SCIENTIFIC PROGRESS AT THE BOUNDARIES OF EXPERIENCE

by

Nora Mills Boyd

B.Sc. in Physics and Philosophy, University of British Columbia, 2008

M.A. in Philosophy, University of Waterloo, 2010

Submitted to the Graduate Faculty of
the Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Nora Mills Boyd

It was defended on

March 13th 2018

and approved by

John D. Norton, History and Philosophy of Science

Robert Batterman, Philosophy

Christopher Smeenk, Philosophy at Western University

James Woodward, History and Philosophy of Science

Dissertation Director: John D. Norton, History and Philosophy of Science

SCIENTIFIC PROGRESS AT THE BOUNDARIES OF EXPERIENCE

Nora Mills Boyd, PhD

University of Pittsburgh, 2018

My dissertation introduces a new empiricist philosophy of science built on a novel characterization of empirical evidence and an analysis of empirical adequacy appropriate to it. I analyze historical and contemporary cases primarily, though not exclusively, from the space sciences attending carefully to the intricate practices involved in data collection and processing. I argue that the epistemic utility of empirical results as constraints on theorizing depends on the conditions of their provenance and that therefore information about those conditions ought to be included in our conception of empirical evidence. I articulate the conditions requisite for adjudicating the empirical adequacy of a theory with respect to some evidence and argue that much more background information is required for this adjudication than has been widely appreciated. Although my account is strictly anti-realist, this project is a defense of a sense of epistemic progress in science. Empirical evidence, as I have defined it, genuinely accumulates over the history of human inquiry. We learn that whatever theoretical framework we propose for understanding what the world is like will have to be consistent with this growing evidential corpus.

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

I owe huge debt of gratitude to my advisor John Norton, who expertly shepherded me through the entire process of graduate school. John, thank you for being an incredible mentor and especially for always providing instantaneous constructive feedback that pushed me to do my best work and to find a bold philosophical stance that I really care about. I am grateful also to three Jims (Woodward, Bogen, and Weatherall) for their support and for encouraging me to think in directions that I am sure I will continue to grapple with for a long time to come. I could not possibly thank David Colaço and Aaron Novick with sufficient profusion for reading countless drafts over the past six years, for their criticisms and advice, and for being my intellectual brothers. I am also immensely grateful to fellow travelers Michael Miller, Siska De Baerdemaeker, Katie Creel, Dana Matthiessen and Liam Kofi Bright for their friendship, dialog, and for being there when I needed them most.

I also want to acknowledge several people who have contributed to my education over the years without whom I would not be where I am today. Thank you to Matthew Capdevielle, a friend and mentor who taught me both physics and philosophy in high school and to Alan Richardson, my undergraduate philosophy of science professor who provided much needed encouragement and support in pursuing my higher education in philosophy. Much gratitude also to Derek Storm, Doug Will, and Greg Harper at the nuclear physics lab under whose guidance my love of experimental physics was cemented. Doug and Greg, you are family, thank you for being there fore me. Finally, I want to acknowledge my master's degree advisor Doreen Fraser who rocketed me into the Pittsburgh HPS scene.

I have been extremely privileged to have the benefit of substantive support from my family. Both of my parents have been role models for me throughout my education. Watching my mother Beth Mills earn her PhD while I was in middle school was a very formative

experience for me. I have always been inspired by her as a teacher and a scholar and I am so grateful for her help and confidence from the beginning. Among innumerable other things, I owe my dad Andrew Boyd immense gratitude for introducing me to what is wonderful, awesome, and puzzling about physics and the cosmos from an early age. Dad, you are my favorite model of the life-long learner. A thousand thanks both to my dad and to my stepmother Cristal Weber for sending me to university and for always supporting me in my education. Thank you also to my grandfather Robert Boyd, for fueling my love of science with stories.

Finally, I want to extend my deep gratitude to Zander Winther. Zander, thank you for introducing me to some of the best philosophy, for moving with me multiple times for school and for adventure, and for your wisdom.

1.0 INTRODUCTION: EPISTEMIC PROGRESS IN SCIENCE

Philosophers of science who engage with the problem of scientific progress have (at least since Kuhn) looked to the history of science for their philosophical ore. This historically-oriented philosophy of science has often focused on apparent discontinuities in the scientific record marked by revolutions in scientific theory. Philosophy of scientific progress has been primarily concerned with giving accounts of the familiar trajectory beginning with Ptolemy, passing through Copernicus, Kepler, Galileo, Newton and ending with Einstein, with the trajectory from alchemy through the chemical revolution, and so on...in short: understanding how mature scientific fields emerged from a graveyard of diverse worldviews.

