Synthese (2011) 182:89–100
DOI 10.1007/s11229-009-9614-9

Data meet theory: up close and inferentially personal

Ioannis Votsis

Received: 20 May 2009 / Accepted: 3 June 2009 / Published online: 8 July 2009

Abstract In a recent paper James Bogen and James Woodward denounce a set of views on confirmation that they collectively brand 'IRS'. The supporters of these views cast confirmation in terms of Inferential Relations between observational and theoretical Sentences. Against IRS accounts of confirmation, Bogen and Woodward unveil two main objections: (a) inferential relations are not necessary to model confirmation relations since many data are neither in sentential form nor can they be put in such a form and (b) inferential relations are not sufficient to model confirmation relations because the former cannot capture evidentially relevant factors about the detection processes and instruments that generate the data. In this paper I have a two-fold aim: (i) to show that Bogen and Woodward fail to provide compelling grounds for the rejection of IRS models and (ii) to highlight some of the models' neglected merits.

Keywords Data · Phenomena · Confirmation · Inferences · Bogen · Woodward

1 Introduction

James Bogen and James Woodward's much discussed work on the relationship between data, phenomena and theories denounces, among other things, positivist-inspired views of confirmation (Bogen & Woodward 1988, 1992, 2003; Woodward 1989). In their most recent article, they classify such views under the general heading 'IRS', for IRS supporters cast confirmation in terms of Inferential Relations between observational and theoretical Sentences. According to Bogen and Woodward the most paradigmatic of such views are the hypothetico-deductive and the positive instance models

I. Votsis (✉)
Philosophisches Institut, Universität Düsseldorf, Universitätsstr. 1, Geb. 23.21/04.86,
40225 Düsseldorf, Germany
e-mail: votsis@phil-fak.uni-duesseldorf.de

of confirmation. Countering IRS models, Bogen and Woodward launch two chief objections: (a) inferential relations are not necessary to model confirmation relations since many data are neither in sentential form nor can they be put in such a form and (b) inferential relations are not sufficient to model confirmation relations because the former cannot capture evidentially relevant factors about the detection processes and instruments that generate the data.

In this paper I have a two-fold aim: (i) to show that Bogen and Woodward fail to provide compelling grounds for the rejection of IRS models and (ii) to highlight some of the models' neglected merits.

2 The IRS model and its presumed failures

In their now classic 1988 paper, Bogen and Woodward draw a distinction between *data* and *phenomena*. Data are roughly what is observed or detected with or without the use of instruments. Examples include drawings made by surgeons, temperature readings and bubble chamber photographs. Phenomena are the objects, events and processes (in the world) under investigation. Examples include sensory processing dysfunction, the melting point of mercury and particle interactions. In Bogen and Woodward's view we use theories to systematically explain, infer or predict phenomena; we typically cannot use them to explain, infer or predict data. This view is a reaction to what they conceive of as the positivist-inspired view that "[e]xplanation, prediction, and theory-testing involve the deduction [or more generally the inference] of observation sentences from other sentences, some of which may be formulated in a 'theoretical' vocabulary..." (p. 303).

While in their 1988 paper, Bogen and Woodward focus on the first of these dimensions, i.e. explanation, in a recent paper they turn their attention to the theory-testing dimension, branding their nemesis 'IRS'. According to IRS, "the epistemic bearing of observational evidence on a scientific theory is best understood in terms of Inferential Relations between Sentences which represent the evidence and sentences which represent hypotheses belonging to the theory" (2003, p. 223). As already explained in the introduction, the most well-known versions of IRS are the hypothetico-deductive account and the positive instance account of confirmation. The latter is understood broadly so at to include a number of inductive models including bootstrapping accounts (see, for example, Glymour 1980) and several Bayesian accounts (see, for example, Howson & Urbach 1989).[1] Against such views, Bogen and Woodward emphasise their conviction that the character of the evidential relation between data, phenomena and theories is not one best cast in inferential terms. In their eyes, "... the epistemic import of observational evidence is to be understood in terms of empirical facts about particular causal connections and about the error characteristics of detection processes. These [facts] are neither constituted by nor greatly illuminated by considering the formal relations between sentential structures which IRS models focus on" (2003, p. 223).

[1] Bogen and Woodward claim that in principle Bayesian accounts can solve some of the difficulties IRS models face but in practice most fall prey to them (2003, f43, p. 255).

