Synthese (2011) 182:165–179
DOI 10.1007/s11229-009-9618-5

Data and phenomena: a restatement and defense

James F. Woodward

Received: 20 May 2009 / Accepted: 3 June 2009 / Published online: 7 July 2009

Abstract This paper provides a restatement and defense of the data/ phenomena distinction introduced by Jim Bogen and me several decades ago (e.g., Bogen and Woodward, The Philosophical Review, 303–352, 1988). Additional motivation for the distinction is introduced, ideas surrounding the distinction are clarified, and an attempt is made to respond to several criticisms.

Keywords Data · Phenomena · Inductive inference

1 Introduction

Two decades ago, Jim Bogen and I published a paper (Bogen and Woodward 1988) in which we introduced a distinction between data and phenomena and claimed that this had important implications for how we should understand the structure of scientific theories and the role of observation in science. This initial paper was followed by a series of papers on related themes (Bogen and Woodward 1992, 2005; Woodward 1989, 2000). In the intervening years, our ideas have been embraced by some (e.g., Kaiser 1991) and extensively criticized by others (Glymour 2000; McAllister 1997; Schindler 2007). In this essay, I want to place some of our ideas within a more general context and provide some additional background and motivation for them. Along the way I will attempt to clarify and, where appropriate, correct some claims in our original paper. I will also attempt to respond to some of the criticisms that have been leveled at our account.

J. F. Woodward (✉)
Division of the Humanities and Social Sciences, California Institute of Technology,
MC 101-40, Pasadena, CA 91125, USA
e-mail: jfw@hss.caltech.edu

Bogen and Woodward 1988, advocated a three-level picture of scientific theory or, more accurately, of those theories that were in the business of providing systematic explanations. Explanatory theories such as classical mechanics, general relativity, and the electroweak theory that unifies electromagnetic and weak nuclear forces were understood as providing explanations of what we called phenomena—features of the world that in principle could recur under different contexts or conditions. For example, the magnitude of the deflection of starlight by the sun and the gravitational red shift are phenomena that are explained by General Relativity (GR) and the existence of neutral currents is explained by the electroweak theory. The melting of lead at 327.5°C is a phenomenon which is explained by the character of the electron bonds and the presence of so-called "delocalized electrons" in samples of this element. As explained in our original paper, in speaking of theories as "explaining" phenomena, we had in mind the capacity of theories to provide relatively detailed and systematic explanations, often taking the form of derivations from basic laws and principles. We distinguished such explanations from so-called singular causal explanations, in which it is merely specified that some causal factor plays a role in the production of an outcome, but without specifying in a detailed, quantitative way how the outcome depends on this factor or which other factors influence the outcome. Thus, GR does not just tell us that the gravitational field of the sun "causes" starlight deflection, but allows for a principled derivation of the magnitude of this deflection.

Data are public records produced by measurement and experiment that serve as evidence for the existence or features of phenomena. The data from Eddington's expedition of 1919 that provided evidence for a value for the deflection of starlight by the sun that was predicted by GR took the form of photographic plates reflecting stellar positions and some of the data serving as evidence for the existence of neutral currents took the form of bubble chamber photographs. Data serving as evidence for the value of the melting point of lead might take the form of a record of temperature readings taken from a thermometer of some particular design. Investigators attempt to design experiments and arrange measurement apparatus in such a way that data produced by these reflect features of the phenomena they are trying to detect. However, typically (even in a well-designed experiment or measurement) data will reflect the influence of many other causal factors as well, including factors that have nothing to do with the phenomenon of interest and instead are more idiosyncratic to the measurement apparatus and experimental design employed and its local environment. Eddington's photographic plates changed in dimension because of changes in the temperature of the surrounding air—an influence for which he needed to correct in reporting his estimate of starlight deflection—and this influenced his data. Eddington's data also reflected the operation of the photographic and optical technology he employed, his decisions about the positioning of his telescopes and cameras and so on. In the case of the melting point of lead, repeated measurements, even with a thermometer of the same design and the same procedure for taking measurements, will result in a scatter of data (perhaps reflecting variations in the temperature of the surrounding air, idiosyncracies of the visual perception of the observers taking the measurement and so on). Schematically, if we think of the melting point as a fixed quantity M, then even if the measurement procedure is working properly the data d_i observed on different occasions of measurement will be some function f of M and of additional causal factors u_i which will take different values on different occasions of measurement: $d_i = f(M, u_i)$. The researcher will be able to observe the measurement results d_i but will usually not be able to measure or observe the values of all of the additional causal factors u_i. Her goal will nonetheless be to extract from the data d_i, an estimate of the value of M. This may be accomplished if the researcher knows or has good reason to believe that various additional assumptions are satisfied: for example, as discussed briefly below, if the researcher knows that f is additive (e.g., $d_i = M + u_i$) and the distribution of the u_i is normal, then the mean value of the thermometer readings d_i will, in a sense that can be specified (see below) yield a reliable (unbiased, minimum variance) estimate of M.

