



Universiteit
Leiden
The Netherlands

Phenomena and Patterns in Data Sets

McAllister, J.W.

Citation

McAllister, J. W. (1997). Phenomena and Patterns in Data Sets. Retrieved from <https://hdl.handle.net/1887/8118>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/8118>

Note: To cite this publication please use the final published version (if applicable).

JAMES W. McALLISTER

PHENOMENA AND PATTERNS IN DATA SETS

ABSTRACT. Bogen and Woodward claim that the function of scientific theories is to account for 'phenomena', which they describe both as investigator-independent constituents of the world and as corresponding to patterns in data sets. I argue that, if phenomena are considered to correspond to patterns in data, it is inadmissible to regard them as investigator-independent entities. Bogen and Woodward's account of phenomena is thus incoherent. I offer an alternative account, according to which phenomena are investigator-relative entities. All the infinitely many patterns that data sets exhibit have equal intrinsic claim to the status of phenomenon: each investigator may stipulate which patterns correspond to phenomena for him or her. My notion of phenomena accords better both with experimental practice and with the historical development of science.

1. THE PROBLEM WITH PATTERNS

It has long been acknowledged by observers of science that even our most highly regarded scientific theories have difficulty in accounting for data points, which are the outcomes of individual observations and experiments: on any straightforward reading, most theories are contradicted by the vast majority of the data points that are gathered to test them.

We may respond to this finding in two ways. We may admit that, it being indeed the task of scientific theories to account for data points, most of the theories that scientists have ever put forward are unsatisfactory. Alternatively, we may claim that the incapacity to account for data points is not grounds for dissatisfaction with scientific theories, because the function of theories is to account for something other than data points.

A fresh version of the latter response has been advanced by James Bogen and James Woodward (1988): they hold that the explananda of scientific theories are not data points but what they term 'phenomena'. On Bogen and Woodward's account, while phenomena are not observable in any interesting sense of this term, they can be detected by scientists in data sets: for example, the weak neutral current is a phenomenon that is detected by particle physicists in the data contained in bubble chamber photographs. Bogen and Woodward's account of phenomena has been endorsed with minor modifications by James R. Brown (1994, pp. 117–141): as a further example of phenomenon, Brown cites the convertibility

Erkenntnis **47**: 217–228, 1997.

© 1997 Kluwer Academic Publishers. Printed in the Netherlands.

of a unit of mechanical energy into a unit of thermal energy, which was first detected by Joule in data on the heating effects of mechanical agitation and friction.

Bogen and Woodward regard phenomena as fundamental constituents of the world. They believe consequently that what phenomena there are, and whether such-and-such constitutes a phenomenon, is not a matter of stipulation on the part of investigators. For example, Bogen and Woodward (1988, p. 321) write: "It should be clear that we think of particular phenomena as in the world, as belonging to the natural order itself and not just to the way we talk about or conceptualize that order." Brown (1994, pp. 125–128) demonstrates a similar conviction that what phenomena there are is not a matter of stipulation, portraying phenomena as natural kinds of a particular sort.

Bogen and Woodward believe further, perhaps on the grounds that phenomena are fundamental constituents of the world, that the world contains relatively few of them. For example, the reader gains the impression from Woodward (1989, p. 393) that phenomena are scarce enough that we may name and list them. As Brown (1994, p. 125) puts it, "The world is full of data, but there are relatively few phenomena."

Since Bogen and Woodward claim that scientists can detect phenomena in data sets, they must explain in what relation phenomena stand to data. On their account, phenomena manifest themselves in patterns in data sets, presumably in virtue of being causes of such patterns. In other words, Bogen and Woodward assert that each phenomenon corresponds to one or more patterns exhibited by data sets that have been or may be collected. In consequence, the way in which scientists detect phenomena is by identifying the patterns in data sets to which they correspond. Woodward (1989, pp. 396–397) explains: "The problem of detecting a phenomenon is the problem of [. . .] identifying a relatively stable and invariant pattern of some simplicity and generality with recurrent features – a pattern which is not just an artefact of the particular detection techniques we employ or the local environment in which we operate." Detecting elementary particle phenomena from bubble chamber photographs involves recognizing the patterns that correspond to these phenomena in the data contained in the photographs, according to Woodward (1989, p. 419): "Commonly, bubble chamber photographs are first scanned by workers who have no background in physics, but have been taught to recognize certain potentially interesting patterns." The portrayal of elementary particle phenomena as corresponding to patterns in bubble chamber photographs is repeated in a later discussion by Bogen and Woodward (1992, p. 597): "Human perceivers had to look at such photographs, discriminate among them, draw

