

Harold I. Brown
Philosophy Department
Northern Illinois University
DeKalb, IL 60115
USA
hibrown@niu.edu

From *The Cognitive Penetrability of Perception*, edited by A. Raftopolous, Nova Science Publishers, 2005, pp. 49-71.

On the Epistemology of Theory-Dependent Evidence¹

1. Clearing the Deck

In science and everyday life observational evidence provides a major source of information for evaluating hypotheses and beliefs. It is, however, now widely held that observation is “theory-laden,” where this term is used to cover a variety of ways in which our current concepts and beliefs play a role in determining what we observe. Theory-ladenness raises questions about the ability of observational evidence to provide grounds for genuinely independent evaluations of our beliefs, but a proper assessment of the epistemological significance of this dependence depends on how deeply observation is theory-laden. In a previous paper (Brown 1995) I distinguished six versions of the claim that observation is theory-dependent that have been invoked in the literature. I do not think that the list is complete, and I will not review it here. But I want to begin this discussion with an extreme version from that list, a view which holds that our beliefs are so deeply involved in observation that we fail to perceive items that violate our expectations.

This view is sometimes attributed to Kuhn (1962, often referred to as *Structure*), who provides a textual basis for the attribution in his discussion of Bruner and Postman’s (1949) anomalous-card experiment. In this experiment subjects are given brief glimpses of playing cards that they are asked to identify. However, the cards include some in which shape and color are mismatched, such as a red card in the shape of a spade. Initially, many subjects do not notice the anomaly. Kuhn writes (I quote Kuhn,

¹I want to thank Dr. Raymond Brock and Dr. Herman Stark for comments on an earlier draft of this paper, and Dr. Stark for assistance with an article written in German.

rather than the original paper, because my interest here is in Kuhn's use of the experimental outcome):

Even on the shortest exposure many subjects identified most of the cards, and after a small increase all the subjects identified them all. For the normal cards these identifications were usually correct, but the anomalous cards were almost always identified, without apparent hesitation or puzzlement, as normal. (1962: 63)

With increased exposure this easy identification broke down; hesitation and confusion ensued, "until finally, and sometimes quite suddenly, most subjects would produce the correct identification without hesitation" (63). Some of Kuhn's remarks about the relevance of this experiment for understanding scientific discovery are fairly cautious, other are less so. Consider a key example of the latter sort:

Either as a metaphor or because it reflects the nature of the mind, that psychological experiment provides a wonderfully simple and cogent schema for the process of scientific discovery. In science, as in the playing card experiment, novelty emerges only with difficulty, manifested by resistance, against a background provided by expectation. Initially, only the anticipated and usual are experienced even under circumstances where anomaly is later to be observed. (64)

Parts of this passage are subject to either a strong or a modest interpretation: The phrase, "novelty emerges only with difficulty" could mean that anomalies are just not noticed, or it could mean that anomalies are noticed, but their significance is recognized only slowly. However, the remainder of the passage suggests a stronger reading. In the course of this discussion Kuhn does state that we will "occasionally" see scientists behaving in ways that parallel the subject of the experiment (64), but he also provides support for the stronger interpretation when he returns to this experiment later in the text—in the context of a discussion of theory-laden experience: "Until taught by prolonged exposure that the universe contained anomalous cards, they saw only the types of cards for which previous experience had equipped them" (112-13). After referring to other literature from experimental psychology Kuhn

concludes:

What a man sees depends both upon what he looks at and also on what his previous visual-conceptual experience has taught him to see. In the absence of such training there can only be, in William James's phrase, 'a bloomin' buzzin' confusion'. (113)

Kuhn then compares the anomalous-card experiment to situations in which astronomers failed to correctly identify items that were not compatible with the prevailing paradigm, or, in some cases, failed to notice them at all. The latter include cases in which naked-eye Chinese astronomers reported phenomena that were not noted by their European colleagues (115-117). Thus Kuhn provides textual grounds for those who read him as holding that scientists see only what they expect to see. Note especially that Kuhn's invocation of the phrase from James has the effect of drastically limiting the significance of external inputs for the observer. What, we may ask, distinguishes one bloomin' buzzin' confusion from another? What role do the contents of this confusion play in determining which, if any, of our available concepts are to be used to structure this input? I will return to these questions below. At the moment I want to make three points about this strong version of the theory-dependence of observation.

First, the view that scientists typically do not notice anomalies is at odds with a major theme in *Structure*: that anomalies are not always counter-instances (e.g., 77-80). In 1962 this was a major departure from prevailing views since it implies that observational evidence and logical relations are not *sufficient* for assessing scientific hypotheses. Other considerations are required even when we have an observational result that is clearly incompatible with consequences of an accepted theory. This new view was emerging from several directions; it appeared almost simultaneously in work by Putnam (1962), Toulmin (1961), and Sellars (1963, an early version of this paper appeared in 1951). *Structure* is full of examples in which scientists who are fully aware of an anomaly attempt to resolve it within the current paradigm, or to otherwise assess its significance. Kuhn's point that anomalies are often resolved without

paradigm change is an important step in our understanding of science. He also discusses cases in which “the awareness of anomaly had lasted so long and penetrated so deep that one can appropriately describe the fields affected by it as in a state of growing crisis” (67); for example, “The state of astronomy was a scandal before Copernicus’ announcement” (67). Indeed, this awareness of anomaly was present in Ptolemaic astronomy from the beginning; it provided one source of research problems (68). Kuhn extends this point to other examples, maintains that theories are never completely successful, and emphasizes that exactly because paradigm-guided research generates precise, detailed expectations, the ability to detect anomalies is sharpened in normal science (e.g., 64-5). This is hardly a situation in which scientists fail to notice items that violate theory-generated expectations.

Second, while some cases in which scientists fail to notice items that would violate expectations do occur, if it is Kuhn’s considered view that this situation is typical, he is just wrong. *Structure* provides a rich source of evidence against this view, and the literature is full of other examples, such as the occurrence of varying atomic weights for elements in nineteenth-century chemistry, the photoelectric effect before Einstein’s theory, and more recent evidence which seemed to show that some galaxies are older than the universe. Franklin (2001) covers a century of experiment and theory involving neutrinos, and is full of recognized anomalies, alternative theories, challenges to theories, theories that survive such challenges, theories that are rejected as a result of observational evidence, and reinterpretations of observations.

There is also an ironical side to this point. Those who cite Kuhn as a source of evidence against the epistemic power of science often appeal to Duhem-Quine flexibility to argue that we can protect any belief in the face of apparent counter-evidence. But there is no point to this appeal unless it is acknowledged that expectations have been violated in a way that requires *some* adjustment in our beliefs.

Third, and most importantly, the anomalous-card experiment, with its brief glimpses of carefully contrived objects, does not provide a model for either normal perception or scientific observation. In our

everyday experience we do not just take quick glimpses of objects. Often we observe them over time and with input from more than one sense. In an analogous—but much more sophisticated—fashion, scientists evaluate a theory on the basis of multiple interactions with items in that theory’s domain. To be sure, there are cases in which a scientist’s data is limited to brief glimpses followed by memory-laden reports. This was the case for much of the early history of telescopic observation where, as a result of convection and the turbulence of the air, clear glimpses of objects being studied were rare. Newton was aware of the problem:

If the Theory of making Telescopes could at length be fully brought into Practice, yet there would be certain bounds beyond which Telescopes could not perform. For the Air through which we look upon the Stars, is in a perpetual Tremor. . . . Long Telescopes may cause Objects to appear brighter and larger than short ones do, but they cannot be so formed as to take away that confusion of the Rays which arise from the Tremors of the Atmosphere. The only Remedy is a most serene and quiet Air, such as may perhaps be found on the tops of the highest Mountains above the grosser Clouds. (1952: 110-11)

In a more recent discussion, focused on planetary astronomy, we read:

Even at the best sites, perfectly serene air is not attainable, and the passage of eddies in front of the telescope intermittently chops up the image and blurs critical details. The duration of the intervals between interruptions is typically such, in fact, as to make the limits of the perceptual system determine how much, and how accurately, the information can be extracted. The eye, while quick, does not record instantaneously. (Sheehan 1988: 95)

