Evidence Enriched

Nora Mills Boyd*†

Traditionally, empiricism has relied on the specialness of human observation, yet science is rife with sophisticated instrumentation and techniques. The present article advances a conception of empirical evidence applicable to actual scientific practice. I argue that this conception elucidates how the results of scientific research can be repurposed across diverse epistemic contexts: it helps to make sense of how evidence accumulates across theory change, how different evidence can be amalgamated and used jointly, and how the same evidence can be used to constrain competing theories in the service of breaking local underdetermination.

1. Introduction. The epistemology of science ought to include some account of empirical constraints on theorizing about nature. It does not help to say merely that the world 'pushes back' or to appeal as Quine did to the "tribunal of experience." Veiled by these metaphors is something very important—the thing that makes natural science distinctively empirical.

Whatever we philosophers of science want to say about this 'pushing' or 'tribunal', it ought to accommodate not only naked-eye observations but the sort of results germane to the sophisticated machinations of contemporary technology-ridden science. The fact that the output of scientific instrumen-

Received June 2017; revised November 2017.

*To contact the author, please write to: Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260; e-mail: nmb58@pitt.edu.

†I am grateful for extensive and constructive feedback on this work from John Norton, Aaron Novick, and David Colaço. In addition, Robert Batterman, James Woodward, Christopher Smeenk, Thomas Pashby, Casey McCoy, Michela Massimi, Matthew Brown, Porter Williams, and Slobodan Perović all deserve my sincere thanks for their advice and for insightful discussions about this piece. I would also like to acknowledge helpful comments from audience members at the University of Pittsburgh History and Philosophy of Science graduate student work in progress talk series, the Sixth Biennial Conference of the Society for the Philosophy of Science in Practice, the work in progress talk series at the Department of Philosophy, Psychology, and Language Sciences, University of Edinburgh, the London School of Economics Sigma Club, and the 2017 meeting of the Canadian Society for the History and Philosophy of Science.

Philosophy of Science, 85 (July 2018) pp. 403–421. 0031-8248/2018/8503-0004$10.00
Copyright 2018 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

tation eventually needs to make a transcranial journey in order to be of any real epistemic interest ought not mislead us into thinking that the empirical is best understood as 'observable' or 'sensible'. Indeed, this was the sticking point for van Fraassen's constructive empiricism for many of the philosophers of science who engaged with that view. Making what is observable to creatures like us the linchpin of one's empiricist philosophy of science ends up ostracizing much of what scientists actually do in practice and does not seem to get at what makes something distinctively empirical anyway.[1]

If not observations, what does constrain our theorizing about nature such that some theories are empirically viable and some are not? In the hope of replacing observations with something more suitable to science in practice, we might consider the more generic 'empirical results', where 'results' may be understood to include observations and other sensings but also the results of technology-aided detections and measurements, and 'empirical' may be understood in contrast with 'virtual' and 'imagined' and could be cashed out by appeal to a causal story connecting the target of interest to the generation of that result.

This first attempt encounters an immediate worry: empirical results are typically generated and interpreted by recruiting significant theoretical resources. The connectedness, or intertwining, of the theoretical and empirical is often associated with the sort of holism attributed to Duhem and Quine.[2] Thus, the role of Quine's tribunal of experience is to judge not individual statements about the external world but the whole "corporate body" of such statements (Quine 1951, 38). Indeed, according to Quine, "The unit of empirical significance is the whole of science" (39). This holism is then taken to have the consequence that there is much flexibility in accommodating recalcitrant evidence, and indeed that nothing forces one way of accommodating rather than another. Thus, according to Duhem, "the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed" (1954/1974, 187). Duhem and Quine both respond to this quandary with pragmatic resources: for Duhem, the physicist's 'good sense' and for Quine a penchant for conservatism and simplicity. Inviting theory into our conception of the empirical therefore seems to have the unfortunate consequence of making scientific theory choice a matter of pragmatics

1. Van Fraassen himself begins to address this problem in his work on measurement and measuring instruments (2008). I will have more to say about the views of twenty-first-century van Fraassen below in sec. 3.

2. I would like to acknowledge anonymous referees for pushing me to clarify how my view of enriched evidence relates to what is often called the Quine-Duhem problem.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

rather than conformity with experience. The effect of all this is that what was distinctively empirical about empirical science drops out of view.

The intertwining of the theoretical and empirical to which Duhem and Quine brought attention has been absorbed into philosophy of science since the practice turn as the lesson that the epistemic utility of empirical results depends crucially on the details of their provenance. One must understand the concepts and assumptions that have shaped the presentation of the result in order to use it in an epistemically responsible way.

However, it has not yet been widely appreciated that appeal to the auxiliary information associated with the provenance of empirical results solves several questions left open at least since the logical empiricist program dwindled. In particular:

1. How can evidence accumulate across theory change?
2. How can evidence be combined and used jointly?
3. How can the same evidence be used to constrain competing theories?

These questions are not independent of one another; they all concern the relationship between epistemic utility and context. To accumulate, evidence must outlive its original context. To be used jointly, differently sourced evidence must be amenable to the same context. To constrain competing theories, the same evidence must be adaptable to different contexts.

What I want to argue here is that with the right understanding of empirical evidence we can appreciate the sense in which the intertwining of the theoretical and empirical actually affords epistemic activities that we care about, and it does so in such a way that what makes empirical science distinctively empirical remains in view. I will argue that the epistemic utility of empirical results depends on the details of their provenance and that this dependence is what makes possible the accumulation and amalgamation of evidence and indeed the breaking of local underdetermination. The main contribution of this argument will be to show how empiricism can embrace theory-riddled evidence.

