

Bogen and Woodward's data-phenomena distinction, forms of theory-ladenness, and the reliability of data

Dr. Samuel Schindler
Department of Philosophy
University of Birmingham
Edgbaston, Birmingham
B15 2TT, United Kingdom
s.schindler @ bham.ac.uk
<http://www.samuelschindler.org>

Abstract

Some twenty years ago, Bogen and Woodward challenged one of the fundamental assumptions of the received view, namely the theory-observation dichotomy and argued for the introduction of the further category of scientific phenomena. The latter, Bogen and Woodward stressed, are usually unobservable and inferred from what is indeed observable, namely scientific data. Crucially, Bogen and Woodward claim that theories, which we seek to test, predict and explain phenomena, not data. But then, of course, the thesis of theory-ladenness cannot apply. The idea that theory-ladenness does not occur in scientific practice is one of the claims I contest in this paper. More importantly, Bogen and Woodward held that the reliability of the data, which constitutes the first step towards an inference from the data to the phenomena, can be secured without the theory one seeks to test. Again, this appears not to be descriptive of actual scientific practice. In order to show this, I re-visit two case studies that have figured heavily in Bogen and Woodward's publications and others: the discovery of weak neutral currents and the discovery of the zebra pattern of magnetic anomalies (Kaiser 1995). I show that, in the latter case, data can be ignored if they appear to be irrelevant from a particular theoretical perspective (TLI) and, in the former case, the tested theory can be critical for the assessment of the reliability of data (TLA). I argue that both TLI and TLA are much stronger senses of theory-ladenness than the classical thesis and that neither TLI nor TLA can be accommodated within Bogen and Woodward's account.

Introduction

In 1988 Bogen and Woodward introduced a ‘third level’ into the classical dichotomy of theory and observations: scientific phenomena. Unlike the notion of phenomena in the empiricist tradition of ‘saving the phenomena’ defended by van Fraassen (1980) in particular, Bogen and Woodward’s phenomena are usually unobservable (pp. 305-6). However, phenomena can be inferred from observable data, after the reliability of the latter has been secured by various experimental procedures involving statistical inference, data reduction, the exclusion of confounding factors, error and noise control and the like. There is another feature that distinguishes data and phenomena. Whereas phenomena are “stable” over different experimental contexts, data typically lack this stability and are often highly idiosyncratic due to the causes peculiar to those contexts (cf. *ibid.*, p. 317, p. 319). Whereas we have good reasons to believe that there is one true melting point of lead, which has a particular value (i.e. 327.43 °C), it may well be that none of the values of a particular set of thermometer readings will coincide with the true value. What is more, we may not even know the precise reason for the variation of those data points. They depend on a multitude of factors, not all of which we will be able to fully control at all times:

The outcome of any given application of a thermometer to a lead sample depends not only on the melting point of lead, but also on its purity, on the workings of the thermometer, on the way in which it was applied and read, on interactions between the initial temperature of the thermometer and that of the sample, and a variety of other background conditions. (Bogen and Woodward 1988, p. 309)

Even though we often may be ignorant about exactly which causes were involved to which extent in the production of the data, we nevertheless may still have good reasons to believe in the existence of the *phenomenon* of the melting point of lead if we can infer it from the observed values¹. It is this inferred value, i.e. the phenomenon, which is the “thing-to-be-explained”—not the data. In more complex situations than this “toy” example, say in the discovery of weak neutral currents (Galison 1983, Pickering 1984), which has been quoted rather extensively by Bogen and Woodward (see also Woodward 1989, 2000) the inference from the data to the phenomena is not as easy and straightforward as in the example of the

¹ Glymour (2000) has pointed out that in the example of the melting point of lead what Bogen and Woodward call phenomena and data seems to be merely giving other names to the well-defined statistical concepts of population and sample. Although I agree with Glymour, I will however ignore this point and follow Bogen and Woodward in calling the melting point of lead a phenomenon for sake of better illustration.

melting point of lead. There, background effects, which mimicked neutral currents, had to be taken care of before neutral currents could be established as a genuine phenomenon (see below and Schindler (under review) for details). And yet, Bogen and Woodward claim that this case is analogous to the simple example of the melting point of lead. In both cases, phenomena and data possess the characteristics mentioned above and in both cases phenomena are *somehow* inferred from the observable data. Apart from the occasional (rather vague) reference to statistical techniques, Bogen and Woodward don't say much about the nature of those inferences². In any case, they stress that one must make sure that the inferential "base" (i.e. the data) is not flawed in order to carry out inferences to the phenomena:

[T]he question of whether data constitute reliable evidence for some phenomena turns (among other things) on such considerations as whether the data are replicable, whether various confounding factors and other sources of possible systematic error have been adequately controlled, on statistical arguments of various kinds, and on one's procedures for the analysis and reduction of data. (ibid., p. 327)

A point which Bogen and Woodward emphasise very strongly throughout their paper is that the reliability of the data can be secured *without* higher order theory explaining or predicting why particular data happened to be produced by one's experimental apparatus. On the contrary,

[...] the details of the operation of these causes will be both unknown and enormously complex, so that typically there will be no possibility of explaining why this or that individual data-point occurs. (ibid., p. 334)

Nevertheless, despite one usually not being able to provide explanations of how the data were produced, Bogen and Woodward obviously believe that one can identify these causes³ and control for them in the appropriate ways by the various experimental piecemeal procedures mentioned in the above quote. If the tested theory *were* involved in these procedures, one of Bogen and Woodward's motives for introducing the data-phenomena distinction in the first

² Bogen and Woodward explicitly make clear (p. 338) that they do not mean the well-known Inference to the Best Explanation (IBE) according to which one infers the truth of a hypothesis from the fact that it explains the data best (when compared to other hypotheses explaining the same data). Explanation Bogen and Woodward generally regard as being extraneous to data to phenomena reasoning.