In our 1988 paper, Bogen and I claimed that while theories like *GR* aimed to provide systematic explanations of phenomena they "typically" or usually did not provide (and did not aim at providing) similar explanations of data. (I reiterate that "systematic explanation" meant something like the provision of a principled, non-ad hoc derivation from the basic laws and principles. For some reservations about these claims about typicality, see Sect. 3 below.) Fundamental theories like GR fail to provide such explanations of data in part because, as noted above, data commonly reflect the causal influence of many different factors besides those that fall within the purview of any single general theory. For example, GR is a theory of gravitational phenomena, not a theory that purports to explain or provide derivations concerning the behavior of cameras and optical telescopes or Eddington's decisions about experimental design. More importantly, Bogen and I claimed there is often no obvious scientific rationale or motivation for attempting to provide detailed systematic explanations of data. Instead, the role of data is to serve as evidence for the existence of phenomena and to play this role, researchers do not need to exhibit systematic explanations/derivations of data, either from fundamental theories or from such theories supplemented by hypotheses about the operation of measurement apparatus etc.

In denying that fundamental theories typically provided systematic explanations of data, we made it clear that we did not mean to deny that in well-functioning measurements phenomena will figure as one causal factor among many in the production of data. When a thermometer is used to measure the melting point of lead, the researcher certainly hopes that the data obtained will reflect the causal influence of the temperature of the sample as it melts, along with various other factors. In this sense, the melting point M will figure in a singular causal explanation of the data. However, neither the theories that explain why lead has the melting point that it does nor information about the value of M will provide what we have been calling detailed, systematic explanations (or non-trivial derivations of) of the values of the observed data d_i. To provide such explanations one would need (among other things) to identify the values of the other factors u_i that contribute to the d_i and this information is typically unknown. Note that, by contrast, to use the d_i for the quite different purpose of estimating the value of M, one does not need such information about the individual values of the u_i—instead, as explained in more detail below, it is enough if one knows very general facts about the distribution of the u_i. It is thus perfectly possible to be in the position of being able to use the data to estimate the melting point without being able to provide detailed explanations of the data or derivations of the data from independently supported premises.

Some commentators (e.g., Votsis 2009) have responded to these contentions by arguing that even if claims about data are not derivable from explanatory theories by themselves (or even if such derivations are not commonly provided in science), it nonetheless must always be possible "in principle" to derive claims about data from explanatory theories when conjoined with "suitable auxiliaries" including hypotheses about phenomena themselves, the operation of instruments, the distribution of various sources of error that are present and so on. My response to this, again developed in more detail below, is several fold. First, the contention that claims about data are derivable "in principle" from (some) true hypotheses is in itself completely trivial and unilluminating if one is interested in understanding data to phenomena reasoning as it actually figures in science. Suppose that phenomena P is in fact present on some occasion i in circumstances C and datum d_i occurs on this occasion. Then one can derive that datum d_i occurs from the premises (1) If P is present on occasion i in circumstances C, d_i occurs, (2)) P is present on occasion i in circumstances C. Relatedly, within a hypothetico-deductive (HD) framework of confirmation, one might attempt to take the observation of d_i to provide, in conjunction with (1), HD confirmation of (2). Needless to say, the in principle possibility of such a derivation has no tendency at all to show that scientists actually make use of such derivations or that they play some functionally useful role in scientific reasoning. In particular, because of the well-known defects of HD confirmation, the derivability of d_i from (1) and (2) or from other claims that include the occurrence of P, does not by itself warrant any inductive conclusions about P. Our claims about the relationship between data and phenomena were intended as claims about science and scientific theorizing as actually practiced (and what is required for data to provide evidence) and not about what is possible in principle, in a sense that is disconnected from what scientists do or should do.

What is most important from our perspective is that to establish that data are evidence for phenomena it is neither necessary nor sufficient as a normative matter to provide derivations/explanations of data. Indeed, focusing on the possibility of such derivations gets things backwards. What matters is not that we be able to infer "downward" to features of the data from theory and other assumptions; rather, what matters is whether we are able to infer "upward" from the data (and other assumptions) to the phenomenon of interest. Data are scientifically useful and interesting insofar as they provide information about features of phenomena. Thus Eddington infers *from* features of his photographs (and various background assumptions) *to* a value for the deflection of starlight rather than trying to infer or derive characteristics of the photographs from other assumptions. Similarly, the researcher measuring the melting point of lead infers from her data to an estimate of the melting point, rather than vice-versa.