A bit later Sheehan notes just how brief the instances of clear vision may be:

The image of the planet—while generally more or less jumbled owing to these atmospheric disturbances—does nevertheless become sharp for an instant every now and

then, and in these instants the finer details may stand out with startling abruptness. What the observer is faced with is similar to what would be experienced by someone watching a motion picture in which the camera is out of focus except for an occasional sharp frame thrown in at random. (98-9)

These difficulties take on an additional dimension when we recall that, in earlier times, those not at the telescope had access only to verbal descriptions and drawings (see Sheehan 1988, especially Chs. 8 & 9). But many of these limitations are substantially reduced by newer technologies. The Hubble telescope is beyond the atmosphere; X-ray telescopes must be beyond the atmosphere, which blocks the radiation they detect; radio telescopes are not seriously affected by variations in the atmosphere; and adaptive optics allows for the construction of optical telescopes whose mirrors are adjusted to compensate for variations in the air. All of this aims at greater accuracy of observation, but there is another aspect of the technology that is of equal importance: the output from these instruments is not limited to what a single individual sees and remembers. Instead, the outputs are permanently recorded on photographs, video tapes, computer disks, and other long-lived media. As a result, brief subjective glimpses are replaced by stable, intersubjectively available outputs that can be examined, measured, and reexamined as reasons for doing so arise. While individuals may see differently when looking at a photograph, we are dealing here with a public object that provides a touchstone for discussing and evaluating varying impressions. In general, those who design observational procedures work at minimizing those points at which expectations and beliefs could infect the data, and eliminating cases in which the data are available only briefly, and only to a single observer, or small group of observers.

These remarks suggest that instead of focusing on the subjects of the anomalous-card experiment, *a better model of scientific data collection is provided by the activities of the experimenters.* They used 28 subjects, a specially prepared set of cards, and a controlled series of progressively longer exposures. They observed the subjects' behaviors, wrote down their remarks, analyzed the data, and

arrived at intersubjectively available evidence about how the subjects responded to brief glimpses of unexpected items. Consider where the kind of theory-dependence we are currently considering might have entered into the work of these experimenters. Perhaps their prior beliefs affected what they heard when subjects reported which card they were seeing, or perhaps it affected the experimenters' perceptions of the time settings on their tachistoscope or the card that was being shown. This is all quite unlikely, and in a more modern experiment the procedure would have been controlled by a computer, while the subjects' behaviors and remarks might have been recorded on video-tape. All of this would reduce the already minimal possibility that the experimenters' perception of their data would be colored by their beliefs and expectations. I will develop this theme at length in Section 2. First, however, I want to consider three more respects in which observation has been described as theory-laden.

One view holds that our concepts and beliefs play a role in the constitution of our sensory experience, so that they are already involved in the content of our experience (cf. *Structure* Sec. X). This view derives ultimately from Kant, although with some anticipation in Leibniz's remark that there is nothing in the mind that is not in the senses, except the nature of the mind itself (1985: 111). It is also a view that pervades Kuhn's writings, and that he explicitly adopted somewhat late in his career when he described himself as a Kantian with changing categories and unknowable things-in-themselves (1991: 12; cf. Brown 1975, Hoyningen-Huene, 1993). Two points will suffice by way of reply. *First*, the Kantian view does not transfer directly to the situations we are now considering. Kant held this view only for a limited set of presumably a priori concepts, and his argument for this view depended vitally on the claim that these concepts are a priori, universal, and unalterable. Changing concepts are not a priori in the relevant sense, so Kant's arguments do not apply. New arguments are required, and those that have been provided are empirical arguments from psychology and the history of science. The above discussion already suggests the limitations of such constitution, and we will encounter further limitations as we proceed. *Second*, no matter how deep my beliefs, expectations, and desires may be, the physical world

provides powerful constraints on what I can experience. Short of hallucination, nothing in my cognitive contributions to my percepts will turn the water in my glass into cognac, or prevent the glass from falling if I let it go. Kant would surely agree since in his view our cognitive contribution determines only the form of experience, not the massively variable content.

Consider next the view that for observational evidence to be relevant to the evaluation of a theory, that evidence must be described using the using the concepts of that theory. While this claim is correct, its significance is subject to three limitations. *First*, the conceptual framework of a scientific theory always occurs in the context of a much richer body of conceptual resources that we share with other members of our culture. As long as these resources include negation, we are well able to describe observations that contradict the consequences of a theory while using the concepts provided by that theory. *Second*, once we set up an observational situation, nature determines the outcome of the observation. Once we are past the view that we do not notice items that contradict accepted theory, the need to describe observations using the concepts of a theory provides no impediment to recognizing and reporting observations that either support or challenge that theory. *Third*, the history of science reminds us that the persistent pursuit of observational data sometimes provides unexpected results that cannot be accommodated at all by existing theory. Classic examples include the unanticipated discovery of sperm in the seventeenth century, and the successive discoveries of X rays in 1895 and radioactivity the following year; many other examples exist. The inability to describe such items using the concepts of existing theories did not prevent their being noted and their problematic status recognized. Discoveries of this sort provide one major impetus for generating new concepts and new theories.

Consider one more version of theory-dependence: which observations scientists make are determined by accepted theory. This is surely correct and important. The range of items we can explore is too large to be pursued without some guidance, and the issue becomes especially pressing when we deal with observational procedures that require expensive instrumentation. This kind of channeling plays

an important role in the development of science, and deserves an extended discussion that cannot be attempted here. For present purposes it will suffice to underline two points that have already been made: once we decide where we are going to look, our expectations do not pre-determine what we will find; and nature often impinges on us in ways that we did not anticipate.

2. Refitting

Traditional epistemology focuses on what occurs in the individual mind. Standard empiricism, which accords a central epistemic role to evidence acquired through our senses, follows this tradition by identifying empirical evidence with sensory percepts. Even some non-traditional forms of empiricism maintain this focus. For example, when Goldman introduced reliabilism as an account of epistemic justification he raised the question of what we should take as the processes whose reliability concerns us, and limited this notion to what occurs *in the organism*; whatever comes from outside he counts as inputs to these processes (1992: 116). In perception, for example, the photons, pressure waves, and such that impinge on our senses are considered inputs, and epistemological analysis concerns what occurs after these inputs act on a sensory system. Discussions of the theory-dependence of observation have generally adopted this internal perspective, with concern focusing on whether our beliefs affect the content of our percepts—and if so, how and to what degree.² Yet, in a review of the evidence from cognitive science, Brewer and Lambert (1993, 2001) concluded that prior beliefs have a strong influence on what is perceived only when the input is weak or ambiguous. In this section I will focus on the *inputs* provided by scientific data collection, with particular emphasis on the variety and precision of these inputs. I am not denying the importance of studies of internal cognitive processing, I am just going to focus on an aspect of theory evaluation that is, in my view, too often neglected in epistemological discussions.

²From a large literature, I note the debate between Churchland (1988) and Fodor (1984, 1988), and Raftopoulos (2001).

Consider first some aspects of everyday perception that are familiar, but often ignored in the epistemological literature. Much of this literature (e.g., the extensive literature on ideas and sense-data) focuses on momentary perceptual episodes, but when I see an ordinary middle-size object I get to examine it over *time*. If I have doubts about what I am seeing—perhaps because the object contradicts expectations, or is just unfamiliar—I can *actively* pursue my exploration of it in several ways: I can look at it from different directions, touch it, taste it, perhaps pick it up, and also ask other people for their impressions. Except in cases that require a rapid response, no special significance is given to what I notice in a brief perceptual episode.