³ Note that Bogen and Woodward, like Cartwright (1983), seem to presuppose that a line can be drawn between one identifying the causes of certain data and one providing explanations of how those data were caused. Even though they allow for the first conjunct, they claim that the latter more often than not cannot be given (see below).

place would not materialise: rebutting the thesis of theory-ladenness (cf. *ibid.* p. 310, pp. 342-7).

According to the thesis of theory-ladenness, theories interfere on a very low level of cognition: our observations are significantly altered by the theories we seek to test with those same observations. This is usually seen as a challenge to scientific objectivity because if the thesis of theory-ladenness were true, and if we were to hold different theories about the world, we would not be able to agree on what we actually observe. Observations thus could no longer serve as the arbiter between different theories. Bogen and Woodward's scheme rules out the thesis of theory-ladenness because it simply denies that there is any direct epistemological link between theories and observations. It is not observations but rather phenomena that theories predict and explain. Phenomena, in turn, cannot be subject to the thesis of theory-ladenness because they a) are unobservable, and b) inferred from the data in an epistemologically unproblematic way.

Recently, Woodward (this volume) has indicated that he accepts that at least some forms of theory-ladenness (if somewhat weaker than the form discussed above) do occur in scientific practice⁴. According to one of these forms, the design and the conduction of experiments can be motivated by theories (call this the thesis of "theory-drivenness" of scientific practice). Such a claim is epistemologically rather unproblematic. Whether or not the conduct of particular experiments has been motivated by the theory, which these experiments seek to test, is irrelevant to the (logical) question of whether the data obtained in these experiments do or do not support the belief in particular phenomena, and whether the latter in turn, do or do not confirm the theory's predictions. There is another form of theory-ladenness which Woodward (this volume) is happy to accept. A theory may provide the vocabulary for *interpreting* and *describing* the data. Several questions arise here. How can the theory that explains and predicts the phenomena (call this T(p)) but not the data, provide the interpretational framework for the latter? And isn't an interpretation not some sort of explanation, i.e. doesn't the theoretical interpretation of data not clash with their claim that theories do not explain the data? *Why* should T(p) provide such a framework for the data in the first place? After all, T(p) only predicts and explains the phenomena, not the data, so why would an interpretation of the data be required? Lastly, given one interpretational framework

⁴ In his talk at the 'Data-Phenomena-Theories' conference at the University of Heidelberg in Germany (11-13 September 2008), Woodward has tried to weaken this apparent contradiction by mentioning that he and Bogen used the word "typically" too often in their joint paper of 1988 and that that their data-phenomena distinction with all its characteristics was not meant to be a universal account of how science works.

for the phenomena and one for the data, what is their relationship supposed to be? These are questions, I think, Bogen and Woodward need to clarify.

Yet, in this paper, I want to focus on the question of whether Bogen and Woodward's scheme is really as descriptively adequate of scientific practice as they have claimed. In particular, I want to challenge Bogen and Woodward on their claim that the reliability of data can be secured without the theory which predicts and explains the phenomena that are supported by those data. Take for instance "data reduction", one of the procedures, which Bogen and Woodward quote in this context. Here, Bogen and Woodward have referred to the fact that scanning bubble chamber photographs for significant events is often carried out by theoretically naïve personal or even computers:

The extent to which such methods of data reduction are independent of any concern with explanation is illustrated by the fact that the person or machine performing these tasks can carry them out without understanding either the theory which explains the interactions for which the photographs are evidence, or the physical principles by which the equipment works. (p. 333)

I don't think *routines* of data reduction, whose carrying out does not require much theoretical understanding, supports Bogen and Woodward's claim that the reliability of data is secured without the theory whose predictions are tested against the data (Cf. Duhem (1974 [1906], p. 145) and Hacking (1983, p. 179))⁵. Even though someone might have the relevant skills to follow a routine that has been given to her, this routine has to be given to her in the first place. It is hard to imagine that the working out a data reduction routine, which will involve the choice *for* and *against* particular data, could be done without any theoretical understanding of what one *ought to see* in the data (here: certain events on bubble chamber photographs), if the routine in question is supposed to produce any meaningful results. Rather than assuming the independence of experimental from theoretical practice when it comes to the reduction of data (and other forms of establishing the reliability of data, as I shall argue), a thorough interplay between the two appears to be much more plausible.

At this point, one or two comments about the thesis of theory-ladenness are in order. Although philosophers like Thomas S. Kuhn have often been ridiculed as questioning the reliability of our perceptual apparatus and the stability of our perceptions, this is simply not the point about theory-ladenness. Thus, van Fraassen in his *Scientific Image*, asked us not to confuse "*observing* (an entity, such as a thing, event, or process) and *observing that*

⁵ Hacking in fact uses a very similar example of theoretically naïve lab assistants scanning cloud chamber photographs for tracks left by positron events (1983, p. 179).

(something or other is the case)". Van Fraassen illustrates this distinction by calling on a thought experiment in which a member of some isolated tribe is shown a tennis ball. From the behaviour of tribe member, we can infer that she saw the tennis ball (she may throw it, chew it, etc.), but it would be going too far to claim *that* she has seen a tennis ball—because she does not possess the required concepts (van Fraassen 1980, p.15; original emphasis). But even though we and the tribe member may have a common percept of a tennis ball, it is highly unlikely that we see anything other than a tennis ball when we perceive a tennis ball (the tribe member will of course see something else, say an apple with strange properties). This is not a matter of interpretation or any other voluntary or conscious act. *Crucially*, the fact that we and the tribe member will have the same percept is entirely irrelevant to the fact that we and the tribe members will see different things. To quote a more practical example, take the hearing of someone speaking Chinese. Whereas I won't hear much more than strange unfamiliar sounds, someone with the knowledge of Chinese is bound to hear meaningful sentences when hearing the same utterances. Someone with a good knowledge of Chinese will not even have to think about what she hears (in terms of grammatical rules). She will just hear meaningful sentences—despite the fact that the (phonetic) sounds the Chinese speaker utters are the same for her and for me!