The remainder of this paper is organized as follows. Section 2 describes some background motivations for the data/phenomena distinction. Section 3 qualifies some of the claims in our original papers. Sections 4–6 explore some issues about the role of background assumptions and "theory" in data to phenomena reasoning.

2 Background and motivation

My original thinking about the data/phenomena distinction was prompted in part by some issues concerning statistical explanation, and in fact these provide a natural illustration of some of the basic ideas of our 1988 paper. Consider a measurement of some observable (e.g., position) performed on a quantum mechanical system that is not in an eigenstate for the operator corresponding to that observable. With sufficient knowledge of the system, QM allows one to derive the probability with which various outcomes will occur, but not to derive (that is, deduce with certainty) which particular outcome will occur.

Although they differed in important respects, the standard models of statistical explanation developed in the 1960s and 1970s by Hempel, Salmon and others assumed that in this sort of case, if appropriate other conditions were satisfied, quantum mechanics explained why individual outcomes occurred rather than just the probabilities with which they occurred, even when those probabilities were strictly between zero and one.

My contrary (and at the time) heterodox view was that in such cases theories like QM explained only the probabilities with which outcomes occurred but not the occurrence of the individual outcomes themselves. However, it also seemed to me completely uncontroversial (and fully consistent with my heterodox view about statistical explanation) that observations of individual measurements outcomes on quantum mechanical systems when aggregated into claims about the relative frequencies with which such outcomes occurred could serve as evidence for or against the predictions of QM about probabilities. (Standard methodologies of statistical inference show how to use such frequency information to assess claims about probabilities.)

Suppose that one thinks of the outcomes o_i of individual measurements repeatedly made on some observable under the same conditions as data, claims about the probability with which those outcomes occur as phenomena, and QM as the theory which explains such phenomena. Then one has the essential ingredients of the data–phenomena–explanatory-theory picture defended by Bogen and Woodward. The data o_i are evidence for phenomena (probability claims) but only claims about phenomena and not claims about individual data outcomes are derivable from QM. Indeed the latter are not derivable even from QM in conjunction with theories about the operation of measuring instruments and so on since such outcomes are irreducibly stochastic. If one rejects the idea that statistical theories like QM explain individual outcomes when these have intermediate probabilities, then QM explains phenomena (in the form of claims about probabilities) but not the data that are evidence for those phenomena. (One may think of the data as indirectly evidence for QM since they are evidence for probability claims that are evidence for QM.) Even if one rejects these claims about explanation, the claims about the derivability of the o_i remain. Our 1988 paper claimed that a similar set of interrelations among data, phenomena and explanatory theory occurs in many areas of science.

A second motivation for our 1988 paper can be found in the body of ideas that came to be called the New Experimentalism and the associated slogan, due to Hacking (1983), that "Experiment has a life of its own". The central ideas of the new experimentalism are probably sufficiently familiar not to require detailed recounting But very briefly, I understand them to include (at least) the following:

> Experimental techniques, ideas about experimental design, design of instruments, procedures for analyzing experimental data and so on often seem to develop relatively independently of "high" theory (meaning, roughly, explanatory theory in our sense).
>
> Although experiments are sometimes explicitly conducted with the purpose of testing previously formulated theories, it is also common for experimentation to uncover new features of the natural world (phenomena) which were not predicted by any existing theory. Sometimes experiments are (legitimately) conducted in an exploratory fashion, in which researchers look for interesting effects or investigate new features of known phenomena, in the absence of any previously worked out theory that explains those phenomena.

One thought behind our 1988 paper was that similar points could be made regarding areas of scientific investigation that did not involve experimentation (understood as active manipulation of nature) but instead involved the generation of data by more passive forms of observation. Here too one found a variety of assumptions and techniques for data analysis that also seemed to have "a life of their own" in the sense that their development and the considerations that bore on their reliability often seemed relatively independent of high theory. These included many different sorts of statistical techniques for analyzing observational data and handling error found in it (with different techniques being appropriate for different sorts of data—e.g., for time series rather than cross-sectional data), various sort of "data mining" procedures, which may or may not be statistical in character, subject matter specific ideas about possible confounding factors that might affect observational studies and how to deal with these, ideas about how to measure or operationalize quantities of interest (not necessarily provided by theory), subject specific simplifying assumptions in data analysis (such as the spatial smoothing assumptions in fMRI analysis described below) and so on.