certain photographs to the attention of others, recognize characteristic patterns, and measure tracks as part of the complex series of procedures and arguments which led to the detection of neutral currents.” Brown (1994, p. 141) writes similarly: “Phenomena are natural kinds (or patterns) that we can picture.”

I suggest that the claim that phenomena correspond to patterns in data sets renders Bogen and Woodward’s account of phenomena incoherent. More specifically, it is incompatible with their claim that what phenomena there are is not a matter of stipulation.

Any data set – whether it takes the form of a pen trace, sequence of numbers, or array of dots – can be regarded as the sum of two components: a relatively simple and regular pattern and a certain level of noise. This is a consequence of the fact that any mathematical function, however irregular, can be expressed as the sum of a relatively simple and regular function, such as a harmonic function, and another term that accounts for the difference:

$$F(x) = (a \sin \omega x + b \cos \omega x) + R(x)$$

On Bogen and Woodward’s view, as we have seen, the former component is that to which a phenomenon corresponds. Woodward acknowledges that a data set contains a certain level of noise as well as a pattern or ‘signal’ corresponding to a phenomenon:

The sophisticated investigator does not proceed by attempting to explain his data, which typically will reflect the presence of a great deal of noise. Rather, the sophisticated investigator first subjects his data to a great deal of analysis and processing [. . .] in an effort to separate out the phenomenon of interest from extraneous background factors. It is this extracted signal rather than the data itself which is then regarded as a potential object of explanation by general theory. (Woodward 1989, p. 397)

Many general scientific theories [. . .] succeed because investigators first try to filter out the noise, background, and confounding factors in their data, and only then attempt to extract the signal or phenomenon. (Ibid., pp. 451–452)

However, any given data set can be described as the sum of any one of infinitely many distinct patterns and a corresponding incidence of noise. In other words, there are infinitely many distinct descriptions of a given data set as “Pattern *A* + noise at *m* percent”, “Pattern *B* + noise at *n* percent”, and so on. Among these, there is a description of the data set as a pattern exhibited with zero noise – that is, as the (perhaps exceedingly complex) pattern that accords exactly with the data constituting the set. For example, a set of meteorological data may be described as the sum of any one of infinitely many patterns with periods of various lengths (diurnal, annual, secular, and so on) and a corresponding noise level; among these is a pattern of infinite period that accords in every detail with the data constituting the

set, and which is consequently exhibited with zero noise. Similarly, a data set on economic activity exhibits infinitely many patterns corresponding to economic cycles of various durations, each manifested in the data with a certain noise level. A data set on the orbit of Mars exhibits with zero noise a pattern that reproduces the data in all particulars; it exhibits with various nonzero noise levels infinitely many other patterns, corresponding to particular circles, particular ellipses, and more complex curves.

The statement that a data set can be described as the sum of any one of infinitely many distinct patterns and a corresponding incidence of noise holds true regardless of whether the data set has been cleansed of observational, experimental, and other errors. As in information theory, 'noise' here denotes not a margin of error or factual inaccuracy in data, but rather the purely mathematical discrepancy between a given pattern and a data set.

I presume that Bogen and Woodward would claim that only some of the infinitely many patterns that data sets exhibit correspond to phenomena. Otherwise, they would contradict their presupposition that phenomena, being fundamental constituents of the world, are relatively few. We therefore pose the question: on Bogen and Woodward's account, in what respect do the patterns that correspond to phenomena differ from the other patterns that data sets exhibit?