2. Enriched Evidence. The history of philosophy exhibits a variety of empiricisms.[3] The shared gestalt is that knowledge of nature derives from, and only from, experience. Of course "experience" is vague, and what gives

3. Lipton (2001) puts it nicely: "There are almost as many empiricism as there are empiricists, but what these views or approaches have in common is an emphasis on the importance of experience to the formation of concepts and to the acquisition of knowledge. . . . The range of empiricist positions is vast, from the shocking view that all we can think or know about are our sensations to the mundane claims that experience plays some role in the formation of some of our concepts and in the justification of some of our beliefs" (4481).

https://doi.org/10.1086/697747 Published online by Cambridge University Press

any empiricism substance is an explication of this concept. Let me begin by presenting the view that I think is required if empiricism is to remain relevant in the face of the increasingly intricate instruments and techniques prevalent in scientific research today. To this end it will be best to leave behind talk of "experience" right away and speak instead of empirical evidence. Minimally, an empiricist should be committed to requiring that theories of the natural world be consistent with the available empirical evidence. To do otherwise would betray the very heart of the empiricism: it would sever the connection by which the world could possible 'push back', by which the 'tribunal' could possibly judge. Note that requiring that theories be consistent with the evidence does not commit one to naive falsificationism. In particular, that good theories need to be consistent with the available empirical evidence does not mean that whenever a theory encounters anomalous evidence it should be abandoned without further regard since it may be reasonable to work on, or keep around, a theory that is inconsistent with the available evidence as far as we know. But it does mean that when theories are inconsistent with evidence, something has eventually got to give. An inconsistency between theory and evidence cannot persist if the theory is to be empirically viable. The ground-level task of giving substance to empiricism now becomes explicating the notion of empirical evidence. With respect to what exactly are our theories supposed to be consistent?

Given the centrality of the notion of evidence in philosophy of science, it is surprisingly difficult to find explicit characterizations of it. This situation is captured well by van Fraassen (1984): "What is the main epistemic problem concerning science? I take it that it is the explication of how we compare and evaluate theories, as a basis either for theory acceptance or for practical action. This comparison is clearly a comparison in the light of the available evidence—whatever *that* means" (27). Van Fraassen's appraisal remains salient with respect to the contemporary literature, which rarely defines evidence explicitly and often passes over the issue in silence by dealing abstractly with "evidence *e*."

Thorough explication of the view I want to advance, the enriched view of evidence, will have to proceed in several stages and will be aided by the introduction of some new conceptual resources. However, let me state the view right away with the caveat that the unfamiliar terms will be defined and illustrated in due course.

> **Enriched Evidence.** The evidence with respect to which empirical adequacy is to be adjudicated is made up of lines of evidence enriched by auxiliary information about how those lines were generated. By "line of evidence" I mean a sequence of empirical results including the records of data collection and all subsequent products of data processing generated on the way to some final empirical constraint. By auxiliary information, I mean

https://doi.org/10.1086/697747 Published online by Cambridge University Press

> the metadata regarding the provenance of the data records and the processing workflow that transforms them. Together, a line of evidence and its associated metadata compose what I am calling an "enriched line of evidence." The evidential corpus is then to be made up of many such enriched lines of evidence.

This characterization of evidence is sympathetic with the spirit of characterizations given by other philosophers of science who attend carefully to scientific practice. For instance, Bogen and Woodward (2005) emphasize the fact that "evidential relevance depends upon features of the causal processes by which the evidence is produced" (240). I agree with Bogen and Woodward (2011) that philosophers of science need to attend more closely to data-generating processes in our efforts to understand the epistemic relevance of evidence. In their chapter in the edited volume *Evidence, Inference and Enquiry*, Chang and Fisher (2011) argue for "the intrinsic contextuality of evidence" and for the importance of locating evidence within purposeful epistemic activities, operations, and procedures. Perović (2017, sec. 6) argues for a "relaxed stance" toward calibration procedures that incorporate past empirical results, theory, and the outcomes of the very experiments under consideration that is compatible with empiricism broadly construed. I hope that the characterization of empirical evidence introduced in the present work will be a welcome elucidation of a concept of central significance to philosophers working in this problem space. I will say a bit more below to locate my view with respect to van Fraassen (2008) and Leonelli (2009, 2016). Before I do, I should further unpack the notion of enriched evidence. To this end it will be useful to further countenance two important components of the characterization given above: empirical results and metadata.

Empirical results.—Here is a generic sketch of the generation of an empirical constraint. Let us focus on two (roughly delineated) stages of empirical research: data collection and data processing. In the first stage data are collected and recorded. Sometimes the data collected are observational and the collection consists in unaided human perception, which is then codified in some record, as may be the case for naked-eye astronomical observations, such as gazing at the moon. However, as we have already noted above, especially in contemporary science, data are often collected using instruments and/or techniques.

In the second stage, data may be processed in a variety of ways. The original records of data collection typically sustain "cleaning," "cuts," "reduction," and calibration as they are transformed into models of data. For instance, the process of reducing a set of images from a digital telescope might involve (1) correcting each exposure (bias subtraction, flat field correction, bad pixel masking); (2) calibrating each exposure astrometrically and photometrically; (3) modeling the point spread function in each exposure; (4) remapping each

https://doi.org/10.1086/697747 Published online by Cambridge University Press

exposure to a common coordinate system; and (5) co-adding exposures and so forth.[4]

Furthermore, in order to construct an empirical result that is appropriately formulated to constrain some theory, for example, to calculate the empirical value of a particular parameter or to produce a proposition, much more processing than preliminary data reduction will typically be required. The target system under study may have to be modeled and the data interpreted in light of that model. Anderl (2016) gives a nice example of this sort of modeling in radio astronomy: "the recording of data using a single dish radio telescope requires a model of the mechanical and optical properties of the telescope mirror in different positions in order to determine the exact pointing position. For the calibration of data with respect to atmospheric influences, a model of the Earth's atmosphere is needed. Flux calibration presupposes models of the individual stars and planets used in the calibrating observations" (664). In addition, the features of the modeled system may have to be processed further so as to speak to higher-level theories.