Bogen and Woodward trace the motivation for the first IRS models to philosophers like Hempel and Reichenbach, who wanted an objective theory of confirmation, one that avoids the conflation of objective, rational and logical factors with the undesired subjective, psychological ones. They cite, for example, Hempel who seems to want to de-subjectivise and generalise confirmation: "it seems reasonable to require that the criteria of empirical confirmation, besides being objective in character, should contain no reference to the specific subject matter of the hypothesis or of the evidence in question" (1965, pp. 9–10). It is essential to note that, as Bogen and Woodward concede, the IRS project is supposed to be a rational reconstruction of the confirmational practice of science. It is therefore no objection to IRS theorists to argue that their modelling does not mirror actual scientific reasoning. Having said this, IRS theorists can be held accountable if they fail to reconstruct all the evidentially significant factors that make up theory testing.

That the formal tools of IRS models are not able to capture the evidential relations between data and theories is a case that Bogen and Woodward parcel into two objections. First, it is claimed that inferential relations are not necessary for confirmation relations because many data are neither in sentential form nor can they be put in such a form (pp. 232–236). Second, it is claimed that inferential relations are not sufficient for confirmation relations because: (a) they are unable to account for various actual and hypothetical cases including well-known paradoxes of confirmation (pp. 229–231) and (b) they are unable to account for the processes that generate the data and on which standards of reliability depend (pp. 249–255). In the sections that follow I raise doubts on the gloomy picture Bogen and Woodward paint on the prospects of encapsulating confirmation relations in an inferential manner. Moreover, whenever appropriate I present positive reasons for the benefits of adopting a general IRS framework, though I refrain from endorsing any particular manifestation of it.

3 Getting the sentence out of the data

First up, let us consider the objection that inferential relations are not necessary for confirmation relations because several data are non-sentential in form. Such data are manifested in different ways, among them "photographs, drawings, tables of numbers, etc." (p. 226). To substantiate their objection, Bogen and Woodward throw the spotlight on the data gathered by the Eddington expedition in 1919 and used to confirm the general theory of relativity. The data were photographs and required a certain amount of processing, e.g. ascertaining scale, before they were used to calculate the phenomenon at issue, i.e. starlight deflection by the sun's gravitational pull. The result was then compared against the predictions of three competing hypotheses: (1) the General Theory of Relativity (GTR), (2) Solder's Newtonian model and (3) the No Deflection Hypothesis. The detected starlight deflection though not entirely in coherence with the predictions of the GTR was much more in line with those predictions than with the ones made by Solder's model. The upshot, as we all know, was a significant conformational boost that helped establish GTR in the scientific world.

On its own the fact that many data are prima facie non-sentential does not suffice to trouble the IRS-theorist. Such a theorist, recall, wears a reconstructionist hat and hence need only demand that prima facie non-sentential data can be packaged in sentential form. What is really at stake then is whether such data are *irreducibly* non-sentential. To subdue the IRS-theorist, Bogen and Woodward need to argue that at least some of the prima facie non-sentential data—and perhaps even to argue that the vast majority of them—cannot be reduced to sentential form. Though they do not explicitly adopt such a stance, at least one of their arguments seems to target reducibility. To be exact, they argue that IRS "provides no guidance for the construction of the required observation sentences" (p. 235).

Need IRS provide such guidance? It would be a boon, of course, but as Bogen and Woodward regularly point out—see Sect. 5 below—the data are often idiosyncratic in a way that prevents IRS theorists from formulating a general rule or even a rule of thumb that converts non-sentential data into sentential ones. In my view, the most likely scenario is that there is not only one rule but a great number of them, each corresponding to different types of data and conditions under which they are useful, e.g. a rule for converting photographic data when scale is not important, a rule for converting sounds coming from a Geiger counter, etc. A powerful justification for this view comes from the fact that we have learned to automate such tasks with computers. The mere digitisation of data makes them ripe for inferential manipulation—in other words just what the IRS doctor ordered. The subsequent algorithmic analyses of the data to detect phenomena illustrates that even this stage of the testing of a hypothesis is an inferential matter. We can even program the computer to output its results in natural language form, e.g. 'Phenomenon A was detected at time t_1 under conditions ϕ, Ψ, and ζ'. Take the data in the seti@home experiment based at the University of California Berkley. The experiment attempts to detect extra-terrestrial broadcasts (the phenomenon under study) by analysing narrow-bandwidth signals (the data) gathered from the Arecibo radio telescope. The data are recorded, digitised and broken into small manageable chunks which are then sent around the globe to home users who have downloaded a program that analyzes the data for extra-terrestrial signals. All of this is automated, to the extent that a scientist's computer could output each step in the process in argument form.