Commonplace features of ordinary perception also include an overwhelming tendency to find agreement among my senses and among various individuals. Such agreement is so common that it is taken to be the normal situation, and failures of agreement elicit epistemological puzzles. Thus conflicts between the senses provide one familiar class of illusions. A classic example is a partially submerged stick that looks bent. Without invoking general knowledge of the behavior of physical objects, we can generate a conflict just by comparing the visual appearance of the stick with its tactile appearance. We take this to be a problem—which is solved when we invoke refraction to explain why the straight stick looks bent. At this point the conflict is eliminated because the two sensory appearances are no longer considered epistemic equals. Variations on this example occur when we introduce time and motion. Consider two standard examples: A colored patch that looks uniformly pink from a distance, but appears to be composed of red and white dots on closer inspection; and Descartes' tower that looks round from a large distance, but octagonal from nearby. Again we have cases that generate a problem because they yield conflicts where we expect agreement even though the conflicts arise when we consider different views that occur from different positions at different times. The competing percepts are not directly compared, and our general belief that the objects being examined do not change as our position changes is required to generate a conflict. Here too the conflict is resolved when we understand why our sensory

system is unreliable in one of the competing cases. The different apparent sizes of the moon at the horizon and the zenith provides a variation on this case, since now the perceived object moves while the observer remains stationary. The Müller-Lyer illusion will expand our scope a bit. Here there is a conflict between the apparent lengths of the lines on visual examination and the results of measurement. The introduction of measurement takes us beyond just comparing sensory appearances, and our reasons for believing that the visual appearance is illusory are just our reasons for believing that systematic measurement is more reliable in this situation. We can take one more step along this path by including Descartes' point that the perceived sizes of the moon and other celestial objects conflict with the actual size discovered by astronomy. These last cases are of particular interest because, on the customary view of the relation between theory and the evidence of our senses, the latter is the touchstone by which we assess the former. This relation is reversed in these examples, which illustrate the point that there are situations in which our theoretical beliefs are sufficiently well established so that conflicts lead us to conclude that the senses are misleading.

Intersubjective comparisons are more complex and I will comment on them only briefly. In extreme cases, such as when someone hears voices that other people do not hear, we suspect that an hallucination is occurring unless we find an account for why this person has the ability to detect sounds that others cannot detect. Such cases may be complex because there are differences in individual perceptual abilities. For example, on average women can hear higher pitches than men, and some women hear higher pitches than others. In some cases this leads to an inability to work with CRT terminals which emit an unpleasant whine that most of us cannot hear. There are also genuine questions about cross-cultural differences in perception; the questions of whether they occur, and if so, why, are areas of active research. Still, there is wide cross-cultural agreement at the level of behavior. Whatever other variations exist, people generally do not walk into trees, or off cliffs, or on water. These cases raise important and difficult issues that I cannot pursue here. For the moment I want to emphasize two points

about our everyday experience: *First*, when we are dealing with objects in the world, we expect agreement among our various epistemic sources. Failures of agreement lead to further research with several possible outcomes—such as that different epistemic sources are actually responding to different items, or that there is a failure of reliability in one or more of these sources. *Second*, in recognizing and resolving such conflicts, we typically have multiple resources at our disposal. When surprises occur, the appropriate conclusion is often that we need more information.

In science we also seek agreement among our various epistemic sources, and explanations for disagreements that remain. But there is a special feature of science that we do not find in everyday experience (I am not claiming that this is the *only* difference): in science we continually seek out new sources of information, and these may lead to conflicts with established views. We seek new information by several means, which include expanding our observational range, doing systematic experiments, increasing the precision of measuring techniques, and traveling to various parts of the world for such purposes as observing an eclipse or studying a culture. I want to explore the epistemological significance of this search for new information, but it will be useful to begin by establishing some terminology.

Adapting a suggestion from Basu (2003), I will distinguish between *data* and *evidence*. *Evidence* is used to evaluate a theory and, as noted above, evidence must be characterized using the language of the theory in question. *Data*, as I will use the term, is a *description* of the result of some interaction with nature that does not use the language of any theory *under evaluation*. In cases of theory comparison, proponents of competing theories will typically be able to agree on the data, although they may disagree on its role as evidence for or against a particular theory. Preferably, data will consist of descriptions of public items, such as photographs or cultural artefacts, that can be picked out, measured, and otherwise studied in ways that involve no contribution from any of the theories in question. I will refer to such physical objects as *data sources*. Archeologists and historians have long used urns, buildings, and texts as data sources, and biologists regularly use specimens as such sources. Photographs of astronomical

phenomena have provided sources of data for some time now, and more recently this has been extended to photographs of momentary events in particle detectors. Other examples will be introduced as we proceed. Any physical object can become a data source when some researcher recognizes that the object—or some feature of the object—can provide scientifically relevant information. As a subject develops, items that were not previously recognized as relevant to a particular subject can become relevant—as can features of older items that had not previously caught the interest of researchers. For example, when radio telescopes led to the discovery of quasars, some astronomers reexamined old photographs looking for visual images of these items. This last example underlines an important point: In any observational procedure, the items being studied provide the ultimate data source, but in practice access to these items is often limited. In such cases it is records of specific interactions with these items that serve as the data sources (an important qualification is introduced below, see note 4).

I will be working, then, with a three-layer structure: data sources, data, and evidence. Whether an item fits into one of these categories—and which one—can change with the development of a science. The relation between data and evidence is most important for present purposes. Scientists may agree that a particular spot is on a photograph, on its location relative to other items on the photograph, and on other features that can be established by examining and measuring features of the photograph, but may disagree on whether the spot is an image of a quasar. We make the transition from data to evidence when we shift from describing an item as “a spot on a photograph” to describing it as “the image of a quasar”, or from describing a streak on a photograph with a specific length and curvature, to describing “the path of a positron”. In general, the line between data and evidence depends on the particular situation. With regard to a specific theory, we are dealing with data as long as our descriptions do not use the language of that theory. Consider another example. In the original solar neutrino experiment (cf. Bahcall 1989 Ch. 10, Franklin 2001 Ch. 8), neutrinos were detected as a result of interactions in which chlorine atoms in a large tank of fluid were transformed into argon. The resulting isotope of argon is radioactive, and

the quantity produced in a given time was determined by flushing the argon from the tank and recording its decays using a proportional counter. It is the output from this counter that the experimenters examined. The number of counts provides the data; describing these counts as resulting from decay of argon atoms brings theory into the description in a way that begins connecting the data to the subject of the experiment. Further theory-loading gets us from the argon-decay count to the number of solar neutrinos emitted by the sun in a specific period of time.³ When scientists became aware of a major gap between the predicted number of neutrinos and the number detected, one line of research consisted of seeking natural deposits of materials that could have been serving as neutrino detectors for millennia (Bahcall 1989: 363-372). Such deposits then became data sources for research on solar neutrinos.

My use of 'data source' parallels the use of 'data' by Bogen and Woodward (1988). They distinguish *data* from *phenomena*. In their terminology, data consist of specific items that "for the most part can be straightforwardly observed" (305), such as the reading on a particular thermometer, or a bubble chamber photograph. Data provide evidence for the existence of phenomena, such as the melting point of lead or weak neutral currents, which "are not observable in any interesting sense of that term" (306). My use of 'data' differs from theirs in that I use the term for *descriptions* of items that are straightforwardly observed. Bogen and Woodward go on to argue that the use of data as evidence for phenomena often requires more-or-less sophisticated statistical analysis. For example, given the range of systematic and other errors that can occur in temperature measurements, we do not rely on a single measurement to determine the melting point of lead, but on a set of measurements that will be averaged to find the actual melting point. Each measurement contributes data to the determination of this melting point. Bogen and Woodward's main thesis is that "typically" "systematic" scientific explanations explain phenomena, not data (314). There is no doubt that a great deal of scientific explanation is aimed

³The full process includes displaying each decay on an oscilloscope and photographing the display in order to allow for detailed study of the decay characteristics.

at phenomena in this sense, and Bogen and Woodward acknowledge that singular causal explanations also occur in science (e.g., 322, n. 17). Indeed, it would be pointless to deny that scientists attempt to explain the occurrence of a specific earthquake or a particular death. In any case, I am not concerned here with identifying the proper objects of scientific explanation, so I will not pursue this question any further, and I will not make use of their distinction between data and phenomena. I will, however, distinguish between data sources, data, and evidence as indicated above, and I will use ‘phenomena’ as it is commonly used, without any special technical meaning.