The records of the data as transformed by the sequence of data-processing steps—as well as the original records of data collection—are all what I will call "empirical results." The collection of empirical results for a given sequence of data collection and processing stages is what I will call a "line of evidence."

Not all empirical results are useful as constraints on theory. To be useful as a constraint on theory, an empirical result must be well adapted to that theory. To see when a result is well adapted to a theory it is helpful to consider what could make it maladapted. First of all, it is clear that results presupposing concepts, parameters, or other such vehicles that are not found in the theory to be constrained will be maladapted to that theory. Consider ancient Chinese records of astronomical events. These observations were recorded using categories quite different from those of contemporary theorizing. The records refer to *k'o-hsing* ("guest stars" or "visiting stars"), *po-hsing* ("rayed stars" or "bushy stars"), and *hui-hsing* ("broom stars" or "sweeping stars"), not, say, "comets" and "supernovae" (cf. Clark and Stephenson 1977, 40). Contemporary astronomers want to use the content of these records as constraints on their own theoretical frameworks. However, the conceptual vocabulary in which the records are expressed cross-cuts the concepts available in the contemporary framework: the ancient observations are, taken at face value, maladapted to the contemporary epistemic context in which the constraint is to occur. Therefore, if constraints on contemporary theories are to be generated from the ancient results, some work will have to be done to connect those re-

4. This is a partial list from Neilsen's Notes on the Essentials of Astronomy Data: http://home.fnal.gov/~neilsen/notebook/astroImagingDataReduction/astroImagingDataReduction.html.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

sults up to the theories of interest. New and different well-adapted results will have to be generated from the ancient ones.

Another initially plausible thought is that a result is maladapted to the theory to be constrained when presuppositions derived from a genuine competitor theory are incorporated in the data processing that generates that result. But this is not quite right: incorporating presuppositions from a genuine competitor need not generate a maladapted result.

Laymon (1988) discusses just such a case in the context of the Michelson-Morley experiment looking for an effect of aether velocity on the speed of light. According to Laymon, Michelson modeled his experimental apparatus using simple single-ray optics that made assumptions formally inconsistent with the theory to be constrained. However, using the consistent assumptions would have resulted in a fourth-order correction in the context of an experiment that was sensitive only to second-order effects and thus did not make a significant difference (Laymon 1988, 258).

In light of this we will say that in order to constrain some theory, an empirical result must be "well adapted" (meaning well adapted to the context of constraint) and that an empirical result is well adapted either when all of the presuppositions that have been incorporated into it throughout the course of data collection and processing are formally compatible with the theory to be constrained or else their incorporation does not make a relevant difference to the constraint. Here, "not making a relevant difference" means that if the incompatible presuppositions were replaced by compatible ones, the judgment of the consistency of the theory with the resulting constraint would not be affected.[5] That is, the incorporation of the incompatible presuppositions does not influence the constraint thereby obtained in a manner that differs significantly from the influence that formally compatible assumptions would have imparted had they been incorporated instead. Here I use the phrase "formal compatibility" to refer to formal consistency and the sharing of a common conceptual framework and "context of constraint" to encompass both the theory at hand and the norms of constraint belonging to the discipline in question (e.g., conventional standards of statistical significance).

With these preliminaries in hand, let us return to our central question: with respect to what exactly are our theories supposed to be consistent?

Empirical results are not good candidates for explicating the "tribunal of experience" because the evidential corpus composed of empirical results is inconsistent and it would be a fool's errand to require our theories to be consistent with something that itself lacks consistency. Time and time again it looks like science produces result *R* and then promptly not-*R*. Franklin (2002, 35) captures this idea succinctly: "it is a fact of life in empirical science that exper-

5. See Miller (2016) for a discussion of when theoretical and measurement uncertainties make a difference for empirical adequacy.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

iments often give discordant results." Discord is particularly easy to see in the case of empirically derived parameter values. Consider for instance the value of H_0, the Hubble parameter today, the current rate of expansion of the universe. Edwin Hubble's original value derived from observations of Cepheid variable stars in the early twentieth century was a rough 500 kilometers per second per megaparsec (km/s/Mpc), whereas the latest value derived using data from the Planck cosmic microwave background satellite is 67.8 ± 0.9 in the same units (Hubble 1929; Planck Collaboration 2016). These values manifestly disagree. If evidence is discordant, it is not cumulative and it cannot be amalgamated and deployed in joint constraints. Neither are lines of evidence good candidates for explicating the "tribunal of experience" since lines of evidence are just collections of empirical results.

In addition to lines of evidence, we need to include metadata in our conception of the evidence with respect to which empirical adequacy is to be adjudicated. Each empirical result produced in the course of data collection and processing has associated metadata.[6] Let us consider two types: "provenance" metadata (associated with the data collection stage of research) and "workflow" metadata (associated with the data-processing stage of research). In the sense intended here, metadata are auxiliary information about empirical results. For example, in the case of volcanology, where data include rock samples, provenance metadata include identifiers signifying the field campaign and the researcher who collected the sample, geographical information system coordinates of the sample collection site, date and time of collection, description of surrounding environment and weather conditions, description of the specimen condition at the time of collection, and narrative field notes that record anomalous conditions and other details deemed relevant (Palmer, Weber, and Cragin 2012, 7–8).