By distinguishing between data and phenomena, raising the question of how scientists reasoned from the former to the latter and taking seriously the possibility such reasoning often had its own distinctive characteristics (which might be different from the reasoning used to establish connections between phenomena and explanatory theory), we hoped to provide a framework for the exploration of the various sorts of considerations and techniques described above. My view is that forms of inference and discovery that involve reasoning with data and that are at least somewhat independent of prior detailed theoretical understanding have become even more prominent in science than they were when we wrote Bogen and Woodward (1988). In areas ranging from genomics to neurobiology to climatology, we have huge amounts of data and often limited theoretical understanding—the current level of development in these areas is, as the cliché, has it, "data rich and theory poor". The development of techniques for extracting information from data under such circumstances is thus a matter of considerable importance. Understanding the distinctive features of data to phenomena reasoning also should be high on the agenda of philosophers of science.

3 Corrections and restatements

Although much of “Saving “ seems to me to stand up fairly well, we also advanced claims that now strike me as exaggerated and insufficiently qualified and others that (with the benefit of hindsight) I wish we had expressed differently. For starters, we were far too willing to formulate our central contentions as claims about what “typically” or even always happens in “science”—thus, we said phenomena are “typically” or usually not observable, that theories don’t (ever?) explain data and so on. These formulations now strike me as overgeneralizations and (more importantly) unnecessary. When one considers the enormous variety of activities and disciplines that fall under the heading of “science” and the great variation these exhibit over time, it seems implausible that there will be many interesting exceptionless truths about how science always works. Formulating claims in terms of what is “typical” or “usual” rather than what always happens makes them vaguer and harder to assess empirically, but arguably does not much enhance their plausibility or usefulness. Moreover, such sweeping formulations now seem to me largely unnecessary for the points we were trying to make. As emphasized above, our goal was to provide a framework that made sense of various sorts of data-based reasoning, one that grants that such reasoning can be relatively independent of certain kinds of theory. In support of this framework, it is enough to show that reasoning from data to phenomena *can* (and not infrequently does) successfully proceed without reliance on theoretical explanation of data. It was an unnecessary diversion to claim that this was always or even usually the case. A similar point holds in connection with the claim in our original paper that phenomena, in contrast to data, are typically not observable (in senses of observable closely linked to human perception). As Paul Teller has argued[1] there are plausible examples of observable phenomena, at least if we do not restrict this notion in question-begging ways: for example, various phenomena associated with the operation of the visual system such as color constancy in changing light and pop-out effects in visual attention. Similarly, for some optical phenomena such as the so-called Poison spot which appears at the center of the shadow cast by a circular object illuminated by a point source. What we should have said (I now think) is that phenomena need not be observable and that in many cases standard discussions of the role of observation in science shed little light on how phenomena are detected or on the considerations that make data to phenomena reasoning reliable—asking whether phenomena are observable is often not the right question to ask if one wishes to understand how such reasoning works. This is because the reliability of such reasoning often has little to do with how human perception works.

Another, related point that should have been clearer in our original paper is that (contrary to what some commentators have supposed) our goal was not to replace “top-down” theory-dominated views of science with an equally monolithic view in which scientific reasoning is always understood as “bottom-up” or purely data-driven. Instead our goal was to advance a more pluralistic understanding of science, in which, depending on goals and contexts, sometimes data-driven reasoning *and* sometimes

[1] In a talk at PSA 2008, Pittsburgh, November, 2008.

theory played a more leading role. It goes without saying that different areas of science face very different sorts of problems and are in different epistemic situations and these problems may require very different strategies for their solution. There is no general answer to the questions like how data-driven or theory-driven scientific research should be—it all depends on what the problem is, what sorts of information is available and so on.

4 The role of background assumptions and goals in data to phenomena reasoning

I turn next to some general remarks about the structure of data to phenomena reasoning and about the role played by "background" assumptions as well as goals or interests in such reasoning. Some commentators have interpreted Bogen and me as claiming that researchers reason to conclusions about phenomena from data alone without the aid of any additional assumptions that go beyond the data—as claiming that such conclusions somehow emerge just from data and nothing else. Other commentators have taken us to be claiming that the researcher's goals or interests play no role in this process. This is absolutely not our (or at least my) view.

Data to phenomena reasoning, like inductive reasoning generally, is ampliative in the sense that the conclusion reached (a claim about phenomena) goes beyond or has additional content besides the evidence on which it is based (data). I believe it is characteristic of such reasoning that it *always* requires additional substantive empirical assumptions that go beyond the evidence. Thus the general form of an inductive inference is *not*

(4.1) Evidence-→ conclusion.

but rather *always* involves (at least) something like this:

(4.2) Evidence + substantive empirical assumptions → conclusion.

This is true of data to phenomena reasoning as well.

In fact, even schema (4.2) is incomplete. In many cases, inductive inference (including a great deal of data to phenomena reasoning) also relies on more or less explicit assumptions about epistemic *goals* or ends, including attitudes toward risk and costs associated with various mistakes.