Although theory-ladenness leaves intact the stability of our perceptions, it does not merely refer to the fact that theories often motivate certain experiments or that data must be theoretically interpreted. If that were the case, it would be rather strange why we would need a different concept in the first place⁶. But this is not to say that the concept of theory-ladenness is superfluous. Rather, theory-ladenness implies something stronger than the mere interpretation of data or the theory-drivenness of scientific practice but something weaker than a change of actual perception caused by theory. For someone with a distaste for the rather casual examples I quoted above, consider an example due to Duhem, which is worth quoting in full for its vividness:

Go into the laboratory; draw near this table crowded with so much apparatus: an electric battery, copper wire wrapped in silk, vessels filled with mercury, coils, a small iron bar carrying a mirror. An observer plunges the metallic stem of a rod, mounted with rubber, into small holes; the iron oscillates and, by means of the mirror tied to it, sends a beam of light over to a celluloid ruler, and the observer follows the movement of the light beam on it. There, no doubt, you have an experiment; by means of the

⁶ These are nevertheless the senses in which Heidelberger (2003) uses the concept in his discussion of theory-ladenness.

vibration of this spot of light, this physicist minutely observes the oscillations of the piece of iron. Ask him now what he is doing. Is he going to answer: “I am studying the oscillations of the piece of iron carrying this mirror?” No, he will tell you that he is measuring the electrical resistance of a coil. If you are astonished and ask him what meaning these words have, and what relation they have to the phenomena he has perceived and which you have at the same time perceived, he will reply that your question would require some very long explanations, and he will recommend to you take a course in electricity. (Duhem 1974 [1906], p. 145)

Bogen and Woodward’s account does not seem to be able to accommodate the classical thesis of theory-ladenness. In fact, their account was (amongst other things) explicitly designed to rebut any form of theory-ladenness: theories don’t explain and predict what is observable (data) but just what is unobservable (phenomena). But this shall not be the major criticism of this paper. There are much stronger versions of theory-ladenness, which have not found much discussion in the literature, and which threaten to undermine Bogen and Woodward’s account quite severely.

Strong versions of theory-ladenness

After the above clarifications about the traditional concept of theory-ladenness (which I think is correct), and the epistemologically unproblematic concept of theory-drivenness, consider even stronger versions of theory-ladenness: the *principled neglect of data* due to theoretical predispositions (TLI), and theoretical reasons for belief in the reality of phenomena, which can prove to be critical in the *assessment of the reliability of the data* and the eventual acceptance of this phenomenon as being real (TLA). In the following I shall substantiate these two stronger forms of theory-ladenness by considering two historical cases which—rather ironically—have been quoted in support of Bogen and Woodward’s distinction. Neither TLI nor TLA seems to be compatible with Bogen and Woodward’s account, because theories, according to Bogen and Woodward, simply have no business in establishing the reliability of the data. So even if Bogen and Woodward were to accept the traditional form of theory-ladenness discussed above (which their account doesn’t seem to permit without major modifications), I don’t think it can handle either TLI or TLA.

Case study I: the zebra pattern of magnetic anomalies

The first case I am going to discuss is the discovery of the zebra pattern of magnetic anomalies, which has been discussed by Kaiser (1995) to defend Bogen and Woodward’s

notion of data and phenomena⁷. In the late 1950s and early 1960s it was found that the sea floor is not only positively but also negatively magnetised and that these magnetisations occur in stripes running in a north-south direction. Negative anomalies were conventionally shaded thus giving the impression of a “zebra” pattern (see FIG).



Fig. 1: The Zebra pattern of positive and negative magnetic anomalies off the East coast of Northern America.

This pattern received its currently accepted (and perhaps final) explanation in the so-called Vine-Matthews-Morley hypothesis. This hypothesis is essentially a combination of the hypothesis of geomagnetic field reversals (GFR) and the hypothesis of sea floor spreading (SFS). GFR is the idea that the poles of the earth’s magnetic field swap throughout time so that the North Pole becomes the South Pole and vice versa. SFS is slightly more complicated. It is the hypothesis that sea floor is formed at ridges by hot mantle rock that cools down at the surface of the Earth’s mantle after being elevated by convection currents within the mantle. The sea floor, which has been formed, then is pushed towards the periphery by succeeding hot mantle rock and submerges beneath the continental plates where it causes earthquakes and eventually is turned again into magma by the high temperatures within the mantle (see FIG).

⁷ For a discussion of Kaiser’s view and for many more details about this case can be found in Schindler (2007).

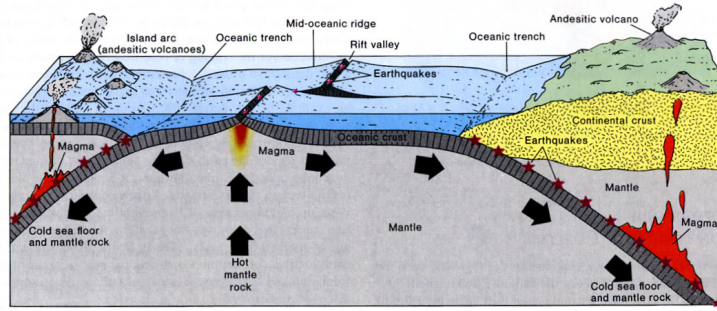


Fig. 2: Sea floor spreading

Magnetites within the molten rock which is elevated from the mantle to the surface align with the current geomagnetic field, and remain like that when they cool down. Thus, the past geomagnetic field is “fossilised”, as it were, and the striped pattern of FIG can be explained by an evolving sea floor whose material is magnetised according to past geomagnetic fields (see FIG).