Workflow metadata might include (in the case of potassium-argon dating, for instance) details about the atomic absorption spectrophotometer used to date the rock samples (including a description of the apparatus and procedure used), corrections for atmospheric contamination, background information on radioactive isotopes including isotopic abundances and decay series, formulas for calculating time since the rock cooled from quantities of isotopes in the sample, and a variety of assumptions including lack of contamination from nonradiogenic ^{40}Ar (cf. McDougall and Harrison 1999).[7]

6. See Leonelli (2014) for a discussion of the importance of metadata for assessing the epistemic relevance of biological data shared in online databases.

7. Leonelli (2009) characterizes metadata for biological data shared in databases as "including 'evidence codes' classifying each data set according to the method and protocol through which it was obtained, the model organism and instruments used in the experiment, the publications or repository in which it first appeared, and the contact details of the researchers responsible, who can therefore be contacted directly for any question not answered in the database" (741).

https://doi.org/10.1086/697747 Published online by Cambridge University Press

I will refer to lines of evidence considered together with their associated metadata as "enriched lines of evidence." One can discern enriched lines of evidence in fields from climate science to molecular biology to particle physics.[8] The data management strategies and techniques will of course vary from field to field, but the broad-brush elements are shared across the sciences. One can think of an enriched line of evidence in analogy with Peter Railton's notion of an ideal explanatory text.[9] Railton (1981) suggests that acceptable explanations, which genuinely convey explanatory information, need not be maximally specific. An informative answer to the question "Why is this one lobster blue?" need not invoke all details of evolutionary theory and particular conditions associated with the individual, but could be simply "It's a random mutation," very rare (239). Similarly, although all of the presuppositions that contribute to the generation of an empirical constraint are implicated in the epistemic relevance and adaptedness of that constraint to theoretical contexts, in practice the entire enriched line of evidence need not be hauled out for appraisal every time. For instance, researchers may have good reasons to believe that the instrument used to collect data was well calibrated without checking all the available information relevant to that calibration. However, reason to be suspicious of the instrument's calibration could always arise later on, and revisiting the information available about the calibration could become epistemically imperative. We can often take things for granted. Until we can't.

Thus one can think of enriched lines of evidence as including the rich (perhaps bottomless) reservoirs of background information implicated in the production of an empirical constraint. Different circumstances will call for interrogating this reservoir to various extents.

Before I go on to discuss some benefits of adopting an enriched view of evidence in the next section, allow me to briefly comment on the relation of this view to the work that Leonelli has done on data, database curation, and traveling facts (cf. Leonelli 2009, 2013, 2015, 2016; see also Howlett and Morgan 2010). I am broadly sympathetic to the approach that Leonelli takes. In particular, I share her interest in understanding how it is that the products of empirical science are in fact fruitfully and responsibly shared across epistemic contexts—how such products are reused and repurposed. Indeed, I think that focusing on understanding such successful transfer across contexts gets at issues of interest to many philosophers of science; for instance, those interested in epistemic progress and theory change after Kuhn (1975), generalization,

8. See http://lhcb-elec.web.cern.ch/lhcb-elec/html/architecture.htm for a description of the front-end electronics implemented in the LHCb experiment. Jenni et al. (2003) is the full technical report on the ATLAS trigger and data acquisition system. See, e.g., Perović (2017) for a philosophically informed discussion of calibration at the Large Hadron Collider, especially secs. 3–5.

9. Chris Smeenk and Porter Williams independently suggested this analogy to me.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

replication, triangulation, ecological validity, and other such epistemic issues in the epistemology of experiment.[10]

An important aspect of Leonelli's account of how biological data travel to different epistemic contexts involves two moves: *decontextualization* and *recontextualization* (cf. Leonelli 2016, sec. 1.2.3). According to Leonelli, in the decontextualizing move, data "are at least temporarily decoupled from information about the local features of their production" (30). In Leonelli (2009, 746), she discusses this move as the "liberation" of data from the details of their provenance. In particular, she argues,

> Data that travel through databases become nonlocal. They travel in a package that includes information about their provenance, but they can be consulted independently of that information. This is a way to 'free' data from their context and transform them into nonlocal entities since the separation of data from information about their provenance allows researchers to judge their *potential relevance* to their research. This is different from judging the *reliability* of data within a new research context. This second type of judgment requires researchers from the new context to access information about how data were originally produced and match it up with their own (local) criteria for what counts as reliable evidence, as based on the expertise that they have acquired through their professional experience in the lab. What counts as reliable evidence depends on scientists' familiarity with and opinion of specific materials (e.g., the model organism used), instruments, experimental protocols, modeling techniques, and even the claims about phenomena that the evidence is produced to support. Thus, data judged to be reliable become local once again: what changes is the research context that appropriates them. (747–48)

I take it that the picture is something like this: potential data users can reasonably window-shop curated databases without having all of the details of the provenance of the data encoded there ready at hand; but when those users want to get down to the business of actually repurposing some data in a new context, the background provenance information (and new information associated with the new context) must be involved. This picture is consistent with the enriched view of evidence I have articulated. As I stated above, in practice the entire enriched line of evidence need not be hauled out for appraisal in every circumstance.