For the time being to counter the irreducibility claim it will thus suffice to tackle Bogen and Woodward's counterexamples on a case by case basis. Take the GTR example. The evidentially relevant part of the eclipse photographs consists in the distances between the images of the stars. These represent the apparent distances between the stars under starlight deflection. Foil photographs were also taken at night-time to record the 'actual' distances of the stars, i.e. the distances without starlight deflection from the sun. In sentential form these data will look something like this: 'The image of star A is n microns distant from the image of star B'. Similar reconstructions of the data can be given to Bogen and Woodward's favourite cases, namely that of the calculation of the melting point of lead and that of the detection of weak neutral currents. The data collected from measurements of the melting point of lead can be sententially dressed as follows: 'Sample x_1 measured by thermometer l_1 under experimental conditions c_1 at time t_1 registered value y_1'. The data from the Gargamelle heavy-liquid bubble chamber at CERN were also photographs. As with all particle accelerator photographs the telltale signs of particles and of particle interactions are streaks that satisfy certain geometrical properties. In sentential form these could look something like this: ‘Photograph x_1 taken at the Gargamelle bubble chamber at time t_1 under experimental conditions c_1 exhibits geometrical properties y_1’.

Bogen and Woodward will surely complain that there is much else of evidential relevance in these cases. That’s absolutely right. There are calculation techniques, correction factors, signal threshold assumptions, etc. Nothing prevents us from stating these in sentential form. Take the GTR case again. One of the central evidentially relevant assumptions employed was a correspondence relation between features of the photographs and features we want to attribute to the stars. The correspondence can be stated in sentential form and will look something like this: ‘Distances in microns between images of the stars correspond to distances in astronomical units (AU) between the stars’. Another central assumption that was evidentially relevant was a correction for parallax effects between the eclipse and comparison photographs. In sentence form this assumption will not be very different from the following: ‘For eclipse photographs taken at (long.-lat.) coordinates λ_1, ϕ_1 at time interval $t_2 - t_1$, the distance measured between images of stars A and B in microns n needs to be corrected by factor g before they can be compared to the night-time photographs taken at (long.-lat.) coordinates λ_2, ϕ_2 at the same time interval’. The irony of it all is that sometimes Bogen and Woodward unwittingly offer their own sentential formulations of evidentially relevant considerations. For example, speaking of the scale complications between the eclipse and the comparison photographs they state that this was solved by a “correspondence of radial distances between stars shown on an accurate star map to linear distances between star images on the photographs” (p. 232). A similar treatment can be given to the other high-profile cases Bogen and Woodward like to discuss in their work. Indeed, I provide examples of sentential formulations of evidentially relevant factors from the melting point of lead case (Votsis, forthcoming).

Without a doubt these reconstructions are not easy to perform. Life does get a bit easier for the IRS theorist, however, in that evidentially relevant factors need not be reconstructed in detail. As Bogen and Woodward themselves note some factors are either not explicit or not systematically understood (p. 244). In such cases all one needs to do is encapsulate whatever little information we have about the factor in sentential form. For example, not knowing the exact mechanism behind a reliable instrument does not forbid us from asserting (justifiably or unjustifiably) that the features in its outputs correspond to features in the objects that we target with it. Bogen and Woodward’s discussion of Galileo’s telescope fits like a glove here as they point out that the telescope was reliable “even though Galileo lacked an optical theory that explained its workings” (p. 240). At this point one may well wonder why we would trouble ourselves with a reconstruction of evidentially relevant factors of which we have little to no knowledge about. The answer is the same as that given for factors we do have knowledge about, namely that by reconstructing them we can more easily examine their inferential role and ultimately hold them accountable.

A final remark should secure enough nails on the coffin of the irreducibility objection to seal it tight. If a set of data is to play as weighty a role as confirming or disconfirming a phenomenon (and by Bogen and Woodward’s lights only indirectly a theory) then it had better be something that we can express assent or dissent towards.