I will return to this point below but for now want to further explore the role of empirical assumptions in (4.2). These may take a huge variety of different forms, which vary greatly from case to case. There is thus no one single grand assumption (e.g., nature is uniform) which serves as foundation for all inductive reasoning, but rather many different possible assumptions, the appropriateness of which will depend on the case at hand. Sometimes these assumptions will be highly specific. For example, when data consisting of the amount of carbon-14 in a fossil is used to arrive at a date for the fossil (as part of a reconstruction of, e.g., a migration pattern, or a phylogenetic tree, the phenomenon of interest), assumptions about the half-life of carbon-14, and the way in which soil conditions and atmospheric exposure may affect the presence of carbon in bones are among the background assumptions employed. In other cases, the relevant assumptions may be more generic, in the sense that they may be motivated by very general features of the situations to which they apply. For example, in statistical inference, assumptions about the probability distribution governing the error—whether it is normal, whether the error is correlated with other factors and so on—will often be crucial since the reliability of various estimating procedures will depend on such assumptions. Thus when measuring the melting point of lead, general empirical assumptions about the factors influencing the measurement results may be used to motivate the assumption of a particular error distribution. For example, the assumption that there are many factors influencing the measurement results, that these operate independently and that they have a combined effect that is additive may be used via the central limit theorem to motivate the assumption that the error distribution is normal. As a final example, consider the common use of spatial smoothing procedures in the analysis of fMRI data. Each individual voxel measurement is noisy; it is common to attempt to improve the signal to noise ratio by averaging each voxel with its neighbors, weighted by some function that falls with distance. This procedure depends (among other considerations) on the empirical assumption that the activity of each voxel is more closely correlated with nearby spatial neighbors.

In each of these cases the "additional" assumptions employed "go beyond the data" in the sense that they are not warranted just by the data generated in the particular experimental or observational context at hand, but rather have some independent source. These assumptions are conjoined with data (or assumed in some way in the data analysis) to reach conclusions about phenomena. However, that such assumptions "go beyond the data" does *not* mean they are arbitrary, empirically unfounded, untestable, or matters for stipulation or convention. Instead, assumptions like those described above are ordinary empirical assumptions, which are either true or false of the situations to which they are applied. When necessary, such assumptions often can be tested by ordinary empirical procedures. Such assumptions are sometimes "theoretical" in the sense that they concern factors that are not observed. (For example, the error distribution is not usually regarded as something that is "observed".) However, in many cases such assumptions will not be "theoretical" in the sense that they are part of a theory whose role is to provide substantive, detailed explanations. For example, one does not usually think of assumptions about the distribution of the error term as playing such an explanatory role—instead the error term is responsible for the "unexplained" variance in measurement. Arguably a similar point holds for the use of spatial smoothing assumptions in fMRI analysis—their role is not one of explanation.

I have stressed these points about the need for additional assumptions in data to phenomena reasoning because Bogen and I are sometimes interpreted as radical inductivists. Schindler (2007, p. 165) writes:

> Bogen and Woodward [believe] that unobservable phenomena are—without the mediation of the theory—inferred from observable data.

Elsewhere Schindler describes us as arguing for a "'bottom-up" construction of phenomena from data without the involvement of theory' (2007, p. 160).

If without the "mediation" or "involvement" of theory means "without any additional substantive empirical assumptions" I repeat that this is just not our view. Our view is that such assumptions are *always* required in data to phenomena reasoning

(I will also add, in anticipation of discussion below, that there are also cases in which data to phenomena reasoning is "mediated" by the very theory that explains those phenomena. What is excluded by the view defended in our 1988 paper is that this mediating role takes the form of theory providing *explanations* of the data.[2])

I said above that in addition to substantive empirical assumptions, inductive inference (including data to phenomena reasoning) often relies on (or is guided by) evaluative considerations having to do with the investigator's choice of goals, interests, and attitudes toward risk. Putting aside implausibly strong forms of realism about epistemic values, it seems natural to suppose that such choices can be reasonable or unreasonable, but that (unlike the empirical assumptions that figure in inductive inference) they are not straightforwardly true or false—they are not, as it were, forced on us by nature. (Nonetheless the reasonableness of such evaluative choices may depend in part on assumptions that are straightforwardly true or false: to take an example discussed below, whether it is reasonable to employ a certain estimator may depend on whether certain descriptive assumptions about the distribution of the error term hold.)

The goals and values figuring in data to phenomena reasoning can take many different forms, some of which are obvious and unremarkable, others of which may have a rich conceptual/mathematical structure.

(4.3) An investigator who wishes to determine the melting point of lead will have this as her goal or cognitive interest and this will structure her investigation in certain ways—she will proceed differently if instead she wishes to measure the boiling point of water. The investigator's project will be misguided if certain factual assumptions on which it rests are false (for example, if there is no such thing as "the" (unique) melting point of (all pure samples of lead). Nonetheless, the choice of this particular goal is not dictated by nature (or by the data) and the investigator might reasonably have chosen other goals instead.