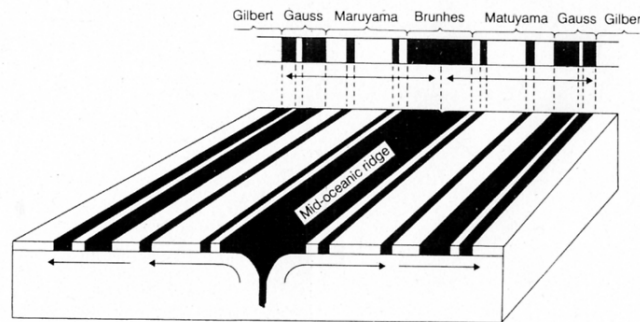


Fig. 3: The Vine-Matthew-Morley hypothesis as a combination of geomagnetic field reversals and sea floor spreading

It must be noted that neither GFR nor SFS were accepted as real phenomena within the scientific community at the time but were regarded as rather speculative. This changed when VMM combined SFS and GFR in order to explain the zebra pattern. Before VMM, the discoverer of the zebra pattern, Ron Mason of the *Scripps Institute of Oceanology* in California offered his own explanation of the pattern which is highly interesting for the purposes of this paper. Mason reasoned that the positive anomalies were likely to be caused by slabs consisting of “material highly magnetic by comparison with adjacent formation and almost certainly a basic igneous rock” (Mason 1958, 328). The concrete form of these slabs depended on various variables like their depth within the seafloor, their thickness, their width and their susceptibility to magnetisation (see FIG).

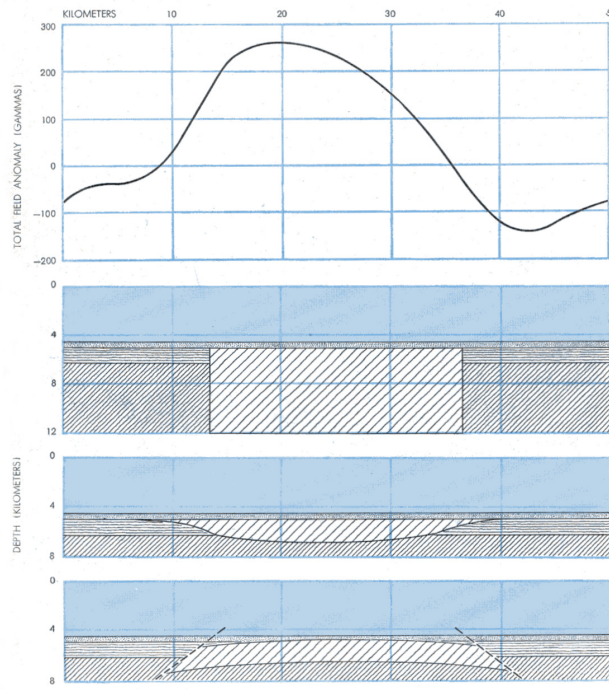


Fig. 4 A profile of the magnetic anomalies and Mason's models accounting for positive anomalies. Notice that positive and *negative* values are depicted. The next three figures are models of the sea-floor that are to account for the *positive* anomalies. The layers of each model starting from the top represent water, sediments, volcanics, and crust. Diagram adapted from Raff (1961, 154).

Crucially, Mason did not even try to account for the negative magnetisations. Retrospectively, Mason explained this thus:

I could have kicked myself for not thinking of the idea [VMM hypothesis], particularly because, *had I looked more carefully at our map, I would have realized that some of the seamounts might be reversely magnetized*, and this might have headed my thoughts in the right direction. (Mason 2003, 41)

Yet, as it turns out, Mason's own retrospective attempt to explain why he ignored the negative anomalies can be shown to be implausible. First, Mason shaded only the positive anomalies of the map of magnetic anomalies, not the negative ones. He must have been thus aware that there were negative anomalies. This is also clear from the description of the maps. For instance, Raff, Mason's co-worker wrote in the *Scientific American*:

The Lineated Pattern [...] is startlingly apparent when the positive anomalies are shown in grey and the *negative anomalies in white*"(Raff 1961, 147)

Notice that Raff spoke of a *lined* pattern, rather than of an (alternating) zebra pattern. This becomes particularly clear in the following quote by Mason:

This paper has been mainly concerned with the *north-south anomalies* because they are the *most striking feature of the magnetic map* [...] (Mason 1958, 328; my emphasis)

Now, despite appearances it needs to be emphasised that Mason and Raff's description and explanation of the lineated (rather than the "zebra") pattern cannot be easily dismissed as being just bad science. There is no record that anybody dismissed what they said, their articles made it into the most respected peer-reviewed scholarly journals, and no alternative proposals (until the VMM hypothesis in 1963, i.e. 5 years after Mason's first publication about the pattern) are reported anywhere. Even though one could of course question the judgment of the scientific community at the time, I think it is worth asking why they neglected the negative anomalies. Schindler (2007) has argued that Mason and Raff didn't explain the whole pattern (but just the subset of positive anomalies) because they either didn't possess the concept of SFS, or they didn't know how to apply it to the pattern in any case. Nowhere in their publications did they mention SFS. Even though they were well aware of GFR and even tried to apply it to the pattern, GFR is not explanatory of the pattern if not combined with SFS, as VMM did later. Because they didn't possess SFS, they also couldn't know about a corollary of SFS, namely transform faults (see FIG, right).