Subjective percepts can also serve as sources of data, and we can apply the distinction between evidence and data to them, although the situation becomes difficult because percepts are short-lived and not available for public examination. In this case, asking whether percepts can provide data amounts to asking whether percepts that yield evidence relevant to the evaluation of a theory can be described in ways that are neutral with respect to that theory. The answer may depend on whether those percepts are infected by the theories under evaluation, but I am going to leave this question aside in the present paper and focus instead on data provided by public items. When evaluating a theory, data are gathered by attempts to interact with the items that form the subject of that theory, and that are presumed to exist *independently of the theory*. This last remark addresses a common theme in the literature where we often read that in science we study mind-independent reality. This will not do since psychology is within the realm of scientific study, and much psychological research studies items that are not independent of minds. The important point is that the subject matter of a scientific theory is presumed to exist independently of that theory; we test the theory by deriving consequences about that subject matter, and gathering evidence by attempting to interact with it. Note especially that such tests involve *attempts* to interact with the items of interest; sometimes these attempts do not succeed, and may even lead to the

conclusion that an hypothesized item does not exist.⁴ Gathering appropriate evidence for evaluating a theory requires gathering data that can be characterized without using the language of that theory. Such data will be theory-dependent in the sense that the theory in question played a role in determining what kind of data we collect—but, again, this kind of theory-dependence does not pre-determine what we will find, or how we must describe it.

One of my main theses in this paper is that *the development of science exhibits a continuing drive to increase the scope and precision of the data that provide the touchstone for evaluating theories.*

Historically, the key step in the improvement of data collection has been the introduction of instrumentation that take us beyond the limits of our senses. I want to consider three classes of instruments that accomplish this goal in progressively more dramatic ways.

First, there are instruments such as a ruler or astrolabe which provide measurements that are more precise than we can achieve with our unaided senses. The characteristic feature of these instruments is that they do not intervene in the causal process between the item being studied and our senses. Instead, they improve the quality of the data by adding an appropriate item to the set of items we can sense. Extra items that improve the quality of the data we acquire with our senses are not always themselves measuring instruments. One place where the issue arises is in reading the output of various instruments. For the moment I want to ignore what a particular instrument measures, and just consider what occurs when we read its output. A common arrangement involves a pointer that moves over a scale. Reading the instrument requires noting the mark on the scale that is below the pointer, but there is room for ambiguity in this case because exactly what is read depends on the position of the reader's head. A standard way of reducing this ambiguity is to place a mirror parallel to the scale and below the pointer. In reading the dial we line up the pointer with its mirror image. This aim of reducing errors generated by

⁴This point about attempted interactions is the qualification that I mentioned three paragraphs earlier. In such cases, the public record of these attempts provides the data source.

our perceptual systems has taken a different path in more recent instrumentation with the introduction of digital readouts, which further limit the chances for misreading. Another example will underline this persistent drive to eliminate errors caused by our perceptual systems. In the early eighteenth century Bradley introduced a method of timing celestial events that depends on looking through a telescope while listening to a clock tick seconds. This was considered highly reliable until Bessel discovered individual variations among astronomers using this method. Bessel attempted to deal with the problem by calibrating astronomers (as in $\text{Smith} = \text{Jones} + .2 \text{ seconds}$), but people are not sufficiently consistent for this to work. (The classic discussion is Boring 1950 Ch. 8; see Brown 1987 Sec. 6.5 for further discussion.) In contemporary astronomy the problem is solved when people are replaced by electronic equipment and our senses enter into the procedure in ways that are less likely to introduce errors. All of these techniques, and more, reduce uncertainties and ambiguities that would be introduced into the data if we relied only on unaided sense perception.

The *second* class of instruments I want to consider includes telescopes and (perhaps) thermometers. It consists of instruments that intervene in the process between the observer and the item being observed, but are limited to improving our information about items we can detect with our unaided senses. These instruments improve our ability to study those items that van Fraassen includes among the observables (1980, 1985). On van Fraassen's account, great distance from an item does not make it *unobservable* as long as our senses would respond to it if we moved sufficiently close; small size does make an item unobservable if it is so small as to make it undetectable by our senses alone. Thus electron microscopes are not instruments in this class, nor are light microscopes in many of their typical uses; optical telescopes and ordinary eyeglasses are instruments in this class. Thermometers provide an interesting case. We can detect heat with our unaided senses, although these senses do not provide accurate quantitative measurements of temperatures. Simple thermometers provide quantitative information by taking heat as their causal input and generating a quantitative output that we detect with a

different sense. Historically the output is visual, but with modern equipment it could be auditory or even Braille. It is not clear to me whether van Fraassen would consider the property made available by this quantitative output to be observable, but the example will serve to introduce the *third* class of instruments that concerns me: those that respond causally to items we cannot detect with our senses and provide outputs we can detect.

Instruments of this third type are especially important because, we have learned, the world is full of items to which our senses do not respond. The path towards this discovery began with doubts about the adequacy of our senses as windows on the world that led first to the distinction between primary and secondary qualities, and the view that there is *less* in the world than appears in our senses. Recognition that there is *more* in the world than in our senses took longer, but once this point was established, instruments that detect such items took on an overwhelming importance. An early example of the kind of discovery that concerns me, perhaps the first, is Herschel's discovery of infrared radiation in sunlight (1800; cf. Hacking 1983: 176-8); it was rapidly followed by Ritter's discovery of ultraviolet radiation (Guiot 1985, Wetzels 1990). Consider how these discoveries occurred. Herschel was studying the spectrum of light from the sun using different colored filters, and he noticed that heat and light sometimes occur together, but that he sometimes felt a sensation of heat with little light, and sometimes light with little heat. This led him to explore the association of heat with light in some detail. In one set of experiments he used a prism to break sunlight into its spectral colors, and thermometers to measure the temperature in different colors, and he extended the series of thermometers at measured distances beyond the edges of the visible spectrum. Herschel found that the temperature was greater towards the red end of the spectrum, continued to rise for a distance beyond the red end, reached a peak, and dropped off. Measurements at the violet end of the spectrum showed that "the power of heating is extended to the utmost limits of the visible violet rays, but not beyond them; and that it is gradually impaired, as the rays grow more refrangible" (Herschel 1800: 291). He initially concluded that light and radiant heat are the

same—that what we call ‘light’ is just that part of spectrum that our eyes detect—and that “the invisible rays of the sun probably far exceed the visible ones in number” (291-2), although later studies led him to doubt this conclusion. The existence of possible rays beyond the violet edge of the spectrum was taken up by Ritter after reading Herschel’s paper. Ritter knew that hornsilver (silver chloride) darkened in the presence of light, and darkened more intensively in light towards the violet end of the spectrum. So he dampened a strip of paper with hornsilver and placed the strip in the spectrum from sunlight projected in an otherwise darkened room. The strip quickly darkened, especially in the violet and beyond, allowing Ritter to conclude that the radiation continues in this direction too. Scientific data collection was launched in a new direction, one with many ramifications for our understanding of the universe. *One of these ramifications is a major increase in the empirical constraints on scientific theories which must eventually account for all of the data that we collect.*

Consider a more recent example. Stars emit energy throughout the electromagnetic spectrum, and theories of stellar energy production and stellar evolution must account for all of this output. Radiation in the narrow range that is visible to us is part of this output, but it does not play a special role in stellar behavior. The study of solar neutrinos that began in the 1960s extended the range of data beyond electromagnetic radiation, and has produced a major confirmation of current theories of stellar energy production, along with a challenge to the standard model of the fundamental particles and interactions. The challenge arises because these studies (along with studies of atmospheric neutrinos) generated anomalies that have been resolved by attributing mass to neutrinos. The standard model distinguishes “left-handed” and “right-handed” particles, where these notions are defined in terms of directions of spin and velocity. While all other fundamental particles occur in both varieties, only left-handed neutrinos are included in the standard model because of evidence that these are the only kind that exist (cf. Rolnick 1994: 173, Table 10.1). But, because of the way handedness is defined, the limitation to left-handed neutrinos requires that neutrinos move at the speed of light, which in turn requires that

they have no mass; if neutrinos have mass, there must be right-handed neutrinos.⁵ Radioactivity provides another indicator of the contributions from studies of items we cannot sense: We have learned vastly more about the nature of matter since the discovery of radioactivity in 1896 than we did in the previous millennia when we were limited to studying properties of matter that are available to our senses; results of this more recent research have often been startling.