(4.4) An investigator wishes to estimate the point value of some quantity from statistical data. A standard practice within frequent statistics is to choose an estimator satisfying certain evaluative criteria. For example, it is usually thought desirable that estimators be *unbiased,* where a random variable x^* is unbiased estimator for parameter X, if $E(x^*) = X$. Moreover, among unbiased estimators, it is standard to choose an (or the) estimator that is "best" in the sense of having minimum variance. If we require also that the estimator be a linear function of the data we arrive at the familiar notion of a best, linear unbiased estimator (BLUE). Again, the adoption of these criteria for what makes an estimator good is not a choice that is forced on us by nature. In some circumstances it might make sense to choose an estimator that is slightly biased but has smaller variance over an unbiased estimator of higher variance.

[2] We also do not require that the theories which figure in data to phenomena reasoning be independent of the theory which explains or predicts those phenomena—see Sect. 6.

5 Skepticism and foundationalism in data to phenomena reasoning

If additional assumptions are always required to get from data to phenomena (and such inferences may also be guided by goals and interests), this creates an opening for the skeptically minded. This is the dialectic exploited by James McAllister in a series of papers criticizing our views (e.g., McAllister 1997). McAllister observes that claims about phenomena do not follow just from the data alone and that something more is required for such inferences. According to McAllister, this "something more" has the character of a stipulation. In particular, according to McAllister, such reasoning involves the decomposition of data into a signal or pattern (corresponding to a phenomenon) and a "noise level", the latter being "stipulated", given the investigator's cognitive interests and attitudes toward risk An indefinitely large number of different stipulations and decompositions are legitimate. It follows that we must conceive of the world as "radically polymorphic"—any body of data contains or corresponds to indefinitely many phenomena.

For reasons that will emerge below, I do not find McAllister's notion of choice of a noise level very clear. If we interpret this as reflecting the investigator's interests, or attitudes toward mistake, then (as I have already said) I fully agree with the general idea that these play an important role in inductive inference, including data to phenomena reasoning. However, as have seen, substantive assumptions figure too. I believe McAllister is simply wrong in supposing that substantive assumptions like those described above (the assumption that carbon-14 has a certain half-life or the assumption that the error normally distributed in a certain measurement) are just matters of stipulation (of noise level or of anything else). Instead such assumptions are empirical claims that, given a particular context, are either true or false. And in the case of many possible phenomena claims (e.g., the claim that the melting point of lead is 627 K), the empirical assumptions required to license their reliable inference from data are false. True empirical assumptions will not license inconsistent phenomena claims from the same data, although different true assumptions may license different but consistent phenomena claims from the same data. We need to distinguish the role of goals and interests (which may have a stipulative element) from the role of empirical assumptions in data to phenomena reasoning. McAllister's notion of a "noise level" encourages a conflation of these two kinds of considerations.[3]

More generally, we can say that McAllister's focus on the need for additional assumptions/stipulations in data to phenomena reasoning draws attention to general features of *all* inductive inference. These features are inescapable; it is not a flaw in our account that it recognizes a role for them. Our account was never intended to answer the sort of skeptic who demands to be shown how conclusions about phenomena follow from data alone, without further substantive assumptions. In (Bogen and Woodward 1988), our interest was in describing scientific contexts in which (we assumed) not just data but a justifiable basis for other sorts of substantive assumptions that go beyond the data were available. We saw our task as identifying those assumptions, elucidating

[3] Noise level might refer to something like choice of significance level which arguably reflects a stipulation or convention but it might also reflect assumptions about the distribution of the error term which are not a matter of stipulation.

how they figured in data to phenomena inference, and (in some cases) showing how they were justified, but not one of providing a presuppositionless foundation for them.

Once it is recognized that goals and interests play a role in guiding data to phenomena reasoning (and inductive inference more generally), an important issue is where they (and whatever "relativity" or "subjectivity" they introduce) should be "located". One possibility, which I take to be the alternative favored by McAllister, is to locate these "out in the world", so to speak, or at least the way we think about the world—that is, in the phenomena themselves, or in the way in which we conceptualize phenomena. In other words, one takes the fact that goals and interests affect which data to phenomena inferences investigators make to show the phenomena themselves (or the truth-value of claims about them) are in some way relative to the interests, goals or perspective of the investigator. This line of thought encourages the idea that reality itself should be regarded as radically polymorphic, containing different phenomena corresponding to each of the goals and stipulations the experimenter is willing to adopt.