As a consequence, some philosophers have adopted non-epistemic accounts of scientific progress (what Bird (2007) calls "functional-internalist" accounts). For instance, those who understand the history of science as a series of paradigms and revolutions have a difficult time reconciling their view with the plausibility of cumulative epistemic progress. Notoriously, Kuhn (1975) resorted to a kind of pragmatic understanding of scientific progress as progress *in problem-solving capabilities* rather than increased knowledge about the natural world. According to Kuhn, scientists adopt a new paradigm when the growing pile of unsolved anomalies besetting their old paradigm becomes unbearable; "[t]he scientific community is a supremely efficient instrument for maximizing the number and precision of the problems solved through paradigm change" (1975, 168).

Both Larry Laudan and Dudley Shapere furnished accounts of scientific change in the wake of the reception of Kuhn's work. Both philosophers were concerned to construct philosophical accounts that did not fall prey to the difficulties that beset logical empiricism and the sort of relativism that Kuhn's philosophy inspired. Laudan (1977, 1996) concerned himself with the interplay between rationality and progress in science. He too understands scientific progress in terms of problem-solving effectiveness and he eschews the characterization of progress as cumulative. On Laudan's view, scientific progress is evaluated according to *our* standards, which are not the necessarily the same standards of the historical figures of science (1996, 138). Similarly, Shapere (1984) argued that science "involves no unalterable assumptions whatever, whether in the form of substantive beliefs, methods, rules, or concepts" (238). In other words, for both philosophers, the 'rules' of scientific methodology change with historical context.

Kukla (1990) presented a ten-element taxonomy, which modified and expanded Laudan's classification of scientific progress. His primary criticism of Laudan's scheme was that it did not allow for the fact that the *creation* of empirical problems can constitute advances in science (Kukla, 1990, 459). Kukla describes six different types of scientific activities related to theory development:

> (1) inventions that increase the scope of a theory (which I will call *constructions*[)], (2) discoveries that increase the scope of a theory (*expansions*), (3) discoveries that decrease the scope of a theory (*contractions*), (4) inventions that increase the probability of a theory (*simplifications*), (5) discoveries that increase the probability of a theory (*amplifications*), and (6) discoveries that decrease the probability of a theory (*degradations*) (462).

All of these activities are "theoretical", according to Kukla, in that "they can be conducted without breaking the contact between armchair and backside" (463). He names four further activities (*empirical amplification, empirical degradation, empirical expansion, and empirical contraction*) that are "more athletic" (Kukla, 1990, 464). Interestingly, Kukla describes all four of these "empirical" activities as "guided by theoretical concerns" (ibid.). However, he does add one further category, which he does not include in the ten-element taxonomy of progress:

> There is also a relatively atheoretical type of research exemplified by the activities of natural historians. This kind of research does not attempt to confirm or disconfirm a theoretical point of view, nor does it seem to affect the explanatory scope of any theory. Natural historians collect data which strike them as intuitively important or interesting...for the sake of taxonomic elegance, I prefer to treat it as a limiting case of empirical expansion (Kukla, 1990, 464).

This taxonomy is promising in the sense that it aims to give an exhaustive account of the modes of scientific activity that contribute to progress. Furthermore, although Kukla down-plays the role of the natural historian-like activities involved in scientific reasarch, he is right to identify them. In fact, we will find that such activities are especially important in progressing frontier science. However, in weighting his description of scientific activity so heavily in the direction of "theoretical concerns", like Laudan, Kukla has failed to provide an account of cumulative epistemic progress transcending the framework of particular theories in which scientific inquiry has increased our knowledge about the natural world. Indeed, Mizrahi and Buckwalter (2014) note that "few contemporary philosophers of science directly associate scientific progress with knowledge" (152).[1] In particular, several important questions remain unanswered. I contend that any account of epistemic progress in natural science should provide answers to the following questions:

1. What is the nature of the scientific knowledge that we actually have?
2. How has such knowledge improved over time?
3. What sorts of improvements may reasonably be anticipated in the future?

On my view, scientific progress importantly involves constraining viable theoretical interpretations by increasing empirical access. We gain greater empirical access to the natural world by way of further observation, detection and experiment. This access yields an accumulating evidential corpus that in turn constrains the landscape of viable interpretive approaches by which we attempt to understand what the available evidence means for how the world is. Scientists do not just become better problem-solvers, they really do learn more about the natural world. Certain alternatives are not consistent with the available evidential corpus—and that counts as knowledge about the world.