Before developing these ideas and their consequences further, I want to pin down some key points. The development of modern instrumentation has allowed us to interact with, and thus gather data from, a much larger portion of the world than we could probe at earlier stages in the history of science. This point holds throughout the sciences—for example, in biology, from the study of bacteria through the study of DNA and human brains and nervous systems. In addition, while expanding the sources of data, scientists have also increased the precision of this data. As a result, the empirical constraints on theories have been increasing. Moreover, we now have greatly increased computing power which makes at least two contributions: it permits more sophisticated analysis of data than was previously possible, and, at least in physics, it permits the derivation of more precise consequences from theories. Without these more accurate consequences, the increased precision of the data would have little bearing on the evaluation of theories. When scientists encounter anomalies the usual response is to examine the

⁵There are further complications. The standard model includes three neutrinos, and this is of considerable importance since the model also requires that each neutrino be matched to another lepton (electron, muon, or tau), and that the total number of leptons equal the total number of quarks. There are also well-established empirical results that support the currently accepted number of fundamental particles (Rolnick 1994 Ch. 15). Yet the recognition that neutrinos have mass has led to measurements of mass-differences between neutrinos which suggest the possibility of a fourth neutrino (and maybe more). Two experiments—MiniBooNE, which is currently taking data, and MINOS, which is still under construction, will yield important new tests of the number of neutrino types. There are web pages describing each of these experiments.

sources of the data more carefully, increase the variety of means of probing those sources, and improve the techniques of data analysis. This data must pass through our senses for us to become aware of it, but it can still be described and used in ways that do not allow for the kind of molding by prior beliefs that has been a major focus of discussions of theory-dependence. Indeed, we have a substantial and continuing history of such data challenging our most deeply held beliefs. An adequate epistemological theory must explain how such challenges arise. It should be clear that such an account will not depend on the purity of subjective percepts, but on the variety and precision of the data we collect. I now want to examine some additional examples in order to bring out three features of scientific data collection.

First, I want to emphasize the importance of *long-term collaborative research*. Bogen and Woodward pointed out that statistical analysis plays a central role in much contemporary empirical research. Consider how this works in a fairly extreme case: the confirmation that top quarks exist (Abachi, *et al.* 1995). As is common in high-energy physics, the experimental apparatus produces outputs that can be studied for the characteristic signature of interactions that involve a top quark, but there are also interactions that produce the same signature without the presence of this particle.⁶ These outputs require further study to determine whether top quarks occurred. In some experiments in high-energy physics, further analysis leads to the identification of specific instances in which an item of interest occurred, but in other cases no such instance is ever identified. The top quark is of the latter sort. The conclusion that top quarks exist derives from a statistical analysis which shows that there is an extremely high probability that some of these cases involved top quarks. Collection and analysis of sufficient data to justify this conclusion at the probability level that physicists consider reliable (five standard deviations, that is, a probability of 99.9994% that there are top quarks among the relevant events) required several years and the combined work of hundreds of people. This deployment of many people over considerable periods of time is not particularly unusual. At an earlier stage in the

⁶This is one aspect of the problem of *backgrounds*, see Galison 1987 for an important discussion.

development of high-energy physics, when bubble chambers were still in operation, experiments typically produced hundreds of thousands of photographs that had to be painstakingly examined to look for events of interest. The examination was often done by technicians using special equipment who, when they found a candidate, called in the physicists for the final assessment (see Galison 1997 Ch. 5 for a detailed account).

Second, I want to consider *unexpected phenomena that do not fit into current theories* in a bit more detail. For example, the discovery of sperm in seminal fluid (by van Leeuwenhoek and Hartsoeker in the 1670s) was wholly unexpected and it was unclear what role, if any, sperm play in reproduction. Sperm did not fit into the prevailing view which held that pregnancy results from the mixture of male and female fluids (Farley 1981, Gasking 1967: 54).⁷ For a substantial period after their discovery, many naturalists believed that sperm are parasites of the testes playing no role in reproduction—a view that survived well into the nineteenth century (Farley 1982: 43-7). I will not explore the many developments on the route to our present understanding of reproduction, but I will note that in the mid-eighteenth century, when the motility of sperm had been established, Maupertuis held that the female contribution to reproduction is a fluid formed in the uterus, and that semen carries the male contribution; but he believed that all of the semen is involved and the essential element consists of solid particles in the semen. Sperm, he thought, are “motile particles whose function was to agitate the commingled mass of the two semina, and thus facilitate the mixture of essential parts” (Gasking 1967: 83).

Other examples of unexpected discoveries that did not fit existing theories include X rays and radioactivity. A less well-known case has been discussed in some detail by Franklin (1986 Ch. 2). Some

⁷Harvey challenged this view in 1651, but sperm did not fit into his account either. Although he held that animals develop from eggs produced by the female (mammalian eggs would not be observed until 1828), and that begin to develop as a result of the influence of male semen, he denied that the semen makes physical contact with the egg (Farley 1982: 17; Gasking 1967: 25-28).

experiments on electron scattering done from 1928-1930 violated existing expectations and led to a body of further experimental and theoretical work that continued for more than a decade. In the course of this research, which involved several experimenters and theoreticians, the initial experiments were largely forgotten as the focus of the research shifted to a particular discrepancy between experiment and theory. Franklin argues that part of the reason for the loss of interest in the original experiments was a failure to reproduce their results. From a later perspective these results can be interpreted as an early detection of the non-conservation of parity in weak interactions. We can also see that the failure to reproduce the original discrepancy resulted from a change in the source of electrons. In the original experiments the electrons were provided by β -decay—a weak interaction. But the relevance of this difference was not clear until later theoretical developments had occurred. Franklin describes a period in which it was well understood that the overall experimental-cum-theoretical situation was in a confusing state. Scientists recognized an anomaly, and it took a substantial period of time—including new theoretical developments—before it was resolved. Moreover, the early data remained, and can now be interpreted as evidence for a phenomenon that came to be understood only later. I submit that we will not get any deeper insight into such cases by focusing our attention on the role of subjective percepts in this research.

Third: as the last example suggests, *data persist even through radical theory changes that lead to its drastic reinterpretation.* I want to add some familiar examples to those already mentioned. One case is provided by falling bodies and projectiles. The Aristotelian, Newtonian, and relativistic accounts of these familiar phenomena are quite different, but all these theories offer explanations of these phenomena, which can be described in language that is independent of all of these theories.

Consider a more recent example in which the persistent data are not identified quite so easily. A major problem in nineteenth-century chemistry was to isolate pure samples of the chemical elements and determine their atomic weights. The work proceeded under the assumption that each element is characterized by a unique weight, and varying weights were interpreted as evidence that the sample was

not pure. But the discovery of radioactivity undermined this guiding assumption when it became clear that transformations occur in which an element emits an alpha particle and two beta particles (in any order). Beta decay was treated as raising the atomic number without affecting the weight (since electrons make no significant contribution to an element's weight). Alpha decay drops the atomic number by two units and the weight by four units. Thus the sequence of emissions just mentioned leaves the element's slot in the periodic table unchanged while its weight is reduced. These results led chemists and physicists to rethink the significance of atomic weight, and to Soddy's introduction of the new concept of an isotope. Soddy, along with his immediate predecessors, believed that the nucleus is composed of protons and electrons. The number of nuclear protons was viewed as determining the atomic weight, and enough electrons were included to neutralize some of the proton charges and get the correct atomic number (Bruzzaniti and Robotti 1989, Soddy 1913, 1932). But this view generated anomalies as new data, concepts, and principles were introduced. Key steps included the concept of spin—which was introduced to explain features of spectral lines; the distinction between Fermi-Dirac and Bose-Einstein statistics; and Pauli's spin-statistics theorem—which requires that particles with integral spin accord with Bose-Einstein statistics while particles with half-integral spin conform to Pauli-Dirac statistics. The most common isotope of nitrogen now generated a problem: its atomic number is seven, its weight is fourteen; thus its nucleus was believed to be made up of fourteen protons and seven electrons. Electrons and protons each have a spin of $1/2$, and the spins of protons and electrons can be parallel or anti-parallel; if two spins are parallel, they add to one; if they are anti-parallel they cancel. Experiment showed that the nitrogen nucleus behaves as particle with one unit of spin, but there is no way to get this spin out of parallel and anti-parallel combinations of 21 particles each with spin $1/2$. The problem was resolved with the discovery of the neutron in 1932, and the redescription of the nitrogen nucleus as consisting of seven protons and seven neutrons (cf. Pais 1986: 299-303). The entire story involves a typically complex interaction between data collection and theoretical development, but the point I want to highlight is the

way in which old data (the varying weights of samples of the elements) remained relevant while the overall body of data grew and was reinterpreted. As the theoretical context changed, the old data became evidence for quite different claims. To be sure, sometimes data that were considered relevant to a particular topic are later reevaluated as irrelevant, but this requires specific reasons for dropping consideration of that data.⁸