An alternative view, which I find more plausible, is take the relativity to goals and interests that figures in data to phenomena reasoning to attach to *investigators* and their choices rather than to the phenomena themselves. On this alternative view, someone who decides to measure the melting point of lead rather than the boiling point of water is influenced by her interests, but this does not mean that the value of the melting point of lead is an "interest-relative" matter or that given one set of cognitive goals or stipulations about an appropriate error level, the melting point might legitimately be regarded as, say, 327°C and given another set of interests and goals it might be legitimately regarded as 627°C. Similarly, in an investigation of the bias of a coin in which the data are repeated coin flips, a researcher who employs a significance test with a significance level of 0.05 has a different attitude toward the costs of a certain kind of mistake (she adopts a different noise or error level) than a researcher who employs a significance level of 0.001. This might lead the second researcher to reject the hypothesis that the coin is fair in circumstances in which the first researcher does not. But this does not mean (at least on the usual interpretation of significance testing) that whether the coin is biased itself depends on the researcher's attitude toward error or that there is no such thing as the true bias of the coin, independently of the researcher's interests. It just means that the researchers have made different choices about (or have different attitudes toward) the possibility of making a certain kind of mistake about the true (interest -independent) bias of the coin.

6 The "involvement" of explanatory theory in data to phenomena reasoning

I suggested above that explanatory theories often aim at providing explanations of phenomena and that it is difficult to provide systematic explanations of data from general theory, even in conjunction with theories of instruments, non-trivial auxiliaries etc. Moreover and more importantly, for data to provide reliable evidence for phenomena it is not necessary to provide such explanations. I want now to expand on these claims. I begin by emphasizing what they do *not* say. Suppose that we have a theory T that explains some claim about phenomenon P and P is established by reasoning from data D. Some commentators have taken Bogen and Woodward, 1988, to be contending that theory T cannot in any way be "involved" in the reasoning from D to P. We do not make this claim and I think it is obviously false. Instead what we claim is T that commonly does not play (and need not play) one particular kind of role in reasoning from D to P—that of providing systematic explanations of D. This is not to say that T plays no role at all in reasoning from D to P.

This is not mere hair-splitting; a major problem with discussions of the role of "theory" in data to phenomena reasoning (or, for that matter, in connection with the so-called theory-ladeness of observation) is that philosophers fail to make discriminations among the very different ways that theories can be "involved" in this process. They also fail to distinguish among very different things that go under the heading of "theory". Different "involvements" by different sorts of theories can have quite different epistemological implications for data to phenomena reasoning. We need to keep these distinct.

To begin with a very simple case, suppose theory T explains some phenomenon P that a researcher wishes to measure/detect from data D. Suppose also T provides a *motivation* for measuring P: the researcher measures P because she wants to test T or because T says P is an important quantity that plays a fundamental role in the nature—the researcher would not have attempted to measure P if he did not regard T as a serious possibility. Perhaps also T provides a vocabulary for characterizing the results of measuring P. In this case, there will be an obvious sense in which T is "involved" in reasoning from D to P—the researcher would not have engaged in this reasoning at all or would not have used concepts drawn from T to describe its results of this reasoning if she did not accept T or at least take it seriously. However this sort of involvement of T in data to phenomena reasoning does not necessarily mean that T is being used to explain D or that D cannot be evidence for P unless T is conceived as playing this explanatory role. For example, it seems plausible that Eddington would not have performed the measurements on his solar eclipse expedition and would not have described the results of these measurements in the way that he did (that is, as measurements of the deflection of light by the sun's gravitational field) if he did not accept various "theoretical" claims—most obviously, that gravity influences the behavior of light. In an obvious sense these theoretical assumptions help to structure Eddington's measurement activities—they play both a motivating role (in suggesting a goal or target to aim at) and a vocabulary for describing the results of the measurement. This does not mean, however, that GR or the other explanatory theories about gravity and its coupling with light under test played the role of explaining Eddington's data. Still less does it follow that Eddington's data analysis required that he assume that one of these theories played this explanatory role or even that one of them was correct.

I said above that different "involvements" by different sorts of theories in data to phenomena reasoning can have very different epistemological implications. One particular philosophical worry is that the involvement of theory in such reasoning is of such a character as to introduce a "vicious" circularity of some kind, with the involvement of T in the data analysis somehow guaranteeing that the results of such analysis will be to support T regardless of what the data are, or at least making it impossible to detect that T is false when it is. It is thus important to note that it is entirely possible for T to be involved in some way in data to phenomena reasoning without introducing this sort of vicious circularity. For example, in Millikan's oil drop experiment, the mere fact that theoretical assumptions (e.g., that the charge of the electron is quantized and that all electrons have the same charge) play a role in motivating his measurements or a vocabulary for describing his results does not by itself show that his experimental design and data analysis were of such a character as to guarantee that he would obtain results supporting his theoretical assumptions. His experiment was such that he might well have obtained results showing that the charge of the electron was not quantized or that there was no single stable value for this quantity. A similar point holds for Eddington's measurements