This I take to be a basic requirement of scientific discourse about data. But to express assent towards a set of data D_1 or dissent towards a set of data D_2, both with respect to a phenomenon P, just means to believe in the first but not in the second. Since beliefs, at least for most philosophers, are propositional attitudes, it is reasonable to assume that the content of the foregoing beliefs are propositions. But that's precisely what sentences express.[2] Hence to express assent or dissent to a set of data is to implicitly (and perhaps even contrary to one's wishes) endorse the view that what is evidentially of essence in that set can be put in sentential form.

To sum the section up, Bogen and Woodward's qualms are powerless to faze the IRS theorist. All that matters according to that theorist is whether we can (re)construct sentences from prima facie non-sentential data without loss of evidential content. I am sympathetic to this view and furthermore challenge Bogen and Woodward to display *one* datum whose evidential worth cannot be sententially packaged in any of the aforementioned ways.

4 Restoring sufficiency. Part I—A paradox and a thought experiment

Let us now turn to the objection that IRS models are insufficient to capture all evidential relations. This objection amounts to the idea that there are cases where the theory and the data have the right inferential relation but no corresponding evidential relation exists. In the next section I examine how Bogen and Woodward's views about standards of reliability threaten the sufficiency of IRS models. In this section I concentrate on two of their counterexamples to the sufficiency of IRS models that are discussed independently of reliability considerations. One is the well-known raven paradox. A white shoe does not seem to be evidence for the hypothesis that 'All ravens are black'. Yet, under the IRS model of confirmation it is evidence because it stands in the right inferential relations to that hypothesis. The other example is a thought experiment which involves what I brand 'doppelganger data'. Imagine a case where a set of data D_1 generated by some unknown process is identical to a set of data D_2 generated at Gargamelle. Suppose further that D_2 positively identifies a weak neutral current event. D_1 does not seem to be evidence for weak neutral currents although it stands in the right inferential relations to them. What lands the IRS models into hot water, according to Bogen and Woodward's diagnosis, is that they neglect to take into account the actual processes that generate the data (p. 250).

Confronting the IRS theorist are two options. First, such a theorist can bite the bullet and argue that data standing in such inferential relations do confer some support. This option is harder to sell in the case of the doppelganger data but it is a rather uncontroversial one in the debate over the significance of the raven paradox. Thus it has been argued that white shoes provide some support for the raven hypothesis even though much less than the support one gets from finding black ravens. The degree of support

[2] We can even argue in the following way. Assent or dissent towards a set of data expresses a belief in the truth or falsity of the data respectively. But truth and falsity are properties of sentences. More subtly, since in most cases data encode information that is to some extent degraded, assent towards D_1 amounts to, among other things, the belief in the truth of a counterfactually close data set D_1' which is devoid of such imperfections.

depends on the relative size of the two classes, in this case that of non-black things and that of black ravens. Perhaps more subjectively, the degree of support depends on whether such objects were randomly picked. There is no reason why these ideas cannot be encoded as auxiliaries or even as specific rules of inference. Bayesians, for example, model the different degrees of support—via degrees of belief—afforded by each datum by assigning different values to the expectedness of the evidence.

The second option available to the IRS theorist is to accept (as Bogen and Woodward anyhow insist) that such data are not evidentially relevant but at the same time to employ inferential filters to dismiss them. That way the data no longer stand in the right inferential relations to the theory and hence no insufficiency objection can be launched. This option is open to both the doppelganger data and the raven paradox data. In the raven case, we can use rules to block Bogen and Woodward's assumption that "evidence which confirms a claim also confirms claims which are logically equivalent to it" (p. 230). In cases of doppelganger data, we can use rules to block inferential relations that do not instantiate the right causal relations in much the same way as causal-reliabilist theories in epistemology block knowledge when an agent's belief in a certain proposition p is not causally connected with one or more fact(s) about p. I can conceive of two ways of implementing the second option. The first involves the postulation of rules that can be used as steps in a non-monotonic argument to prohibit a certain class of inferences. Such rules are therefore internal to the argument. The second involves the postulation of rules that discriminate between relevant and irrelevant data prior to the construction of the particular inferential model. These rules are external.