10. Two notable examples are Colaço (2018), which engages with these topics in the context of biology, psychology, and neuroscience, and Matthiessen (2018), which discusses how theoretical and practical knowledge support repurposing data across diverse epistemic contexts.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

If this is the right way to understand Leonelli's position, then I would submit that it is misleading to speak of "decontextualization" and "liberation" as she does. The epistemic utility of empirical results depends crucially on the details of their provenance. Epistemically responsible use of empirical results (such as data) depends on access to their associated metadata: data can never be permanently decoupled from their associated enriching information and retain epistemic utility. Epistemically useful data are never fully liberated of the details of their provenance; their utility derives from their enrichment by such details.

3. Benefits of Enriched Evidence. Let us take stock. Enriched evidence in the sense articulated in the previous section is an account of what our theories of the natural world are supposed to be consistent with that accommodates sophisticated contemporary scientific research, theory-informed practice and all. Moreover, it does so in a manner consonant with empiricist scruples, that is, without invoking 'good sense' or extra-empirical virtues such as conservatism or simplicity à la Duhem and Quine. In the remainder of this article, I want to draw out what I think are three major benefits that adopting the enriched view of evidence affords; namely, adopting this view helps to make sense of how evidence accumulates across theory change, how different evidence can be amalgamated and used jointly, and how the same evidence can be used to constrain competing theories in the service of breaking local underdetermination.

Accumulation.—Empirical results are bound to be lost in the transition out of their native epistemic contexts when they are maladapted to the receiving context. However, it may be possible to salvage a constraint in the new context, as long as enough information is available about how the result in question was generated to backtrack through the stages of data processing in order to find a product of an earlier stage that is adaptable to the theory to be constrained and reprocess using its own resources, thereby generating a well-adapted result. In this way, enriched lines of evidence provide the resources with which a particular empirical result can be brought to bear on frameworks besides those originally used in the generation of that result.

Recall the ancient Chinese astronomical observations records, expressed in categories *k'o-hsing*, *po-hsing*, and *hui-hsing*, which cross-cut contemporary ones, "comets" and "supernovae." Of astronomical events recorded using these historical terms, contemporary astronomers would like to know which, if any, are relevant to supernovae. The *hui-hsing* are the easiest to rule out: they are described as a star with a definite tail, and we would categorize them as comets today. In contrast, *po-hsing* "is the standard term to describe an apparently tail-less comet" (Clark and Stephenson 1977, 40). However, there is the possibility of mistakenly translating an observation of a *po-hsing* as an observation of a comet when it is in fact a record of a nova. There are

https://doi.org/10.1086/697747 Published online by Cambridge University Press

some records of motionless *po-hsing*, and a motionless new star without a tail could have been a nova. Regardless, when the duration of the visibility of these new stars was recorded, they are too short to be supernovae, so *po-hsing* can also be ruled out. For instance, translating *ko-hsing* observations is not always straightforward. Clark and Stephenson offer the following:

> Ko-hsing (which will be subsequently abbreviated to ko) seems to have been the general term to describe a new star-like object. The well known new stars of AD 1006, 1054, 1572, and 1604 were identified in this way and we might thus expect ko to be synonymous with novae and supernovae. On the other hand, there are frequent references to moving ko throughout oriental history (more than 20 are catalogued by Ho Peng Yoke, 1962), so that usage of the term must be treated with caution. The nucleus of a comet resembles a star, so that if no tail is evident confusion seems possible. (40)

Astronomers mining these historical records need to be wary of the possibility of comets interloping as novae and supernovae.

Nevertheless, with enough enriching information it can be possible to generate constraints on contemporary theorizing using these historical records. Quantitative modeling of the evolution of supernovae and their remnants depends on precise dating of stages of the process. To take just one example, careful historical work on Chinese records of the supernova of July 4, 1054, has allowed researchers to precisely date the end of the visibility of the event. In particular, by carefully interpreting a Chinese observation record, Stephenson and Green extract the date of April 6, 1056 (Green 2015, 97).[11]

Will it always be possible to adapt initially maladapted results to the context of interest? Unfortunately not. Consider a data record that is maladapted to some epistemic context. One can come to know that the record is maladapted in the first place by having access to the associated provenance metadata that include information about in what way the record is maladapted. This very information would tell us that it will be impossible in practice to generate a useful constraint on theory from those data. In a sense this means that the evidence associated with the data must be lost in the transition between epistemic contexts under consideration.

This loss is not as epistemically problematic as the loss of empirical results more generally. If as a part of our philosophy of science we characterize evidence as detached empirical results or as unenriched lines of evidence, then evidence appears to be lost all over the place. However, construing empirical science as replete with such loss is both descriptively inadequate with

11. For further success stories see Clark and Stephenson (1977) and Stephenson and Green (2002).

https://doi.org/10.1086/697747 Published online by Cambridge University Press

respect to actual scientific practice and ill-advised epistemically. With so much evidence "lost," the cheapness of empirical adequacy would look dangerously like cherry-picking. Yet, as I have noted above, scientists do manage to repurpose results across epistemic contexts, and it is desirable to do so when possible because this generates more empirical constraints. However, if some constraints that we would like to have as a matter of fact cannot be generated, there is little to be done except move on to generating constraints in another way. So it goes.

Furthermore, with the resources of an enriched view of evidence, we can account for how it is that empirical adequacy is supposed to be adjudicated with respect to a corpus of evidence that contains discordant empirical results. If pieces of empirical evidence really were discordant with one another, then evidence again would not be cumulative. However, the collection of empirical results considered together with auxiliary information about how they were generated is not internally inconsistent, just as there is no contradiction between "If *x* then *p*" and "If *y* then not *p*," even though there is one between *p* and not *p*. Thus, returning to the example of the discordant values of the Hubble parameter, Hubble's estimated value of a rough 500 km/s/Mpc conditioned on the presuppositions with which it was generated should not be inconsistent with the Planck satellite value of 67.8 ± 0.9 conditioned on the presuppositions with which it was generated.