If Bogen and Woodward's account is to succeed, they must be able to answer this question. They claim that phenomena are the explananda of scientific theories and that phenomena can be detected in data sets: if they cannot specify which patterns exhibited by data sets correspond to phenomena, their account is ultimately unable to identify what the explananda of scientific theories are. At the same time, however, Bogen and Woodward are precluded from answering that a pattern corresponds to a phenomenon merely in virtue of the fact that a given scientific theory portrays it as corresponding to a phenomenon. Bogen and Woodward's contention is that scientific theories describe phenomena: if they then identified phenomena as the features of the world that particular scientific theories describe, their account would become vacuous. Thus, Bogen and Woodward must be able to say without relying on scientific theories in what respect the patterns that correspond to phenomena differ from the other patterns that data sets exhibit.

To answer this question, they will have to identify some property that patterns that correspond to phenomena have and that other patterns lack. Woodward (1989, pp. 396–397) suggests for example that the patterns that correspond to phenomena are those that are "stable and invariant", have "some simplicity and generality", and show "recurrent features". However,

identifying a property of the patterns that correspond to phenomena is not enough. Any property that we may choose is shown by infinitely many patterns, differing from one another by an additive or multiplicative factor; and all these patterns are exhibited in any data set, with noise levels that range from zero to infinity. For example, to say that a particular pattern in a given data set is stable and invariant means presumably that this pattern can be discerned also in a larger data set, or even in the set of all possible data. But this larger data set will exhibit infinitely many patterns, with noise levels that range from zero to infinity. So the property of stability and invariance fails to identify any finite number of patterns as those that correspond to phenomena. The property 'some simplicity and generality' has the same shortcoming: claiming that the patterns that correspond to phenomena are those that show at least such-and-such a degree of simplicity implies that infinitely many patterns in any data set, exhibited with noise levels that range from zero to infinity, correspond to phenomena. (In passing, I note that claiming that the patterns that correspond to phenomena are those that show at least a particular degree of simplicity makes what counts as a phenomenon depend on a stipulation, contrary to Bogen and Woodward's account, since the degree of simplicity in question would have to be stipulated by investigators.) The property of showing recurrent features suffers from the same problem, as would any other property of patterns that we might specify.

In order to make the number of patterns that correspond to phenomena finite, Bogen and Woodward would have to say that the patterns that correspond to phenomena are those that, as well as having some specified property, are exhibited by data sets with a noise level no higher than a specified value. They could claim, for example, that the patterns that correspond to phenomena are those that, as well as being stable and invariant, having a certain degree of simplicity, or showing recurrent features, are exhibited in data sets with no more than a specified noise level. By fixing this noise level at a suitably low value, the number of patterns that correspond to phenomena can be made as small as Bogen and Woodward wish.

This noise level could be fixed at zero or at a specified nonzero value. Neither of these options can be endorsed by Bogen and Woodward, however. If they deemed phenomena to correspond only to the pattern that a data set exhibits with zero noise, Bogen and Woodward would frustrate their own aims. They set out to identify some feature in data sets for which scientific theories account better than for individual data points, but the pattern that a data set exhibits with zero noise reproduces exactly the data constituting the set. If they deemed phenomena to correspond only to the patterns that a data set exhibits with some specified nonzero noise level,

then – since a data set lends itself equally readily to being described as containing any amount of noise – this noise level would have to be stipulated by investigators. This stipulation would fix which phenomena the world contains, so Bogen and Woodward would be unable to maintain that what phenomena there are is not a matter of stipulation.

Thus, the premises of their account make it impossible for Bogen and Woodward to answer the question that I pose, in what respect do the patterns that correspond to phenomena differ from the other patterns that data sets exhibit. Yet their account is unacceptable if it does not provide an answer. From this I conclude that Bogen and Woodward's account of phenomena is incoherent.

