Predispositions in Experimental Physics: Einstein and

the Gyromagnetic Experiments, 1915–1925', *Historical Studies in the Physical Sciences* **12**, No. 2, University of California Press, Berkeley, pp. 285–324.
Galison, P.: 1983, 'How the First Neutral-Current Experiments Ended', *Reviews of Modern Physics* **55**, 477–509.
Galison, P.: 1985, 'Bubble Chambers and the Experimental Work Place', in P. Achinstein and O. Hannoway (eds.), *Observation, Experiment and Hypotheses in Modern Physical Science*, MIT Press, Cambridge, Massachusetts.
Glymour, C.: 1984, 'Explanation and Realism', in J. Leplin (ed.), *Scientific Realism*, University of California Press, Berkeley, 1970.
Hacking, I.: 1983, *Representing and Intervening*, Cambridge University Press, Cambridge, England.
Haçking, I.: 1981, 'Do We See Through a Microscope?', *Pacific Philosophical Quarterly* **62**, 305–22.
Hanson, N.: 1958, *Patterns of Discovery*, Cambridge University Press, Cambridge, England.
Heilbron, J.: 1982, *Elements of Early Modern Physics*, University of California Press, Berkeley, California.
Holton, G.: 1978, 'Subelectrons, Presuppositions and the Millikan-Ehrenhaft Dispute', reprinted in, *Historical Studies in the Physical Sciences* **9**, 166–224.
Holton, G.: 1978, *The Scientific Imagination: Case Studies*, Cambridge University Press, Cambridge, England, pp. 25–83, (page references are to reprinted version).
Hughes, I.: 1985, *Elementary Particles*, Cambridge University Press, Cambridge, England.
Kitcher, P.: 1982, 'Explanatory Unification', *Philosophy of Science* **48**, 507–31.
Klayman, J. and Y. Ha: 1987, 'Confirmation, Disconfirmation, and Information in Hypothesis-Testing', *Psychological Review* **94**, 211–28.
Klotz, I.: 1980, 'The N-Ray Affair', *Scientific American* **242**, 168–80.
Kuhn, T.: 1970, *The Structure of Scientific Revolutions*, 2nd ed., University of Chicago Press, Chicago.
Kragh, H.: 1981, 'The Concept of the Monopole: A Historical and Analytical Case-Study', *Studies in History and Philosophy of Science* **12**, 141–72.
Laymon, R.: 1985, 'Idealizations and the Testing of Theories by Experimentation' in P. Achinstein and O. Hannaway (eds.), *Observation, Theory, and Hypothesis in Modern Physical Science*, MIT Press, Cambridge, Massachusetts.
Losecco, J. M., Reines, R. and Sinclair, D.: 1985, 'The Search for Proton Decay', *Scientific American* **252**, 54–62.
Marr, D.: 1982, *Vision*, W. H. Freeman, New York.
Mayo, D.: 1986, 'Cartwright, Causality and Coincidence', in A. Fine and P. Machaner (eds.), *PSA* 1986, Edward Brothers, Ann Arbor, Michigan.
Myratt, D., M. Doherty and R. Tweney: 1977, 'Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inference', *Quarterly Journal of Experimental Psychology* **29**, 85–95.
Nagel, E.: 1961, *The Structure of Science*, Harcourt, Brace and World, New York.
Pagels, H.: 1985, *Perfect Symmetry*, Simon and Schuster, New York.
Pais, A.: 1986, *Inward Bound*, Oxford University Press, Oxford, England.
Perrin, J.: 1923, *Atoms*, trans. by D. C. Hammick, Van Nostrand, New York.

Pickering, A.: 1981a, 'Constraints on Controversy: The Case of the Magnetic Monopole', *Social Studies of Science* **11** 63–93.
Pickering, A.: 1981b, 'The Hunting of the Quark', *Isis* **72**, 216–36.
Pickering, A.: 1984, 'Against Putting the Phenomena First: The Discovery of the Weak Neutral Current', *Studies in the History and Philosophy of Science* **15**, 85–117.
Pickering, A.: 1984, *Constructing Quarks*, University of Chicago Press, Chicago.
Pinch, J.: 1986, *Confronting Nature*, D. Reidel, Dordrecht.
Rosenberg, A.: 1985, *The Structure of Biological Science*, Cambridge University Press, Cambridge, England.
Schaffner, K.: 1969, 'Correspondence Rules', *Philosophy of Science* **36**, 280–90.
Stuewer, R.: 1985, 'Artificial Disintegration and the Cambridge-Vienna Controversy', in P. Achinstein and O. Hannoway (eds.), *Observation, Experiment and Hypothesis in Modern Physical Science*, MIT Press, Cambridge, Massachusetts.
Suppe, F.: 1977, 'The Search for Philosophical Understanding of Scientific Theories', in F. Suppe (ed.), *The Structure of Scientific Theories*, University of Illinois Press, Urbana, Illinois.
Taubes, G.: 1986, *Nobel Dreams*, Random House, New York.
Thagard, P.: 1978, 'The Best Explanation: Criteria for Theory-Choice', *Journal of Philosophy* **75**, 76–92.
Travis, G.: 1981, 'Replicating Replication? Aspects of the Social Construction of Learning in Planarian Worms', *Social Studies of Science* **11**, 11–32.
van Fraassen, B.: 1980, *The Scientific Image*, Oxford University Press, Oxford, England.
Watkins, P.: 1986, *Story of the W and Z*, Cambridge University Press, Cambridge, England.
Weinberg, S.: 1974, 'Unified Theories of Elementary Particle Interaction', *Scientific American* **231**, 50–59.
Will, C.: 1986, *Was Einstein Right?*, Basic Books, New York.
Woodward, J.: 1986, Review of Robert Ackermann, *Data, Instruments, and Theory*, *Philosophy of Science* **53**, 455–58.
Woodward, J.: forthcoming, 'The Causal Mechanical Model of Explanation', in P. Kitcher and W. Salmon (eds.), *Minnesota Studies in Philosophy of Science*, volume 13.
Ziman, J.: 1978, *Reliable Knowledge*, Cambridge University Press, Cambridge, England.

California Institute of Technology
H&SS 101–40
Pasadena, CA 91125
U.S.A.