To see more concretely how the enriched view of evidence helps to make sense of how evidence can accumulate across epistemic contexts, let us briefly consider an example from the history of particle physics from Franklin (2015, 159) (and discussed by Galison [1987]), the experiment that eventually discovered the existence of weak neutral currents: "When the experiment was initially conceived, it was a rule of thumb in particle physics that weak neutral currents did not exist. The initial design included a muon trigger, which would be present only in charged current interactions. In a charged-current event a neutrino is incident and a charged muon is emitted, in a neutral-current even there is a neutrino in both the initial and final states, and no muon is emitted. Thus, requiring a muon in the event trigger would preclude the observation of neutral currents." In other words, the original experimental design would have essentially filtered for interactions that produce muons and thus filtered out the weak neutral currents that the Weinberg-Salam electroweak theory posited. Fortunately, as Franklin explains, the experimentalists realized this problem in time and changed the experimental design. But suppose the original experimental design had been retained. Any viable theory would still have had to be consistent with the empirical evidence that would have thereby been produced. That is, any empirically viable theory would have had to be consistent with the results of the counterfactual experiment considered together with the presuppositions that went into their generation. If results consistent with no neutral currents had been produced from the original

https://doi.org/10.1086/697747 Published online by Cambridge University Press

experimental design, such results would still have been consistent with the existence of neutral currents since the experiment was organized in such a way that regardless of whether neutral currents existed or not, the experiment would not have been sensitive to them on account of the muon trigger. So it is not the case that the prediction of neutral currents derived from the Weinberg-Salam theory would have been inconsistent with the enriched evidence produced in the counterfactual experiment. In fact, had the experiment been performed as originally intended, ill-advised muon trigger and all, the enriched evidence thereby produced would still belong in the cumulative evidential corpus. Indeed, the enriched evidence associated with this experiment would have been something that any theory—theories positing weak neutral currents and those omitting them—would have to be consistent with to be empirically viable, that is, viable at all for an empiricist.

Amalgamation.—That the epistemic utility of empirical results depends on the presuppositions incorporated into those results throughout data collection and data processing might cause one to worry about the feasibility of combining evidence in an epistemically responsible way. An enriched view of evidence also helps to make sense of how evidence produced using significantly different instruments and techniques might be fruitfully combined. In fact, there is a danger that if enriching information is not taken into account, results used in joint constraints could interact in epistemically problematic ways.

Consider the multiprobe approach to constraining theorizing about dark energy in contemporary cosmology. "Dark energy" is a placeholder for whatever is responsible for the accelerated expansion of the universe, inferred from telescopic observations of distant supernovae. Very little is presently known about the nature of dark energy. Indeed, cutting-edge research is largely concerned with trying to discern whether dark energy behaves as a cosmological constant or if its contribution to the energy density budget of the universe evolves over cosmic time. To tackle this question, cosmologists are combining different data sets gathered in a variety of ways. For instance, the approach taken in the Dark Energy Survey (DES) combines cosmic shear, galaxy-galaxy lensing, galaxy clustering, Baryon Acoustic Oscillations, galaxy cluster number counts, and Type Ia supernova (Krause 2017). However, as the DES cosmologists are aware, it is not always appropriate to simply calculate the constraints on the theoretical parameters of interest for each probe in parallel and then combine the constraints thereby derived afterward. Care must be taken in combining the different galaxy survey probes, because they "are highly correlated with each other in that they are tracers of the same underlying density field, and in that they share common systematic effects" (3). Effectively combining results from these different probes requires paying attention to the details that have gone into analyzing them. Without conscientious treatment of how the systematic errors associated with each probe interact, the joint constraints

https://doi.org/10.1086/697747 Published online by Cambridge University Press

could be constructed in a way that obscured the shared systematics and thereby delivered the wrong pronouncement on the parameters given the empirical results.

In other words, combining results from DES probes in a responsible way requires knowing what presuppositions have gone into those results. Note, though, that knowing what presuppositions have gone into the results would be required even if the results were suitably independent from one another such that they could be straightforwardly combined after parallel processing. Knowing that results can be straightforwardly combined requires knowing that nothing has been baked into those results during analysis that will cause problems in the epistemic context of interest. This is true not just of the results from DES probes but of results generally. Whether and how results can be combined and used in joint constraints depends on the presuppositions those results have incorporated.

Breaking underdetermination.—Temporary underdetermination is a ubiquitous feature of scientific research. There are often multiple empirically viable theories (or models or hypotheses) of some target. In addition, scientists often want the same empirical evidence to constrain multiple alternatives, for instance, the same observational evidence used to constrain competing theories of dark matter, including theories that cast the ontology of dark matter in radically different terms—as a particle/substance or as a feature of gravitation. Given that empirical results are often heavily processed and often involve presupposing resources from the very theory that they are generated to constrain, how is it that the same evidence could be used to constrain alternative theories? On the enriched view of evidence, the answer is clear: with the help of enriching information, elements of a line of evidence can be repurposed to many contexts of constraint. For instance, the same galaxy rotation curve data can be processed in multiple ways to constrain parameters relevant to different proposals for dark matter particles and to different gravitational theories.