Reflection on all the above examples suggests that, while empiricists are correct in holding that data gathered through interactions with nature provides the ultimate touchstone for evaluating scientific theories, *traditional empiricism went astray because it treated data as subjective percepts, rather than as a body publicly available constraints provided by nature*. I want to develop this idea further, and consider the relation between data and evidence in more detail.

In actual encounters between advocates of fundamentally different views, it is typically possible to pick out a body of data that is described in language acceptable to all parties of the dispute. Indeed, major innovators—such as Copernicus, Galileo, Descartes, Newton, Darwin, and Einstein—were all masters of the older view that they sought to supersede. This put them in the position of being able to identify situations in which their view yields results that differ from those of the prevailing view, and to describe these situations in terms that would be acceptable to their peers. For example, Galileo understood why Aristotelian physics predicts that a rock dropped from the top of the mast of a moving ship will land at the back of the ship; his physics predicts a different result. The outcome of such an experiment would be observable by all, and could be described in common language that does not

⁸For example, Laudan (1977: 29) points out that Cartesian planetary theory explains why all the planets move in the same direction, while Newtonian theory does not explain this phenomenon. But the phenomenon is not just ignored in Newtonian theory. Rather, it is a feature of Newtonian theory that the direction of motion is determined by initial conditions, not by fundamental theory. Of course, cases also occur in which previously accepted data are rejected because they are found to result from a mistake.

presuppose either theory—as I just did. If an outcome supports a new theory, this result might provide a reason for putting in the effort required to learn that theory. I submit that it will always be possible to find a mutually agreeable description of an outcome, even when that data receives different further interpretations from the perspectives of different theories. Two examples will help pin down the idea.

Consider *first* an example on which Kuhn places considerable weight:

Since remote antiquity most people have seen one or another heavy body swinging back and forth on a string or chain until it finally comes to rest. To the Aristotelians, who believed that a heavy body is moved by its own nature from a higher position to a state of natural rest at a lower one, the swinging body was simply falling with difficulty.

Constrained by the chain, it could achieve rest at its low point only after a torturous motion and a considerable time. Galileo, on the other hand, looking at the swinging body, saw a pendulum, a body that almost succeeded in repeating the same motion over and over again *ad infinitum*. (1962: 118-19)

Did Galileo and an Aristotelian have different visual experiences when looking at the swinging body? I submit that for purposes of understanding the role of evidence in the evaluation of scientific theories, we need not answer this question. It is sufficient that we can describe what is occurring in a way that is neutral between the two theoretical accounts. Kuhn provides just such a description in the initial sentence of the quoted passage. In this case the datum does not provide a basis for choosing between the two theoretical accounts, and the next task is to find some data on which the two theories differ. Often this process begins with the identification of a data source. Hanson emphasized the need for a common data source when he initiated the discussion we are pursuing. Having asked, “Do Kepler and Tycho see the same thing in the east at dawn?” (1958: 5), he tells us that:

Unless both are visually aware of the same object there can be nothing of philosophical interest in the question whether or not they see the same thing. Unless they both see the

sun in this prior sense our question cannot even strike a spark. (7)

Now imagine a little dialogue that is more informative than the one Hanson presents (6):

Tycho: See that bright object near the eastern horizon?

Kepler: Yes.

Tycho: It is moving westward around the earth.

Kepler: Not so, it looks like it is moving because the earth is turning from west to east.

Tycho: Hmmm, what tests would decide between these two views?

In the case of the earth's annual motion, stellar parallax provides an example whose import was understood since the ancient world. Its absence in naked-eye (and early telescopic) astronomy was generally recognized, although whether this counts as evidence against an earth that moves around the sun depends on additional considerations.

Second, consider an example from more recent science: the use of astronomical red-shift data to determine recession velocities. The example is particularly interesting because the move from an agreed value of the red-shift to a recession velocity depends on the Döppler effect, but classical mechanics and special relativity give different formulas for converting a red-shift to a velocity. If S is the red shift and β the recession velocity expressed as a fraction of the velocity of light, then the classical formula is $\beta = S$ while the relativistic formula is $\beta = (S^2 - 1)/(S^2 + 1)$. Since neither theory limits the size of S , the classical formula does not include any limit on the resulting velocity; a limitation to values less than the speed of light is a necessary consequence of the relativistic formula. So specific red-shift data can become evidence for two quite different velocities, while the data themselves are not in dispute.⁹

Still, the use of instrumentation in contemporary data collection leads to further epistemological worries since a description of the output of an instrument will depend on the accepted theory of that

⁹There are also some astronomers who deny that cosmological red shifts are to be interpreted as indicating recession velocities, while still accepting the measured values of these red shifts.

instrument. There may be little chance of disagreement in recognizing that the numeral on a digital display is 5, but the move to 5 miles per hour, or 5 degrees Celsius, or 5 neutrinos in the last month, requires a theoretical understanding of the instrument. In many cases this worry can be put aside because the design and operation of the instruments used (including the computers) are not in dispute.

Greenwood, who calls theories of the instrument “exploratory theories”, notes:

The same exploratory theories about telescopes and photography were employed in the evaluation of Newton’s and Einstein’s theories; the same techniques of chemical separation and purification were employed by defenders and critics of Prout’s hypothesis; the same theories of X-ray diffraction were employed in the testing of competing theories of the structure of DNA. (1990: 570)

Instrumental outputs that occur in the form of photographs, digital displays, and computer printouts are easily identifiable public objects. Even if a photograph looks different to people with different theoretical commitments, the photograph is there on the table, available for continuing examination and discussion. At this level we can usually find a basis for agreement among those who disagree on its significance—including those who deny that it has any relevance at all.

There is, however, another concern. Consider a case in which a test of a theory reduces to a prediction of what numeral will appear on a display, and suppose that the prediction is falsified; from a logical point of view, one can respond by challenging the accepted theory of the instrumentation. This is a genuine issue, but its significance is ameliorated because of the special role that data play in science. The empirical side of science includes a commitment to explain these results; since these results are provided by nature, in the long run they cannot simply be ignored. Moreover, in science it is not enough just to note the logical possibility that we may misunderstand how the instrument functions; exactly what is wrong must be determined, and the subject remains in an unsatisfactory state until this is done. I want to examine two examples in which the process of coming to understand an unexpected outcome did lead

scientists to modify their views on the operation of their instruments.