Bogen and Woodward might attempt the following responses. First, they might argue that investigators approach data sets with conjectures about which phenomena exist, arising from theoretical commitments, background assumptions, and the like; and that these conjectures lead investigators to believe that, say, patterns *A* and *B* in a given data set are those that correspond to phenomena. Clearly, this response falls short of what is required. It suggests a route by which investigators come to believe that given patterns correspond to phenomena (a claim about the psychology of investigators), whereas my challenge is to specify the respect in which the patterns that correspond to phenomena differ from other patterns (a question about the ontology of the world). For Bogen and Woodward's account to succeed, they must be able to tell us independently of the beliefs of investigators in what respect the patterns that correspond to phenomena differ from those that do not: otherwise, their concept of phenomenon is unusable. Nor can Bogen and Woodward claim that investigators have learned to recognize the patterns that correspond to phenomena from earlier experience analyzing data sets: this would be possible only if the investigators knew which patterns correspond to phenomena in these earlier data sets – exactly the point at issue.

For a similar reason, as we have already seen, it is not open to Bogen and Woodward to respond that the patterns in data sets that correspond to phenomena are those that are explained by or embedded in the well-confirmed scientific theories of a particular age or the theories on which science may ultimately converge. The claim that Bogen and Woodward defend is that the objective features of the world for which scientific theories offer an explanation are phenomena. They may not thereafter define phenomena as the features that correspond to patterns in data sets that are explained by particular scientific theories. If their account is not to be vacuous, they must be able to pick out independently of the content of scientific theories which patterns in data sets correspond to phenomena.

Nor, lastly, is my challenge averted by Bogen and Woodward's assumption that the patterns in data sets that correspond to phenomena are those that are caused by phenomena. Presumably all patterns in data sets are caused to some extent by phenomena, and none is caused exclusively by a given phenomenon. So this response fails to pick out any subset of patterns as those that correspond to phenomena. Even if Bogen and Woodward were to claim that only some patterns in data sets are caused by phenomena, this response would still offer no way of distinguishing which patterns these are unless we knew which phenomena there exist.

Bogen and Woodward's account could withstand my challenge only if some of the patterns that data sets exhibit appeared labelled in some way as those that correspond to phenomena. Woodward (1989, p. 438) suggests that he believes this to occur: "Detecting a phenomenon is [. . .] like fiddling with a malfunctioning radio until one's favorite station finally comes through clearly." The flaw in this analogy lies in the fact that, as we tune the radio, one of the patterns exhibited by the electromagnetic field – the station's broadcast – appears labelled as the pattern to which we should pay heed. In contrast, all the infinitely many patterns that are exhibited in a data set collected in scientific research have equal claim to correspond to phenomena.

2. THE CHOICE OF PHENOMENA

I now propose an account of phenomena alternative to that of Bogen and Woodward. Let us consider afresh which of the infinitely many patterns exhibited by a data set are those that we might regard as corresponding to phenomena. The objections against regarding phenomena as corresponding either to the patterns that data sets exhibit with zero noise or to all patterns that data sets exhibit are overwhelming: as we have seen, the former notion of phenomena would duplicate that of individual data points, while the latter would have the consequence that infinitely many phenomena exist. The sole remaining option is to say that phenomena correspond to the patterns that a data set exhibits with a noise level no higher than some specified nonzero value. As a data set lends itself equally readily to being described as containing any amount of noise, this value will have to be stipulated by investigators. This stipulation amounts to fixing which patterns correspond to phenomena, and thus which features of the world count as phenomena. In consequence, this option yields an investigator-relative notion of phenomena.

This is the option that I select. My account of phenomena is thus the following. Investigators analyzing data sets encounter infinitely many pat-

terns exhibited with various noise levels. Each investigator designates some of these patterns, picked out on any criterion that he or she may choose, as the patterns that correspond to phenomena. Far from denoting a small number of fundamental constituents of the world, the term 'phenomenon' is on my account a label that investigators apply to whichever patterns in data sets they wish so to designate. Thus, on my account, which patterns count as those corresponding to phenomena is entirely a matter of stipulation by investigators. Typically, the patterns that investigators regard as corresponding to phenomena are those that they intend to study or hope to explain.