The availability of this answer is a benefit that the enriched view has over the view that van Fraassen articulates in his 2008 book *Scientific Representation: Paradoxes of Perspective*. There, van Fraassen makes a significant step forward in reconciling our conception of evidence with the minimal commitment of empiricism. He countenances checking for the empirical adequacy of theories as an attempt to match the structures of theoretical models and smoothed-out data models. His insight is that the epistemic significance of this matching relies on the relevance of the data model to the theory and that such relevance is appreciated only by contextualizing the data model: "A particular data model is relevant because it was constructed on the basis of results gathered in a certain way, selected by specific criteria of relevance, on certain occasions, in a practical experimental or observational setting, designed for that purpose" (253). Adjudicating the empirical adequacy of a theory requires

https://doi.org/10.1086/697747 Published online by Cambridge University Press

identifying results relevant to that theory. But as van Fraassen rightly recognizes (and as I have argued above), auxiliary information about the particularities of data collection, processing, and analysis is crucial for discerning the relevance of a data model to any theory. Van Fraassen's insight brings into focus the futility of considering bare results in the absence of auxiliary information about their manner of production as empirical evidence at all. Having access to the auxiliary information is critical for (merely) judging the relevance of empirical results. Without auxiliary information, results (such as 125, 10^9 electron volts, 13.8 billion years, a plot, a photograph) are just free-floating.

Although contextualizing results in the manner that van Fraassen suggests is an important step, he does not fully exploit the consequences of this move. I suspect that the reason for this is that empirical adequacy is not the primary problem with which he engages in his 2008 work. Instead, van Fraassen's insight leads him to a solution of what he calls the *Loss of Reality Objection* (258). According to van Fraassen, the objection is a sort of puzzle for any empiricist account of science; namely, how can it be that our theories are constrained by the way that the natural world is, when empirical adequacy is adjudicated by matching models of theory to data models rather than to nature itself? His own answer rests heavily on including representation users in our understanding of representations. Instead of casting representation as a two-place relation (between, e.g., a data model and some phenomenon), van Fraassen understands representation as three-place: "Nothing represents anything except in the sense of being used or taken to do that job or play that role for us" (258).

Van Fraassen illustrates this point with an illuminating imagined conversation between a scientist and a metaphysician (254–57). The scientist presents a graph *S* representing the deer population growth in Princeton, which fits with a model of some theory *T*. The metaphysician serves as the voice of the Loss of Reality Objection wondering whether *T* fits the actual deer population in Princeton. Van Fraassen's scientist responds, "Since this is *my* representation of the deer population growth, there is *for me* no difference between the question whether T fits the graph and the question whether T fits the deer population growth" (256). Van Fraassen likens this situation to the "pragmatic tautology" (aka T-schema): "The sentence 'Snow is white' is true if and only if snow is white" (n. 26). For van Fraassen the requisite link between a data model and reality crucially involves locating the representation user, as in "a theory is empirically adequate to the phenomenon as represented by us" (259). Moreover, the pragmatic tautology is supposed to quell the worry that all we can ever say is that theories are empirically adequate with respect to the natural world under some description (which is, after all, not the natural world itself), by collapsing the deer population growth as represented in *S* and the

https://doi.org/10.1086/697747 Published online by Cambridge University Press

deer population (for us). This collapse is supposed to be facilitated by the role of the representation user.

However, I think that van Fraassen misemphasizes what it is that makes results relevant and that consequently his view is unnecessarily restricted. His view does not highlight the ways in which data collected in one context can be relevant in another. I agree with him that a data model is relevant to constraining a particular theory in virtue of the manner in which it was constructed, that is, the manner of data collection, processing, and analysis. However, insofar as these details can be made public, the data model is not relevant to the theory in question merely for me, but also for others who have access to that information. By sharing the information about how data have been gathered and processed, many scientists can assess the relevance of empirical results with respect to theories. Moreover, access to auxiliary information about data collection, processing, and analysis allows many agents not only to appreciate the relevance of data models so produced to the theory or theories for which the data were originally designed to test, but also in some cases to appreciate the relevance of the data to other theories beyond those targeted by the scientists who designed the observations and/or experiments in which the data were collected.

I suspect that van Fraassen would not be hostile to these points. And to be fair, my criticism of his account relies on a fairly strict reading of the passage quoted above (specifically of the phrase "for that purpose"). Nevertheless, it is the case that a data model can be relevant for adjudicating the empirical adequacy of a theory despite the fact that the model was originally constructed for a different purpose. In particular, once results are considered together with the auxiliary information about the manner of their production, it becomes possible to see how maladapted results could be reworked so as to become well adapted. With information about how a result was produced, one can sometimes backtrack through processing stages until one arrives at a result adaptable to one's purpose.

4. Concluding Remarks. I have argued that the characterization of evidence relevant to the adjudication of empirical adequacy is enriched evidence. Empirical adequacy is to be adjudicated with respect to all available data records and the empirical results generated from them considered together with all the available information about how the data were collected and processed. The notion of enriched evidence provides the resources to account for how scientists adhere to the minimal commitment of empiricism by doing due diligence to check the empirical adequacy of their theories. In other words, taking into account auxiliary information about data generation processes, it is no longer so mysterious how theories could be expected to be empirically adequate with respect to initially maladapted re-

https://doi.org/10.1086/697747 Published online by Cambridge University Press

sults and prima facie discordant results, or how there is a sense in which the same evidence can be used to constrain substantially different theories despite the intertwining of the theoretical and the empirical in scientific evidence. In fact, I hope to have shown how it is in fact not *despite* that intertwining but *in virtue of it* that these important epistemic activities are possible at all.

REFERENCES

Anderl, Sibylle. 2016. "Astronomy and Astrophysics." In *The Oxford Handbook of Philosophy of Science*, ed. Paul Humphreys, 652–70. Oxford: Oxford University Press.