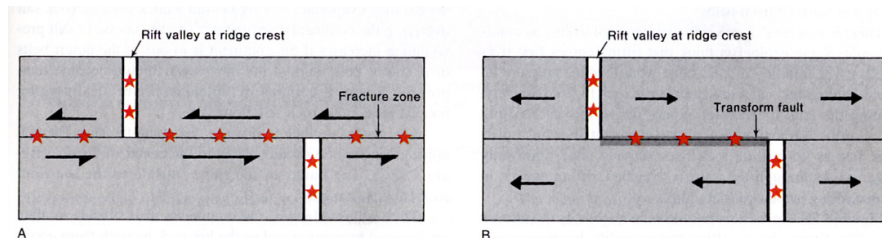


Fig. 5: Two hypotheses about the relationships between fracture zones and ridges. In both A and B the stars indicate expected earthquake distributions; the arrows rock motion. A: *Transcurrent fault*. It is assumed that the two ridges were once continuous across the fracture zone. B: *Transform fault*. It is assumed that the offset of the two ridges is original and invariant. Rock motion is caused by sea-floor spreading away from the ridges. Notice that a) only between the two ridges the respective rocks move in opposite directions and that as a consequence, b) seismic activity is limited to this area. This is what is in accordance with the data.

So rather than realising that the off-set of the ridges within the pattern was original (which is the case in transform faults), Mason and Raff were thus assuming that these off-sets were caused by seismic activity along the fault line (this being a feature of transcurrent faults; FIG, left):

The north-south magnetic lineation appeared throughout the area. *In several places the pattern was sharply broken along an east-west line, so that the striations above and below it did not match.* Although the significance of the *lineation* [not the alternation

of stripes] was—and still is—a mystery, the meaning of the discontinuities was immediately clear. (Raff 1961, 148; my emphasis)

Mason and Raff thus tried to reconstruct what they thought was the original *lineated* pattern by aligning the off-set ridges⁸. Here, of course, the positive anomalies were absolutely sufficient and the alternation between positive and negative anomalies of the pattern rather irrelevant. In other words, their ignoring of the negative anomalies of the pattern can be made due to the set of theoretical ideas that was available to them (or the lack thereof). Although a case of the traditional concept of theory-ladenness can easily be made with this case (perception of the pattern as lineation vs. perception of the pattern as alternation), this case supports an even stronger version of theory-ladenness according to which data are bluntly ignored if they are found irrelevant from a particular theoretical standpoint⁹. Even if Bogen and Woodward were to accept the traditional thesis of theory-ladenness, it is hard to see how their account could accommodate TLI. Bogen and Woodward hold that the reliability of the data can be established without the theory, which is sought to be tested. Now, as pointed out above, the reduction of data, according to Bogen and Woodward, is one of those procedures one applies when trying to achieve the reliability of the data independently of the theory to be tested. Nevertheless, in the case discussed in this section, the theory at stake clearly was involved in the reduction of data. In fact, it led to a *principled neglect* of some data in favour of others. Hence, Bogen and Woodward are simply mistaken on this point.

One may be inclined to dismiss my example on grounds that the theoretically-motivated reduction of the data did not lead to a phenomenon. Leaving aside that TLI can be observed in cases where this sort of data reduction *did* result in an actual phenomenon (see Schindler 2008), Mason & Raff's explanatory models were not attacked or criticised by the scientific community for their descriptions of lineated (rather than an alternating) pattern. And if one wants to defend an account that is descriptive of actual scientific practice (and this is clearly what Bogen and Woodward seek to do¹⁰) one cannot simply discard such cases because, clearly, Mason & Raff's work *was* seen as good scientific practice and the lineation

⁸ They realised however that the “displacements” had rather awkward features, actually characteristic of transform faults: “Horizontally displaced faults have always presented a puzzle: no distinctive topographic features mark their ends; the cracks simply stop” (Raff 1961, 148).

⁹ Kuhn has of course stipulated something like that in his *Structure of Scientific Revolutions*. However, note that Kuhn talked about this phenomenon only in the context of paradigms, which is of course controversial.

¹⁰ E.g.: “[...] we do not regard our discussion as an *argument* [...] Our discussion is rather intended as an empirical description of various features of scientific practice that have been overlooked or misdescribed in the philosophical literature.” (Bogen and Woodward 1988, p. 337; original emphasis)

of anomalies *was* clearly perceived as a genuine phenomenon at the time. It therefore appears plausible to give theories a more active role in establishing the reliability of data and in the discovery of the phenomena. This will become even more apparent in the other case I want to consider in the following.

Case study II: The discovery of Weak Neutral Currents

The discovery of weak neutral current (WNC) figures quite heavily in Bogen and Woodward's publications. It is quoted as a prime example for their notions of data and phenomena. Yet, as I shall argue in the following, the discovery of WNC does not support their account, when correctly construed.

WNC are a form of weak interaction between subatomic particles, mediated by the so-called Z^0 boson. In contrast to their charged current counterpart (mediated by a W^+ or W^- boson), WNC are characterised through their lack of muons (μ^-) in neutrino-nucleon scattering experiments (see FIG; right). Since only charged particles leave tracks in bubble and spark chambers (which are used to detect WNC) and since neutrinos are electrically neutral, the non-production of muons became the main identifier of WNC.