As will be clear, the ontology of my account differs from that of Bogen and Woodward. They hold that the world is composed of a small number of fundamental constituents, 'phenomena', which cause patterns in data sets. On my account, in contrast, the world is a complex causal mechanism that produces data in which infinitely many patterns can be discerned. Some of these patterns are taken by investigators as corresponding to phenomena, not because they have intrinsic properties that other patterns lack, but because they play a particular role in the investigators' thinking or theorizing. The two accounts differ also in their epistemological and methodological implications. Bogen and Woodward hold that investigators discover from data sets which phenomena there exist; on my account, investigators discover the patterns that are exhibited in data sets, but stipulate that some of these correspond to phenomena.

To compare the merits of the two accounts, let us turn to Bogen and Woodward's example of an investigation into the melting of lead. As they write (1988, p. 308), "Even when the equipment is in good working order, and sources of systematic error have been eliminated, the readings from repeated applications of the same thermometer to the same sample of lead will differ slightly from one another, providing a scatter of results. These constitute data." Bogen and Woodward take it for granted that investigators will discern in this data set a pattern consisting of a horizontal line, corresponding to a phenomenon that we might call 'the invariance of the melting point of lead'. The rest of their discussion depends on this assumption. For example, they take it for granted that investigators will calculate the mean of the distribution of data points, reporting this as the melting point of lead.

But consider the alternatives. The same data set exhibits also infinitely many patterns other than a horizontal line, each exhibited with a certain noise level. According to each of these other patterns, the melting point of lead varies, as if it depended on some physical parameter, perhaps as yet unidentified. All these patterns have the same status or degree of privilege: they all have equal claim to the status of phenomenon. Investigators are

free to take any of these patterns as their explanandum and as corresponding to a phenomenon. The pattern that counts as the phenomenon of the melting of lead for a physicist may differ from the one that counts as the phenomenon for a chemist; the pattern that counts as the phenomenon for a twentieth-century scientist may differ from the one that counts as the phenomenon for a twenty-first-century scientist. One scientist may study the approximate invariance of the melting point of lead while another studies its dependence on air pressure, electrostatic charge, or magnetic field. It could never be established that one of these patterns truly corresponds to a phenomenon while the others do not: the data set exhibits them all equally, albeit with differing noise levels. I conclude that ‘phenomenon’ is a status that investigators confer on some patterns in the data set, not a fundamental constituent of the world that manifests itself in particular patterns.

Anyone who believes that investigators are compelled to see a pattern consisting of a horizontal line in the data set on the melting of lead should reflect on the innumerable cases in which we choose to interpret similar data sets as exhibiting patterns other than straight lines. For example, as Brown (1994, pp. 136–138) points out, the data from the Franck–Hertz experiment (taken to provide evidence for the existence of discrete energy levels within the atom) is interpreted by physicists as the sum not of a straight line and a certain noise level, but of a curve showing evenly spaced peaks and a certain other noise level. In this case too, investigators have chosen to attribute the status of phenomenon to one of the infinitely many patterns that they discern in a data set.

Bogen and Woodward might respond that present-day investigators hold to certain background assumptions, such as a theory of molecular structure, which cause them to expect lead to have an invariant melting point. It is in the light of these assumptions that investigators analyzing a data set on the melting of lead pick out the horizontal line as the pattern that corresponds to a phenomenon. In evaluating this response, we must again distinguish between the psychological question, “By what route do investigators come to believe that a given pattern corresponds to a phenomenon?”, and the ontological question, “In what respect does a pattern that corresponds to a phenomenon differ from other patterns?” This response offers an answer to the first question. But it is the ontological question that we face here. According to Bogen and Woodward’s account, phenomena are fundamental constituents of the world whose identity does not depend on stipulations by investigators. Therefore, Bogen and Woodward must tell us independently of the theories and beliefs of investigators in what respect the patterns that correspond to phenomena differ from other patterns exhibited in data sets. They are unable to do this.