A potential objection to the second option calls into question the augmentation of IRS models with additional rules, especially those that have semantic content. To do so, it might be argued, would be to go beyond the original conception of the model as a purely (first-order) syntactical account of confirmation and hence to defend a viewpoint that was never in contention. Although reasonable sounding at first, this objection caricatures the truth on the ground. IRS models, even those touted by logical empiricists, were hardly ever purely syntactical.[3] Take the deductive-nomological model which served both as an account of explanation as well as one of confirmation.[4] Among other things, the model demands the satisfaction of at least one extra-syntactical condition, namely the premises of a D-N argument all have empirical content.[5] In similar fashion, we may demand from other instantiations of IRS models to fulfil any number of additional syntactical or extra-syntactical conditions without forsaking their IRS pedigree.

[3] Cf. Hempel (1943).

[4] Although the deductive-nomological model was primarily meant as a model of explanation, we must not forget that for the logical empiricists explanation and prediction were intimately (and even symmetrically) tied. For that reason the model was employed in a confirmational capacity. It is worth noting that more recent models of explanation, e.g. Kitcher's unificationist account, continue to enjoy proximal relations with confirmation.

[5] Another potentially extra-syntactical condition is the requirement that at least one law of nature be included in the premises of a D-N argument. Its qualification as extra-syntactical depends on whether laws of nature are understood in a purely syntactical manner.

5 Restoring sufficiency. Part II—A question of reliability

The main thrust of Bogen and Woodward's objection that IRS models are insufficient to capture all evidential relations comes not from the examples discussed in the previous section but from their discussion of reliability. Two notions of reliability are at stake here, general and local. An instrument or detection process is *generally reliable* with respect to a set of data "if it has a satisfactorily high probability of outputting, under repeated use, correct discriminations… of competing phenomenon-claims and a satisfactorily low probability of outputting incorrect discriminations" (p. 237). Galileo's telescope is deemed generally reliable because it presumably possesses 'repeatable error characteristics'.[6] Sometimes general reliability is insufficient to establish the evidential import of data. It is then that we turn to *local reliability*. This involves "single-case performances of procedures, pieces of equipment, etc" (p. 226). The relevant phenomenon is thus established without recourse to repeatability but simply "by ruling out other possible causes of the data" (p. 245). So, for example, "...it might be that this particular fossil is contaminated in a way that gives us mistaken data, or that the equipment I am using has misfunctioned [sic] on this particular occasion" (p. 244). The data resulting from the radioactive dating of this fossil might still be locally reliable, provided we can rule out the contamination or the malfunctioning of the instrument as plausible causes.

Bogen and Woodward contend that the IRS cannot model evidentially relevant factors concerning general or local reliability. They buttress their contention by citing two closely related reasons. The first one latches onto the idea that "assessing the general [or local] reliability of an instrument or detection technique does not require that one possess a general theory that systematically explains the operation of the instrument or technique or why it is generally reliable" (p. 240)—see also (1988, pp. 309–310, 312, 317). What is implied here is that without such a general theory the IRS cannot reconstruct the kinds of evidential relations that hold between instruments/detection techniques, data and theories. The second reason concerns the "idiosyncratic" and "highly heterogeneous" nature of the information utilised in making judgments about reliability (2003, pp. 244–245).[7] This information includes such things as the "the performance of the detection process in other situations in which it is known what results to expect, about the results of manipulating or interfering with the detection process in various ways, and so forth" (p. 253). The punch-line, according always to Bogen and Woodward, is that it is unclear how to model such information via the IRS.

A no-frills response to the first reason is that the IRS theorist need not possess a general theory that systematically explains the reliability or the operation of an instrument or technique. So long as we have some evidential basis for our belief in the reliability of an instrument or technique, I do not see why a relevant auxiliary cannot be reconstructed. Such an auxiliary, if you recall from Sect. 2, may encapsulate whatever little

[6] I say 'presumably' because their account of the telescope's general reliability is thin on the details.