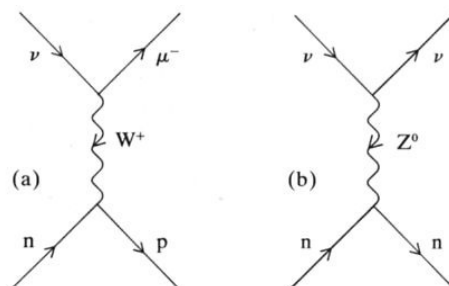


Fig. 6: Neutrino-nucleon scattering.

a) charged current event, mediated by a W^+ boson, which carries a positive charge from the reaction $\nu \rightarrow \mu^-$ to the reaction $n \rightarrow p$ (where ν = neutrino, n = nucleon, p = proton, μ^- = muon); b) neutral current event, mediated by an electrically neutral Z^0 boson (also characterised as massive analogue of the photon in electron scattering). Neutral current events produced in scattering are characterised by the charge remaining the same for incoming and outgoing particles (upper and lower parts of the diagrams respectively). From Pickering (1984).

This way of identifying WNC, however, is highly problematic because the detection of muons was not guaranteed. In spark chamber experiments, where the production and detection of muons is spatially separated, wide-angle muons could escape the detector¹¹. In bubble chamber experiments muons could get stuck in the shielding of the chamber. If one didn't

¹¹ On the other hand, with not enough space between generator and detector, hadrons could penetrate into the detector where they would be counted as muons.

take extra care in estimating these undetected events, one would end up counting as WNC events what in fact were merely charged current events. Moreover, in bubble chamber experiments, incoming neutrinos could knock off neutrons within the shielding, which in turn would propagate into the chamber where they would scatter off hadrons, thus emulating WNC events (FIG). The latter came to be known as neutron background and it became the main challenge in bubble chamber experiments to estimate this neutron background correctly in order to discern the genuine WNC events from the ones that only “mimicked” them.

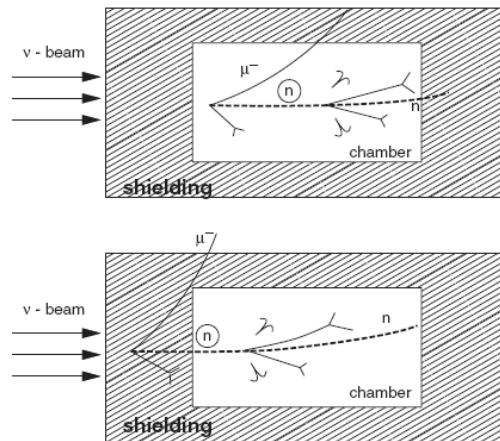


Fig. 7: Neutron stars in bubble chamber. The figure shows two forms of neutron stars triggered by a neutrino beam. *Above*: a neutrino hits a nucleus, producing a muon (μ^-), hadrons, and a neutron (n). The neutron, again, hits another nucleus, producing even more hadrons, but without producing a muon. The event caused by the neutron can unproblematically be associated with the neutrino beam (and hence be identified as a charged current event), which is why these events are called “associated events” (AS) (also called neutron stars (n^*)). All this happens within the visible chamber. *Below*: starting in the invisible shielding, making the muon event (μ^-) undetectable. The latter gives the *appearance* of a neutral current event (non-associated or “background event” (B)). The interaction length of the AS events within the chamber serves as the basis for estimating the number of unobservable B events by means of Monte Carlo programmes. Diagram from Haidt (2004); adapted.

In the history and philosophy of science literature, the discovery of WNC (with the pioneering work of Galison 1983) is not uncontroversial. Pickering (1984) has claimed that already bubble and spark chamber experiments in the 1960s were capable of producing WNC and that they actually did produce data which indicated the presence of WNC. The question then of course is why WNC were not discovered in the 1960s but rather in the 1970s. Pickering has argued that this was so because WNC were a “socially desirable phenomenon” (p. 109) in the 1970s after theoretical physicists had come up with the Salam-Weinberg model which postulated WNC. Pickering concluded that “particle physicists accepted the existence

of the neutral current because they could see how to ply their trade more profitably in a world in which the neutral current was real” (ibid., p. 87; my emphasis). Miller and Bullock (1994) however, have challenged Pickering’s interpretation by denying Pickering’s claim that the 1960s bubble chamber experiments were capable of producing a WNC “signal”. As Schindler (under review) has shown¹², Miller and Bullock’s argument doesn’t go through. On the other hand, I don’t think one needs to fall back on any sociological arguments in order to make sense of the discovery. But then, how does one explain that the discovery was made in the 1970s and not in the 1960s? Which conditions have to be in place for a discovery to be announced if it is not just the experimental evidence (after all, the latter *was* already available in the 1960s)? Perhaps, one should take a closer look at how the eventual discovery was made in 1974. In this year, both the HPWF and the CERN groups each published a paper, both of which are usually regarded as having established WNC as a genuine phenomenon. Here is how both groups, rather surprisingly for “discovery papers”, conclude¹³:

A possible, but *by no means unique, interpretation of this effect* [muonless events] is the existence of a neutral weak current (Benvenuti. et al. 1974, p. 800; my emphasis).

It has to be emphasized that the neutral current hypothesis is *not the only interpretation* of the observed events (Hasert et al. 1974, p. 20; added emphasis).

Likewise, a review article in *Physics Today* read:

Although both groups [Gargamelle and HPWF] suggest that they may be seeing neutral currents, they also offer alternative explanations. *And many experimenters are sceptical that either group has demonstrated the existence of neutral currents.* (Lubkin 1973, p. 17)

In other words, even in the articles that (retrospectively) came to be seen as “discovery papers”, quite obviously a fair amount of scepticism about the reality of WNC remained. But it comes even worse than that. As Perkins, a research at CERN in the mid 1960s, mid 70s, and mid 80s, has pointed out recently (see also FIG):

It is interesting to not that the HPWF result *is actually inconsistent with the Salam-Weinberg theory*, while the Gargamelle result shows a value of R [i.e., NC/CC] that *is*

¹² In this publication, Schindler discusses Bogen and Woodward’s characterisation of the discovery of WNC in terms of their notions of data and phenomena in more detail. There, he also has a few things to say about Mayo’s construal of the discovery in terms of her statistical error account.