First, consider the research that led to Becquerel's *accidental discovery* of radioactivity. His aim was to study fluorescence: a process in which a material absorbs radiation of some frequency and then radiates at a frequency that is characteristic of the material. Becquerel had identified a crystalline uranium salt that, he believed, absorbed sunlight and radiated X rays. His rather simple instrument for studying this phenomenon consisted of a photographic plate wrapped in black paper to protect it from sunlight, with the uranium salt in a dish on top of the protected plate. Sometimes he placed an object between the salt and the dish so that the object's image would be found when the photograph was developed. On February 26 and 27 of 1896 he had prepared an arrangement of this sort, with a copper cross in the dish, but it was cloudy in Paris so he kept the set-up in a dark cupboard. Then, for reasons that have never been explained, Becquerel developed the plates and found an image of the cross. It was now clear that the instrument had not been functioning as he believed, nor had he been studying what he thought he was studying. Still, it was nature that provided the unexpected result, and he attributed this outcome to a previously unknown phenomenon that became a subject of intense research.¹⁰

Second, the solar neutrino experiments provide an example that involves explicit theory testing. The aim of the first experiment in the series was to test the standard theory of stellar energy production by measuring the neutrino flux from the sun. The standard theory predicts specific numbers of neutrinos of specific energies, and this prediction had never been tested when the experiment was developed in the early 1960s—just a few years after physicists became confident that they could reliably detect neutrinos. The experimental apparatus was capable of detecting only the highest-energy electron neutrinos predicted by the theory, and the prevailing understanding of neutrinos indicated that all of the neutrinos

¹⁰Four of Becquerel's papers on this research are translated in Romer 1964. This volume also includes other key papers from this period; the collection is continued in Romer 1970, which contains a valuable "Historical Essay" covering the period.

produced in the sun and emitted in the appropriate direction reached the detector. But the number of neutrinos detected was significantly lower than predicted. This led to more powerful experiments to detect solar neutrinos and, as noted above, was one source of support for an important revision in neutrino theory. The revised neutrino theory predicts that a significant number of these neutrinos will change type as they pass through the sun, and thus accounts for the low result. In the original design of the experiment neutrinos were used as a probe for testing the theory of stellar energy production, and the prevailing view of neutrinos played a central role in understanding the experimental results. The change that has occurred in our understanding of neutrinos thus amounts to a change in the theory of the instrument. However, it was not just postulated that some electron neutrinos are lost. Rather, the revised theory, which involves massive neutrinos, has other testable consequences.¹¹

Many philosophers will still view this dependence of data (and thus of evidence) on the theory of the instrument as just one instance of the general logical point that there is a great deal of flexibility in deciding which previously accepted claims to modify in the face of recalcitrant data—and thus as a source of scepticism about accepted scientific results. I urge, however, that our examples suggest a rather different perspective. As the body of data that must be accommodated grows and becomes more precise, the constraints they impose on scientific theories become more demanding, and it becomes harder to defend favored hypotheses without generating new empirical anomalies (Brown 2001, Greenwood 1990). It is this large and growing set of constraints, not the purity of our percepts, that provides grounds for thinking that science is teaching us about a world that exists independently of our beliefs.

¹¹Franklin's extended study of the history of weak-interaction theory provides another example: many older experiments had to be reinterpreted once the non-conservation of parity in weak interactions was established. Physicists were well aware that the outcomes of previous experiments remained, and that a new interpretation was *required* (2001: 130-3).

From this perspective, theories have an *enabling* function: they make it possible for us to gather evidence that is not available to our unaided senses.¹² Theory-dependent evidence gathering becomes vital once we recognize that the world contains items we cannot detect with the senses we evolved on the surface of this planet, and that the items we can detect with these senses may not be the most informative sources for understanding the universe.¹³ One can, of course, follow van Fraassen and the logical empiricist tradition (e.g. Hempel 1965) in limiting science to the modest goal of just predicting observables, but this is not mandatory even if we agree that this modest approach minimizes the chances of error. Maximizing the probability that our views are correct is one scientific goal, but as Popper argued in 1934 (cf. Popper 1980), theories of wider scope and depth are also goals that one can pursue through science, even at the price of added epistemic risk (cf. Hooker 1985, 1987).

One more form in which evidence may be theory-dependent remains to be considered: empirical research, guided by theory, leads us to seek evidence in some places but not others; again, this seems to bias the results. By now my reply should be, at least, expected: What we find when we look is not predetermined by our theories. Moreover, we have seen that challenges to theories may come from unexpected places, and that *over time* the variety of places in which scientists seek evidence has been expanding. We have every reason to anticipate that this expansion will continue. To be sure, we may never look in all the places that might be relevant, but this is just another reminder of the fallibility of science.

One last familiar objection remains to be considered: that confirmation based on theory-laden

¹²Shapere 1982 is the *locus classicus* of this view; this paper was well known in the philosophy-of-science community for several years before its publication. See also Brown 1979, 1987, 1995, Greenwood 1990.

¹³While theory construction and evaluation have taken us beyond the limits of our evolved senses, it remains an open question what limits are imposed by the fact that our theory-constructing brains are also products of our local evolution.

evidence is viciously circular. I suggest that this conclusion does not follow quite so easily. In two previous papers (Brown 1993, 1994, see also Franklin *et al.* 1989) I have argued that when we look at the *details* of an evidence-gathering procedure, we find that even when a theory plays a central role in the procedure, it may still be possible that the empirical outcome challenges that theory. As a result, such use of a theory in its own test does not automatically bias the test in favor of that theory, so that any circularity we may identify in the procedure is not epistemically problematic. Assessment of whether a vicious circularity occurs requires a detailed analysis of the procedure, including the exact way in which the procedure interacts with nature, and the exact way in which the theory enters into this procedure. This line of argument has been carried a step further by Shogenji (2000) who has shown that for a wide variety of cases that may seem circular on a superficial reading, there is in fact no circle because the theory being evaluated enters into the procedure only in the way in which such theories enter into hypothetico-deductive tests.

3. Sailing On

The main outcome of this paper is to reaffirm the view that empirical evidence, *properly understood*, provides the ultimate basis for evaluating scientific theories, and for believing that our theories are teaching us about a world that exists apart from our beliefs about it. But a proper account of this evidence is rather more complex than what we find in traditional empiricism. Science studies nature—which includes human perceptual and cognitive abilities—and the evidence relevant to the evaluation of theories is gathered by wide ranging interactions with nature. Sense perception provides our interface with nature, but as science has developed we have learned to interact with larger portions of nature—including the vast portions that are not detectible by our unaided senses. We have also learned to increase the precision of the evidence we gather from those interactions. Through these interactions we test our theories against nature—not against our subjective percepts—even though any information we acquire as a result of such interactions must pass through our senses at some stage. Scientists have

learned to reduce the effects that limitations and quirks of our senses play in the acquisition of evidence by limiting their role to tasks where they are least liable to introduce errors—such as reading digital displays and examining photographs and computer-constructed images and graphs. Items of this sort provide data that are stable and available to intersubjective examination. As a result, contributions that established beliefs make in producing the data, and in the transition from data to evidence, are subject to public evaluation and reconsideration over time.

As scientists extend the range of their interactions with nature they encounter results that contradict theory-generated expectations, as well as results that were utterly unexpected. We have seen that scientists typically have no difficulty in recognizing such problematic cases. We have also seen that while theory testing requires describing evidence in the language of the theory being tested, this does not entail describing that evidence as in conformity with that theory. In reporting evidence, scientists make use of logical concepts—including negation—which are not internal to any specific scientific theory. Moreover, the occurrence of outcomes that cannot be interpreted at all in terms of existing theories typically indicates a failure of theory, and the need to construct new theory.

Extension of the range of evidence to the results of interactions with items we cannot sense depends on our theoretical understand of how our instruments work. This introduces another layer of theory-dependence into evidence gathering procedures, but the significance of this theory-dependence is ameliorated by nature's permanent possibility of generating surprises which must eventually be taken into account. As our data-base expands, the constraints that nature places on our theories also expand.

Theory change may lead to major reinterpretations of old data, but there is a growing public body of data that remains intact through these reinterpretations, and that must eventually be accommodated by our theories. The requirement that we deal with the data is a central characteristic of the scientific study of nature, and a feature that distinguishes science from many other human pursuits. It is also a feature that distinguishes scientific fallibilism from relativism. There is no guarantee that this process will

eventually end in a correct account of every aspect of nature—although the situation may be different in different domains. Empirical evidence does not directly reveal how nature is; we seem able to find fundamental features of nature only through theory-construction. Still, empirical evidence provides the final touchstone for accepting theories; in this respect the empiricist tradition is correct. But in order to understand how this touchstone works, we must get past the focus on subjective percepts that has pervaded the history of empiricism. In doing so, we come to recognize that scientific research is more difficult and less certain than many earlier historians, philosophers, and scientists believed. But this is not the same as saying that scientific results are just reflections of the prior beliefs of scientists and their culture. Whatever role these factors may play in the pursuit of science, scientific theory evaluation is constructed so as to give nature a central role.