[7] Bogen and Woodward seem to intend the two reasons to apply to both general and local reliability. For example, after discussing how the two reasons are pertinent to general reliability, they say "[s]imilar remarks apply to conclusions about local reliability" (f37, pp. 239–240).

information we have about a given factor and still prove helpful. Consider the case of Galileo's telescope again. The general reliability of the telescope becomes apparent when one realises that we have independent confirmation for a subset of the data it generates, e.g. features of the moon we can see with the naked eye. But this is surely easy to encode into an auxiliary. In the given case the auxiliary will look something like this: 'Galileo's telescope is generally reliable with respect to domains of phenomena P_1 and P_2 where it produces data sets D_1 and D_2 respectively because we have independent confirmation (i.e. several consistent naked eye accounts) of the validity of D_1'. The benefit of formulating such auxiliaries is that they facilitate the transparent evaluation of scientific presuppositions. In short, they allow us to throw light on the defeasibility conditions of such presuppositions. For example, if we have various independent reasons to believe that Galileo's telescope is not reliable in its production of a given proper subset of data D_2 then we shall have to modify the auxiliary accordingly as well as any of our epistemic commitments that are based on it.

A similar response can be given to Bogen and Woodward's second reason. The specificity and heterogeneity of the relevant information are irrelevant to the question whether we can construct suitable auxiliaries. Initial conditions are both specific and heterogeneous yet they are and have been the staple of IRS models without any objections to the contrary. Why should information about the local reliability of an instrument or a technique be any different? To satisfactorily answer this question we need to delve deep into the sole example Bogen and Woodward provide considerable detail for, namely the attempts to detect gravitational waves in the late 1960s and early 1970s. The key figure in this story is Joseph Weber who in 1969 unilaterally claimed their discovery. The measuring instrument in Weber's experiment was a metal bar which was supposed to vibrate under the influence of gravitational waves. Since other sources could cause such vibrations—e.g. seismic, electromagnetic or thermal activity—the single most important consideration in designing the experiment was to shield against, or at least to correct for, confounding factors. That's precisely what Weber set out to do. In so doing, Bogen and Woodward tell us, Weber was building up a case for the local reliability of his instruments and methods.

> The problem [Weber] faced was that a number of other possible causes or factors besides gravitational radiation might in principle have caused his data. Unless Weber could rule out, or render implausible or unlikely, the possibility that these other factors might have caused the disturbances, he would not be justified in concluding that the disturbances are due to the presence of gravitational radiation (p. 247).

In the years that followed consensus was formed that Weber did not in fact positively identify gravitational waves. At issue were the very methods he employed to establish local reliability. Weber's critics, according always to Bogen and Woodward, were able to challenge his detection claim on grounds that were idiosyncratic and highly heterogeneous.

Two questions emerge at this point. What were these idiosyncratic and highly heterogeneous evidential considerations upon which Weber's detection claim was challenged? Are IRS models incapable of doing such evidential considerations justice? Bogen and Woodward list three evidential considerations that were crucial to the dismissal of Weber's detection claim. Only two of them can be classified as idiosyncratic and highly heterogeneous.[8] They are: (i) one of Weber's detection claims was made on the basis of an erroneous assumption about the time another team's data were recorded—the correction of this assumption annulled the detection claim and (ii) the detections produced by Weber's own data had the desirable feature of sidereal correlation but this feature could not subsequently be reproduced either by him or by any of the other teams.

Let us look at the first problem. Even though Bogen and Woodward mention this problem in passing, we can take a more sustained look informed by Franklin's (1994) excellent historico-philosophical paper on this event.[9] Franklin digs deep into the published records (articles and memoirs) and unearths the following details. A handful of teams other than Weber's run their own detection experiments, collected data and analysed them. The data were then passed around the various groups for closer scrutiny. At one point, Weber professed a detection correlation between a segment of his own team's data and that of data gathered by a team led by David Douglass. According to Weber, the time stamps of the two sets of data had a 1.2 s discrepancy but that could easily be explained away by the fact that the clocks used at each of the different experiment locations were not in perfect synchrony. As a matter of fact the time discrepancy was 4 h, 1.2 s. Weber had erroneously not taken into account that the clocks in the two experiment locations were set to different time zones (one used GMT and the other EDT). Douglass confronted Weber at a General Relativity conference in Cambridge (CCR-5) and Weber conceded the mistake. What is relevant to our discussion here is that we can model the time difference between the two sets of data by appealing to a correction factor. Crucially we can do this by encoding the relevant information in sentential form more or less as follows: 'Data taken by Weber's team can be compared to data taken by Douglass' team only after they are offset by 4 h 1.2 s'.