¹³ The first quote is by the HPWF group the second by the CERN group. Note that the paper by the HPWF group was already submitted August 1973 and that the CERN group published a short paper in 1973, which they elaborated in the article that appeared a year later. See Pickering (1984) for details.

only about two-thirds of the present-day value [...]. The value deduced for $\sin^2\theta_W = 0.38 \pm 0.009$ has to be compared with the present value of 0.23. (Perkins 1997, p. 442)

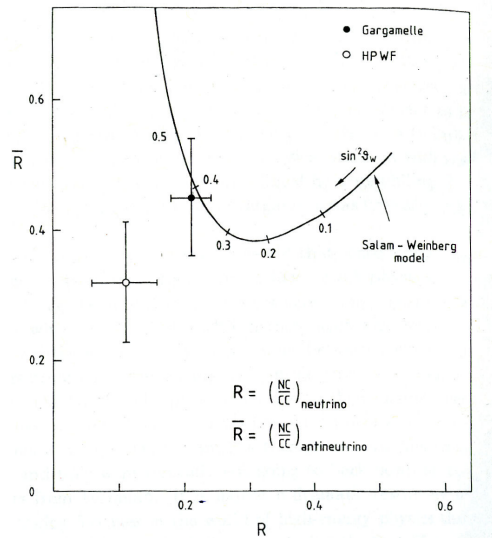


Fig. 8: Comparison of the results of the main Neutral Current experiments. The figure shows the results obtained from Gargamelle and HPWF as compared to the predictions of the Salam-Weinberg model. From Perkins (1997).

Hence, it turns out that the results by the CERN group, which were published in 1974 actually were not at all accurate when compared with currently accepted results, whereas the 1974 results of the HPWF group actually are inconsistent (!) with current results. In other words, the data of 1974 were not at all as compelling as we may think they should be given we take it that WNC were discovered in that year. Hence one may again wonder, with much more urgency than before, how the research community eventually became to be convinced of the reality of WNC in 1974? Since there wasn't any other evidence for WNC than the one provided by the experiments of the HPWF and CERN groups, we can assume that the data themselves obviously were insufficient for making a discovery claim. So the reasons for the researchers to accept WNC to exist must be extra-empirical. Now, I have already said that I'm not willing to resort to any socio-economical arguments here. I simply think they are not necessary. I certainly don't think physicists are as opportunistic as Pickering suggests when he says that they welcomed WNC just because it allowed them to "ply their trade more profitably". Nor do I think one should draw any other relativist conclusions. Rather, I believe the reasons why physicists accepted WNC as a genuine phenomenon—despite remaining doubts about the diminishment of neutral background—has to be sought in the properties of

the theory that postulated WNC. The Weinberg- Salam model¹⁴, proposed in 1967 but re-normalised in 1972, unified theories of electromagnetic and weak interactions into a single theory. But since the Weinberg-Salam model *required* the existence of WNC, the Weinberg-Salam model provided good reasons for the belief in WNC. Of course, the Weinberg-Salam model, by itself, would have been far from being sufficient for physicists to accept WNC as real. But as we saw above, the experiments quite apparently were not sufficient either for a discovery claim. Nevertheless the data combined with the theoretically motivated “need” for WNC did the trick! Although doubts about the conclusiveness of the experiments remained, these doubts were outweighed by the benefits the Weinberg-Salam model offered in terms of conceptually unifying electromagnetic and weak interactions.

Contrary to the conclusions of this section, Bogen and Woodward in their publications assumed all along that a clear case for the existence of WNC could be made on the basis of the experimental evidence only in 1974¹⁵. And yet, the discovery of WNC is only a case supporting their account if one makes this (false) assumption. But quite obviously, the experimental procedures aiming at achieving the reliability of the data were not sufficient for establishing the reality of WNC. Rather, it was the theory that was sought to be tested that provided good reasons for accepting WNC as real. Nothing in Bogen and Woodward’s account, however, allows for such.

Conclusion

In their well known article “Saving the phenomena”, Bogen and Woodward (1988) made four central claims: (i) phenomena in science, more often than not and contrary to traditional intuitions, are unobservable, (ii) the phenomena are inferred from observable data, (iii) the reliability of the data is secured on the basis of experimental and statistical means, (iv) theories do not predict or explain data but rather phenomena. As a corollary to (i) and (iv), Bogen and Woodward also rejected the thesis of theory-ladenness (roughly, what we see is influenced by the theories we hold). Although Bogen and Woodward did not mention any other forms of inference in their original article (making (ii) seem like a necessary condition for the inference of phenomena), they recently retreated on claim (ii) and conceded that their scheme should be complemented by other forms of inferences, most notably by theory-driven inferences. Their retreat on (ii) is largely unproblematic. Although one may be interested in

¹⁴ Even though it is called a “model”, philosophers for all its characteristics would probably prefer to call it a theory.