My account of phenomena also accords better than that of Bogen and Woodward with the historical development of science. It frequently occurs that, while scientists of one epoch attempt to account for patterns that a data set exhibits with a certain noise level, scientists of a later epoch study patterns that are exhibited with lower noise levels. Consider as an example the theories of planetary motions of Aristarchus, Kepler, Newton, and Einstein: each describes and attempts to account for a pattern that is exhibited by data on planetary orbits with a lower noise level than the pattern described by the previous theory. Each of these patterns may be taken to correspond to a phenomenon. Asked in what the phenomenon of planetary orbits consists, Kepler would have replied “In the fact that, with such-and-such a noise level, they are ellipses”, while Newton would have replied “In the fact that, with such-and-such a (lower) noise level, they are particular curves that differ from ellipses, because of the gravitational pull of other bodies”. Thus, phenomena – understood as the patterns in data sets that investigators choose to model – vary from one investigator to another.

Needless to say, physical occurrences (in this example, the planets’ motions in space) are identical for all investigators. But no scientific theory attempts to account for occurrences in all details – this would correspond to accounting for every data point. Instead, theories account for specified aspects of occurrences, such as the planetary orbits’ being, with such-and-such a noise level, ellipses. Which aspect of the occurrences is singled out for explanation varies from one investigator to another, notwithstanding the fixedness of physical occurrences themselves.

Bogen and Woodward’s notion of phenomena cannot make sense of the succession of theories from Aristarchus to Einstein. In view of their supposition that phenomena are fundamental constituents of the world and few in number, Bogen and Woodward presumably hold that there exists in the world a unique phenomenon of planetary orbits which manifests itself in a pattern in a data set. But which pattern is this? All the infinitely many patterns shown by a data set on planetary orbits have equal intrinsic status, so there are no grounds for claiming that either Aristarchus, Kepler, Newton, Einstein, or anyone else has correctly identified the pattern to which the phenomenon of planetary orbits corresponds while the others have failed. Nor are there grounds for claiming that any one of these scientists has correctly recognized which is the genuine explanandum of theories about planetary orbits while the others have been mistaken. It is preferable to claim, as I do, that we confer the status of phenomenon on one or other of these patterns depending on the interests that we have.

These examples show that the following statement by Woodward (1989, p. 397) can be endorsed only with some qualification: “Contrary to what

many accounts of scientific explanation might lead one to expect, by no means everything that happens is a potential object of theoretical explanation. Figuring out what one should even try to explain – what the phenomena are in a given domain of inquiry – and what is mere noise is [. . .] an important aspect of scientific investigation, especially in relatively immature areas of inquiry like the social sciences.” It is clearly true that no investigator attempts to explain all the patterns that a data set exhibits. However, which patterns one should try to explain is not dictated by the world: it is open to investigators to decide. Thus, investigators must figure out what they should try to explain only in the sense that they must select the features of data sets that they wish to attempt to explain, and not in the sense that they must discover which ought to be explained.

In conclusion, I endorse Bogen and Woodward’s claim that the explananda of scientific theories are not data points but phenomena, and their suggestion that phenomena correspond to patterns in data sets. But I reject their simultaneous claim that what phenomena there are, and whether such-and-such constitutes a phenomenon, is not a matter of stipulation on the part of investigators. On the contrary, I suggest that any of the infinitely many patterns that data sets exhibit may be taken as the explananda of scientific theories: which patterns are so taken, and are thereby considered to be the patterns that correspond to phenomena, is stipulated by investigators.

ACKNOWLEDGEMENTS

I thank James Bogen (Pitzer College), James R. Brown (University of Toronto), and two anonymous referees of this journal for comments on previous versions.

REFERENCES

- Bogen, J. and J. Woodward: 1988, ‘Saving the Phenomena’, *Philosophical Review* **97**, 303–352.
Bogen, J. and J. Woodward: 1992, ‘Observations, Theories and the Evolution of the Human Spirit’, *Philosophy of Science* **59**, 590–611.
Brown, J. R.: 1994, *Smoke and Mirrors: How Science Reflects Reality*, Routledge, London.

Woodward, J.: 1989, 'Data and Phenomena', *Synthese* **79**, 393–472.

Manuscript received October 7, 1996

Faculty of Philosophy
University of Leiden
P.O. Box 9515
2300 RA Leiden
The Netherlands