References

- Abachi, S. *et al.* (1995), "Observation of the Top Quark," *Physical Review Letters* 74: 2632-2637.
- Bahcall, J. (1989), *Neutrino Astrophysics*, Cambridge: Cambridge University Press.
- Basu, P. (2003), "Theory-ladenness of Evidence: A Case Study from the History of Chemistry," *Studies in History and Philosophy of Science* 34A: 351-368.
- Bogen, J. and Woodward, J. (1988), "Saving the Phenomena," *Philosophical Review* 97: 303-352.
- Boring, E. (1950), *A History of Experimental Psychology* 2nd edition, New York: Appleton-Century-Crofts.
- Brewer, W. and Lambert, B. (1993), "The Theory Ladenness of Observation: Evidence from Cognitive Psychology," in *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 254-9.
- (2001), "The Theory-Ladenness of Observation and the Theory-Ladenness of the Rest of the Scientific Process," in *PSA00 Part I, Supplement to Philosophy of Science* 68. No. 3, J. Barrett and J. Alexander, eds., S176-S186.
- Brown, H. (1975), "Paradigmatic Propositions," *American Philosophical Quarterly* 12: 85-90.
- (1979), "Observation and the Foundations of Objectivity," *The Monist* 62: 470-481.
- (1987), *Observation and Objectivity*, New York: Oxford University Press.
- (1993) "A Theory-Laden Observation Can Test the Theory," *British Journal for the Philosophy of Science* 44: 555-559.
- (1994), "Circular Justifications," in *PSA 1994* vol. 1, D. Hull, M. Forbes, and R. Burian, eds., East Lansing MI: The Philosophy of Science Association, 406-414.
- (1995), "Empirical Testing," *Inquiry* 38: 353-399.
- (2001), "Incommensurability and Reality," in P. Hoyningen-Huene, and H. Sankey, eds., *Incommensurability and Related Matters*, Dordrecht: Kluwer Academic Publishers, 123-142.
- Bruner, J. and Postman, L. (1949), "On the Perception of Incongruity: A Paradigm," *Journal of Personality* 18: 206-223.
- Bruzzaniti, G. and Robotti, N. (1989), "The Affirmation of The Concept of Isotopy and the Birth of Mass Spectrography," *Archives Internationales D'Histoire des Sciences* 39: 309-334.
- Churchland, P (1988), "Perceptual Plasticity and Theoretical Neutrality," *Philosophy of Science* 55: 167-87.

- Farley, J. (1981), "Sperm," in *Dictionary of the History of Science*, W. Bynum, E. Browne, and R. Porter, eds., Princeton: Princeton University Press, 397-8.
- (1982), *Gametes and Spores: Ideas about Sexual Reproduction, 1750-1914*, Baltimore: Johns Hopkins University Press.
- Fodor, J. (1984), "Observation Reconsidered," *Philosophy of Science* 51: 23-43
- (1988), "A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'" *Philosophy of Science* 55: 188-98.
- Franklin, A. (1986), *The Neglect of Experiment*, Cambridge: Cambridge University Press.
- (2001), *Are There Really Neutrinos? An Evidential History*, Cambridge, MA: Perseus Books.
- Franklin, A., et al., (1989), "Can a Theory-Laden Observation Test the Theory?" *British Journal for the Philosophy of Science* 40: 229-231.
- Galison, P. (1987), *How Experiments End*, Chicago: University of Chicago Press.
- (1997), *Image and Logic*, Chicago: University of Chicago Press.
- Gasking, E. (1967), *Investigations into Generation 1651-1828*, Baltimore: Johns Hopkins Press.
- Goldman, A. (1992), "What is Justified Belief" in *Liasions*, Cambridge, MA: MIT Press, 105-126.
- Greenwood, J. (1990), "Two Dogmas of Neo-Empiricism: The 'Theory-Informity' of Observation and the Quine-Duhem Thesis," *Philosophy of Science* 57: 553-574.
- Guiot, J. (1985), "Zur Entdeckung Der Ultravioletten Strahlen Durch Johann Wilhelm Ritter," *Archives Internationales d'Historie des Sciences* 35: 346-356.
- Hacking, I. (1983), *Representing and Intervening*, Cambridge: Cambridge University Press.
- Hanson, N. (1959), *Patterns of Discovery*, Cambridge: Cambridge University Press.
- Hempel, C. (1965), "The Theoretician's Dilemma," in *Aspects of Scientific Explanation*, New York: Free Press, 173-226.
- Herschel, W. (1800), "Experiments on the Refrangibility of the Invisible Rays of the Sun," *Philosophical Transactions of the Royal Society of London*: 284-292.
- Hooker, C. (1985), "Surface Dazzle, Ghostly Depths," in *Images of Science*, P. Churchland and C. Hooker, eds., Chicago: University of Chicago Press, 153-196.
- (1987), *A Realistic Theory of Science*, Albany, NY: State University of New York Press.

- Hoyningen-Huene, P. (1993), *Reconstructing Scientific Revolutions*, trans. A. Levine, Chicago: University of Chicago Press.
- Kuhn, T. (1962), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.
- (1991), “The Road Since Structure” in *PSA 1990* vol. 2, A. Fine, M. Forbes, and L. Wessels, eds. East Lansing, MI: The Philosophy of Science Association, 3-12.
- Laudan, L. (1977), *Progress and its Problems*, Berkeley: University of California Press.
- Leibniz, G. (1985), *New Essays on Human Understanding*, P. Remnant & J. Bennett, trans. and eds., Cambridge: Cambridge University Press.
- Newton, I. (1952), *Optics* 4 th edition, New York: Dover Books.
- Pais, A. (1986), *Inward Bound*, New York: Oxford University Press.
- Popper, K. (1980), *The Logic of Scientific Discovery*, New York: Routledge.
- Putnam, H. (1962), “The Analytic and the Synthetic,” in *Minnesota Studies in the Philosophy of Science* III, H. Feigl and G. Maxwell, eds., Minneapolis: University of Minnesota Press, 358-397.
- Raftopoulos, A. (2001), “Reentrant Neural Pathways and the Theory-Ladenness of Perception, in *PSA00* Part I, Supplement to *Philosophy of Science* 68. No. 3, J. Barrett and J Alexander, eds., S187-S199.
- Rolnick, W. (1994), *The Fundamental Particles and Their Interactions*, Reading, MA: Addison Wesley Publishing Company.
- Romer, A., ed. (1964), *The Discovery of Radioactivity and Transmutation*, New York: Dover Books.
- , ed. (1970), *Radiochemistry and the Discovery of Isotopes*, New York: Dover Books.
- Sellars, W. (1963), “Is there a Synthetic A Priori?” in *Science, Perception and Reality*, New York: Humanities Press, 298-320.
- Shapere, D. (1982), “The Concept of Observation in Science and Philosophy,” *Philosophy of Science* 49: 485-525.
- Sheehan, W. (1988), *Planets & Perception*, Tucson: University of Arizona Press.
- Shogenji, T. (2000), “Self-Dependent Justification Without Circularity,” *British Journal for the Philosophy of Science* 52: 287-298.
- Soddy, F. (1913), “Intra-atomic Charge,” *Nature* 92: 399-400.
- (1932), *The Interpretation of the Atom*, London: John Murray.
- Toulmin, S. (1961), *Foresight and Understanding*, New York: Harper & Row.

van Fraassen, B. (1980), *The Scientific Image*, Oxford: Clarendon Press.

— (1985), “Empiricism in the Philosophy of Science,” in *Images of Science*, P. Churchland and C. Hooker, eds., Chicago: University of Chicago Press, 245-308.

Wetzels, W. (1990), “Johann Wihlelm Ritter: Romantic Physics in Germany,” in A. Cunningham and N. Jardine, eds. *Romanticism and the Sciences*, Cambridge: Cambridge University Press, 199-212.