¹⁵ Bogen and Woodward are in good company here. See Galison (1983, 1987).

how Bogen and Woodward would spell out those kinds of inferences, which they haven't even mentioned in their publications to date, their retreat on claim (ii) seems to leave their other claims and their mutual consistency intact. Woodward (this volume) has furthermore partially retreated on claim (iv), conceding that data may well be described and interpreted in terms of theoretical vocabulary. Yet, as pointed out above this leads to serious conceptual difficulties, which Bogen and Woodward would need clarify. But regardless of this, there are much more serious questions to be raised about the descriptive adequacy of their account with respect to claim (iii) and its relation to claim (iv). Some of these questions arise from re-visiting two case studies, which Bogen and Woodward and others have quoted in support of the descriptive adequacy of their account. The first case study concerns the discovery of the so-called zebra pattern of magnetic anomalies (Kaiser 1991, 1995) and the second case study regards the discovery of weak neutral currents (Bogen and Woodward 1988, Woodward 1989, Woodward 2000). With respect to the first case study I showed that not only the traditional thesis of theory-ladenness is substantiated but a case can be made for an even stronger version of theory-ladenness according to which the *data are neglected* if they appear irrelevant from a certain theoretical perspective (I called this TLI). In the second case, a positive property of the theory at stake, namely the unifying power of the Salam-Weinberg model, obviously played a critical role in assessment of the *reliability* of the data (I called this TLA). Both case studies directly challenge claim (iii) and claim (iv) of Bogen and Woodward's account. Since Bogen and Woodward and others have depicted the first and the second case discussed here as paradigm examples for their distinction, I don't see how they could dismiss the cases as "untypical", "exceptions" or the like.

On a last note, despite my grave disagreements with Bogen and Woodward that I articulated in this paper, however, where I would agree with them is that there is a distinction to be made between data and phenomena and that most phenomena in science are unobservable. It is here where they have made a significant contribution to the philosophy of science.

Bibliography

- Benvenuti, A. et al. (1973), 'Early Observation of Neutrino and Antineutrino Events at High Energies', *Phys. Rev. Lett.*, 30 (1973), pp. 1084 - 1087.
- Bogen, J. and Woodward, J. (1988), "Saving the Phenomena", *Philosophical Review*, 97: 303-352.
- Cartwright, N. (1983), *How the Laws of Physics Lie*, OUP.

- Duhem, P. (1962), *The aim and structure of physical theory*, translated from the second French edition by Philip P. Wiener, Princeton: Princeton U.P.
- Galison, P. (1983), 'How the first neutral-current experiments ended', *Reviews of Modern Physics*, 55, pp. 477 – 509.
- Galison, P. (1987), *How experiments end*, Chicago: University of Chicago Press.
- Glymour, B. (2000), 'Data and Phenomena: A Distinction Reconsidered', *Erkenntnis*, vol. 52, no. 1, January, pp. 29-37(9).
- Hacking, I (1983), *Representing and Intervening*, Cambridge University Press.
- Hasert, F. J. et al. (1973), 'Observation of neutrino like interactions without muon or electron in the Gargamelle neutrino experiment', *Phys.Lett.* B46 (1973), pp. 138-140.
- Heidelberger, M. (2003), "Theory-Ladenness and Scientific Instruments in Experimentation", *The Philosophy of Scientific Experimentation*, Hans Radder (ed.), Pittsburgh, PA: University of Pittsburgh Press, pp. 138-151.
- Kaiser, M. (1995), "The Independence of Scientific Phenomena", in: In W. Herfel, W. Krajewski, I. Niiniluoto, & R. Wojcicki (Eds.), *Theories and models in scientific process. Poznan Studies in the Philosophy of Science and the Humanities 44* (pp. 179–200). Amsterdam: Rodopi.
- Mason, R. G. (1958), A magnetic survey off the west coast of the United States between latitudes 32° and 36° N. and longitudes 121° and 128°W.: Royal Astronomy Society. *Geophysical Journal*, v.1, pp. 320-329.
- Mason, R. G. (2003), 'Stripes on the Sea Floor', in: Oreskes, N. (ed.) (2003), *Plate Tectonics, An Insider's History of the Modern Theory of the Earth*, Cambridge: Westview Press.
- Miller, A. and W. Bullock (1994), 'Neutral Currents and the History of Scientific Ideas', *Studies in History and Philosophy of Science*, Vol. 25, No. 6, pp. 895-931.
- Lubkin, G. B. (1973), *Physics Today*, vol. 26, November, 17, pp 19-17.
- Perkins, D. (1997), 'Gargamelle and the Discovery of Neutral Currents', in: Hoddeson, L. et al., *The rise of the standard model: particle physics in the 1960s and 1970s*, Cambridge: Cambridge University Press.
- Pickering, A. (1984), 'Against Putting the Phenomena First', *Studies in History and Philosophy of Science*, 15: 85-114.
- Schindler, S. (2007), "Rehabilitating Theory. The Refusal of the bottom-up construction of Scientific Phenomena", *Studies in the History and Philosophy of Science*, Vol. 38, Issue 1, March, pp. 160-184.

- Schindler, S. (2008), "Model, Theory and Evidence in the Discovery of the DNA Structure",
The British Journal for the Philosophy of Science. **59** (2008), 619–658
- Schindler, S. (under review), "How to Discern a Physical Effect from Background Noise: The
Discovery of Weak Neutral Currents".
- Raff, A. D. (1961), 'The Magnetism of the Ocean Floor', *Scientific American*, October, vol.
205, no. 4, pp. 320-329.
- Van Fraassen, B. (1980), *The Scientific Image*, Oxford: Clarendon.
- Woodward, J. (1989), „Data and Phenomena“, *Synthese* 79, pp. 393-472.
- Woodward, J. (2000), "Data, Phenomena, and Reliability", *Philosophy of Science*, Vol. 67,
Supplement, Proceedings of the 1998 Biennial Meetings of the Philosophy of
Science Association, Part II: Symposia Papers, pp. S163-S179.
- Woodward, J. (this volume), "xxxx"