Bogen, James, and James Woodward. 2005. "Evading the IRS." In *Idealization XII: Correcting the Model; Idealization and Abstraction in the Sciences*, Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 86, ed. Martin R. Jones and Nancy Cartwright, 233–67. Amsterdam/New York: Rodopi.

Chang, Hasok, and Grant Fisher. 2011. "What the Ravens Really Teach Us: The Intrinsic Contextuality of Evidence." In *Proceedings of the British Academy 171: Evidence, Inference and Enquiry*, ed. Philip Dawid, William Twining, and Dimitra Vasilaki, 345–70. Oxford: Oxford University Press.

Clark, David H., and F. Richard Stephenson. 1977. *The Historical Supernovae*. Oxford: Pergamon.

Colaço, David. 2018. "An Investigation of Scientific Phenomena." PhD diss., University of Pittsburgh.

Duhem, Pierre. 1954/1974. *The Aim and Structure of Physical Theory*. Repr. New York: Atheneum.

Franklin, Allan. 2002. *Selectivity and Discord*. Pittsburgh: University of Pittsburgh Press.

———. 2015. "The Theory-Ladenness of Experiment." *Journal for General Philosophy of Science* 46 (1): 155–66.

Galison, Peter. 1987. *How Experiments End*. Chicago: University of Chicago Press.

Green, David A. 2015. "Historical Supernova Explosions in Our Galaxy and Their Remnants." In *New Insights from Recent Studies in Historical Astronomy: Following in the Footsteps of F. Richard Stephenson; a Meeting to Honor F. Richard Stephenson on His 70th Birthday*, ed. Wayne Orchiston, David A. Green, and Richard Strom, 91–100. Cham: Springer.

Howlett, Peter, and Mary S. Morgan, eds. 2010. *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*. Cambridge: Cambridge University Press.

Hubble, Edwin. 1929. "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae." *Proceedings of the National Academy of Sciences of the United States of America* 15 (3): 168–73.

Jenni, Peter, et al. 2003. *ATLAS High-Level Trigger, Data-Acquisition and Controls*. Technical Design Report ATLAS. Geneva: CERN.

Krause, Elisabeth, et al. 2017. "Dark Energy Survey Year 1 Results: Multi-probe Methodology and Simulated Likelihood Analyses." http://arxiv.org/abs/1706.09359.

Kuhn, Thomas S. 1975. *The Structure of Scientific Revolutions*. 4th ed. Chicago: University of Chicago Press.

Laymon, Ronald. 1988. "The Michelson-Morley Experiment and the Appraisal of Theories." In *Scrutinizing Science: Empirical Studies of Scientific Change*, ed. Arthur Donovan, Larry Laudan, and Rachel Laudan, 245–66. Baltimore and London: Kluwer Academic.

Leonelli, Sabina. 2009. "On the Locality of Data and Claims about Phenomena." *Philosophy of Science* 76 (5): 737–49.

———. 2013. "Integrating Data to Acquire New Knowledge: Three Modes of Integration in Plant Science." *Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 503–14.

———. 2014. "Data Interpretation in the Digital Age." *Perspectives on Science* 22 (3): 397–417.

———. 2015. "What Counts as Scientific Data? A Relational Framework." *Philosophy of Science* 82:810–21.

https://doi.org/10.1086/697747 Published online by Cambridge University Press

———. 2016. *Data-Centric Biology: A Philosophical Case Study*. Chicago: University of Chicago Press.

Lipton, Peter. 2001. "History of Empiricism." In *International Encyclopedia of the Social and Behavioral Sciences*, ed. Neil J. Smelser and Paul B. Baltes, 4481–85. New York: Elsevier.

Matthiessen, Dana. 2018. "The Role of Local Knowledge in Mobilizing Data." Unpublished manuscript, University of Pittsburgh.

McDougall, Ian, and T. Mark Harrison. 1999. *Geochronology and Thermochronology by the 40Ar/39Ar Method*. 2nd ed. Oxford: Oxford University Press.

Miller, Michael. 2016. "Mathematical Structure and Empirical Content." http://philsci-archive.pitt.edu/12678/.

Palmer, Carole L., Nicholas M. Weber, and Melissa H. Cragin. 2012. "The Analytic Potential of Scientific Data: Understanding Re-use Value." *Proceedings of the Association for Information Science and Technology* 48 (1): 1–10.

Perović, Slobodan. 2017. "Experimenter's Regress Argument, Empiricism, and the Calibration of the Large Hadron Collider." *Synthese* 194 (2): 313–32.

Planck Collaboration. 2016. "Planck 2015 Results XIII: Cosmological Parameters." *Astronomy and Astrophysics* 594 (A13): 1–63.

Quine, Willard Van Orman. 1951. "Main Trends in Recent Philosophy: Two Dogmas of Empiricism." *Philosophical Review* 60 (1): 20–43.

Railton, Peter. 1981. "Probability, Explanation, and Information." *Synthese* 48 (2): 233–56.

Stephenson, F. Richard, and David A. Green. 2002. *Historical Supernovae and Their Remnants*. Oxford: Oxford University Press.

van Fraassen, Bas C. 1984. "Theory Comparison and Relevant Evidence." In *Minnesota Studies in the Philosophy of Science: Testing Scientific Theories*, ed. John Earman, 27–42. Minneapolis: University of Minnesota Press.

———. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford: Clarendon.

Woodward, James F. 2011. "Data and Phenomena: A Restatement and Defense." *Synthese* 182:165–79.

https://doi.org/10.1086/697747 Published online by Cambridge University Press