The second problem also needs some stage setting. Bogen and Woodward explain the appeal of sidereal correlations:

> Weber also relied on facts about the causal characteristics of the signal—the gravitational radiation he was trying to detect. The detectors used by Weber were most sensitive to gravitational radiation when the direction of propagation of given radiation was perpendicular to the axes of detectors. Thus if the waves were coming from a fixed direction in space (as would be plausible if they were due to some astronomical event), they should vary regularly in intensity with the period of revolution of the earth. Moreover, any periodic variations due to human activity should exhibit the regular 24 h variation of the solar day. By contrast, the pattern of change due to an astronomical source would be expected to be in accordance with the sidereal day which reflects the revolution of the earth around the sun, as well as its rotation about its axis, and is slightly shorter than the solar day. When Weber initially appeared to find a significant correlation

[8] The third one is that none of the competing groups of scientists were able to replicate Weber's results.

[9] Franklin discusses additional difficulties afflicting Weber's methods and results.

with sidereal, but not solar, time in the vibrations he was detecting, this was taken by many other scientists to be important evidence that the source of the vibrations was not local or terrestrial, but instead due to some astronomical event (p. 246).

The honeymoon did not last long. None of the other groups that set about to confirm these results were able to detect a sidereal correlation. Even Weber himself was unable to repeat his earlier 'success'. Once more, what is relevant to our discussion is whether IRS models can capture the relevant idiosyncratic and highly heterogeneous evidential considerations. It is not hard to imagine that they can. The relevant formulation will not look very different to this: 'The intensity of any signal *k* that meets threshold *r* should co-vary with the sidereal day and, moreover, such positive correlations should be reproducible by other similarly constructed experiments'.

6 Conclusion

In closing, I would like to summarise some of the main points I have argued for. In Sect. 3 I have argued for the view that prima facie non-sentential data can be reconstructed in sentential form without loss of evidential content. To support this view I furnished sentential reconstructions of Bogen and Woodward's central examples of non-sentential data. What is more, I supplied general reasons why sentential reconstructions are always possible, namely that the widespread automation of data processing implies that in such cases the data are already in sentential form or can easily be converted into it and that the expression of assent or dissent towards a set of data is at least an implicit endorsement of their having propositional content. I concluded that section by challenging Bogen and Woodward to come up with one datum whose evidential worth cannot be sententially recreated.

Arising from Sects. 4 and 5 is, I hope, a clear picture of why Bogen and Woodward's insufficiency objection is hard to maintain. Neither the raven paradox nor the doppelganger data scenario manage to box in the IRS theorist who can always rely on syntactical as well as extra-syntactical tools to restore the model's sufficiency. Considerations concerning general and local reliability fare no better. The IRS theorist need not be in possession of a general theory that explains the workings of an instrument or method. In so far as we have independent reasons to believe in the reliability of the method or instrument, reconstructing a suitable auxiliary does not seem beyond reach. Finally, the cited examples of idiosyncratic and highly heterogeneous evidential considerations appear to be as amenable to IRS formulation as the rest of the cases discussed.

Acknowledgement This work was in part made possible by funding from the German Research Foundation (Deutsche Forschungsgemeinschaft).

References

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review, 97*, 303–352.
Bogen, J., & Woodward, J. (1992). Observations, theories and the evolutions of the human spirit. *Philosophy of Science, 59*(4), 590–611.

Bogen, J., & Woodward, J. (2003). Evading the IRS. Poznan studies in the philosophy of the sciences and the humanities. In R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the model* (pp. 233–268). Amsterdam: Rodopi.

Franklin, A. (1994). How to avoid the experimenter's regress. *Studies in History and Philosophy of Science, 25*(3), 463–491.

Glymour, C. (1980). *Theory and evidence*. Princeton: Princeton University Press.

Hempel, C. G. (1943). A purely syntactical definition of confirmation. *The Journal of Symbolic Logic, 8*(4), 122–143.

Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: The Free Press.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. La Salle, IL: Open Court.

Votsis, I. (forthcoming). Making contact with observations. In M. Suárez, M. Dorato, & M. Rédei (Eds.), *EPSAO7: Launch of the European Philosophy of Science Association*. Springer.

Woodward, J. (1989). Data and phenomena. *Synthese, 79*(3), 393–472.