In general, determining whether T is involved in D to P reasoning (when P is some phenomenon which is used to test T) in such a way as to introduce damaging circularity is a far from straightforward—it is not obvious this can be captured just in terms of logical relationships involving T, P and D.[4] To illustrate this point, consider a schematic example: Theory T is taken seriously but is not known to be true, T predicts (and if true would explain) phenomenon P, and data D provides a prima-facie case that P occurs. (T might be the electroweak theory, P the existence of neutral currents, and D photographs produced by the Gargamelle bubble chamber.) Suppose, however, we are at the limits of available experimental technology: data D are very noisy, we believe there may be unknown sources of error in the experiment, and background assumptions used in analyzing the data are based on simulations in which we are not completely confident. Based just on reasoning from the data and the assumptions adopted in the data analysis (and ignoring the fact that T predicts P), we are far from fully confident that we have detected P. In such a case, the very fact that T predicts P might boost our confidence that the error has been adequately controlled for, the simulations are not radically mistaken and that the experiment has successfully detected P. In other words, the data-based reasoning to P and the fact that T predicts P mutually reinforce our confidence that P is real—the two sets of considerations are in a positive feedback relation with each other. That P apparently obtains might in turn to be important evidence in favor of T. This is a case in which T figures in the overall process leading to acceptance of P in an important way and P in turn supports T. Note, however, that even in this case (i) T does not (or at least need not) play the role of explaining individual items of data and (ii) it is far from obvious that the process described is automatically viciously circular or that it fails to provide a legitimate basis for increased confidence in P (or T, for that matter). Of course there would be no basis for increased confidence in P if (to mention an extreme possibility) T influences the researchers' reasoning in such a way that they would accept P completely independently of whatever data was produced in the experiment—i.e., if their commitment to T is such that it would lead them to interpret virtually any data

[4] Another illustration: in order to test the hypothesis that metals expand linearly with temperature a thermometer is employed which is calibrated on the assumption that the metal it contains (e.g., mercury) increases linearly with temperature. This may seem blatantly circular, but a little thought will show that there are possible results that might falsify the hypothesis—for example, if mercury and the metal tested expand in accord with different functions of temperature, one of which is non-linear, this discrepancy could be detected. See Chang (2004). Two questions: (i) Is this procedure objectionably circular? (ii) If it is not, how should we characterize the kinds of circularity that are objectionable?

as evidence for P (dismissing discrepant data as due to error etc.). However, the story I have told does not require that this be the case.

This last example is a case in which top-down theory-driven reasoning plays an important role in phenomenon-detection and experimental reasoning—it helps to convince researchers that error has been adequately controlled and that the phenomenon is real. Our claims about the data/phenomena distinction should not be interpreted as requiring that the kind of top down reasoning described above never occurs. We do claim, however, that not all data to phenomena reasoning is theory-driven in the way just described. The history of science is full of examples in which phenomena are detected or noticed in observations or exploratory experiments without the investigator being in possession of any prior theory that explains or predicts that phenomenon or, indeed, provides any reason to expect it.[5]

Acknowledgments Many thanks to the participants at the Heidelberg conference on "Data, Phenomena, Theories" for helpful comments.

References

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review, 97*(3), 303–352.

Bogen, J., & Woodward, J. (1992). Observations, theories and the evolution of the human spirit. *Philosophy of Science, 59*, 590–611.

Bogen, J., & Woodward, J. (2005). Evading the IRS. In M. Jones & N. Carwright (Eds.), *Correcting the model: Idealization and abstraction in science. Poznan studies* (pp. 233–267). Holland: Rodopi Publishers.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press.

Glymour, B. (2000). Data and phenomena: A distinction reconsidered. *Erkenntnis, 52*, 29–37.

Hacking, I. (1983). Representing and intervening. Cambridge: Cambridge University Press.

Kaiser, M. (1991). From rocks to graphs—The shaping of phenomena. *Synthese, 89*, 111–133.

McAllister, J. (1997). Phenomena and patterns in data sets. *Erkenntnis, 47*, 217–228.

Schindler, S. (2007). Rehabilitating theory: Refusal of the 'bottom-up' construction of scientific phenomena. *Studies in the History and Philosophy of Science, 38*, 160–184.

Votsis, I. (2009). Making contact with observations. PhilSci Archive. Proceedings of the first conference of the European Philosophy of Science Association. Accessed January 20, 2009.

Woodward, J. (1989). Data and phenomena. *Synthese, 79*, 393–472.

Woodward, J. (2000). Data, phenomena, and reliability. *Philosophy of Science, PSA 1998, 67*, S163–S179.

[5] A number of examples are provided in Hacking (1983)—see especially pp. 152–154.