The view of the epistemology of science that I defend in this dissertation addresses a particular lacuna in the philosophy of science literature surrounding the role of exploratory *observations* in scientific progress. Following Hacking (1983), interventionalists have argued that it is the scientist's intervention on physical systems that grounds increasing knowledge about those systems. Plausible as this idea may be, if it is to apply to the observational or

[1] There has been a recent exchange in the philosophical literature on the topic of scientific progress in which the connection to knowledge has received some attention, including contributions from Bird (2007, 2008); Mizrahi (2013); Mizrahi and Buckwalter (2014); Niiniluoto (2014); Rowbottom (2008, 2015). This exchange has been primarily concerned with exposing intuitions about scientific progress by trading thought experiments. However, it is far from clear how the philosophical approach of considering such hypothetical vignettes relates to questions about progress in actual science.

historical sciences, the intervention in question apparently has to be performed on a model (perhaps via operations on a mathematical representation or a computer simulation) of the system, or by counterfactual reasoning within a particular theoretical framework (Woodward, 2004). However, I believe that there are reasons to think that this suggestion does not yield the epistemic surrogate of the physical manipulation case. In order to understand how epistemic progress can nevertheless be made in such domains, my project focuses heavily (although not exclusively) on examples from astrophysics and cosmology—two paradigmatically observational sciences. By attending to these examples I am in a position to supply a general account of scientific progress, which might be obscured if one simply attended to progress made in fields of experimental science where manipulation and intervention are possible.

Any story of scientific inquiry has to get us from point A to point B and tell us what happens in the middle. Point A is the nature side of the story, point B is the theory side, and somehow or other these two need to be connected in the right sort of way. This dissertation is structured such that, taken together, the chapters trace an arc from the nature side to the theory side. Chapter 2 deals with the production of empirical data via interaction with the natural world and Chapter 3 explicates the path between data records and constraints on theorizing. Chapter 4 is a bit of an intermezzo, which discusses how the details of data collection and processing end up mattering for adjudicating empirical adequacy and some of the consequences this has for data stewardship. Chapter 5 picks the arc back up and carries it forward to the point of contact between evidence and theory: empirical constraints. Finally, in Chapter 6 I discuss how this story of scientific inquiry should be reflected in our epistemic commitments.

2.0 THE MINIMAL COMMITMENT OF EMPIRICISM

Empiricism has a curious status in the philosophy of science. It is at once part of the ordinary and seemingly uncontroversial conceptual currency of the field, and at the same time slippery and equivocal. In this chapter I articulate and defend a characterization of empiricism.

2.1 VARIETIES OF EMPIRICISM

The history of philosophy exhibits a variety of empiricisms.[1] One can find diverse and nuanced empiricisms for instance in the work of Epicurus, David Hume, up through the logical empiricist movement and Bas van Fraassen's constructive empiricism in the 20th century. There are however, a relatively few central themes woven throughout. We can identify three important empiricism theses, which can be divided into the theses constitutive of what we might call Garden-variety Empiricism[2] and a further thesis that, when combined with the first two, yields what we might call Full-bore Empiricism. I will state the theses first and then discuss them.

Garden-variety Empiricism Knowledge of nature derives from, and only from, experience.

- **(Necessity of experience)** To learn about the natural world, we must submit our conceptions to the tribunal of experience.

[1] Lipton, Peter. (2015) 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" [567].

[2] This corresponds to what Lipton, Peter. (2015) calls "Knowledge Empiricism", see p. 569.

- **(Opposition to rationalism)** We do not gain knowledge about nature by contemplation in absence of experience.

Full-bore Empiricism

- **(Opposition to realism)** We ought not believe that our best ways of conceiving of the world are true, but only that they cohere with our experience.

Garden-variety Empiricism is not as controversial a position as Full-bore Empiricism. Realism about scientific theories is compatible with Garden-variety Empiricism. To get to Full-bore Empiricism, one needs an additional premise. In particular, one needs to be committed to the following:

> **(Appearance/world gap)** Experience is mediated and never gives us access to nature *directly* but always via 'appearances' that are contingent on our circumstances in the world, e.g. our limited and particular faculties, scale, and the part of the world that we occupy in space and time.

If this premise is true we will never be in an epistemic position that would warrant belief that our theories are true, where truth is construed as correspondence. In other words, if one is committed to the position that knowledge of nature derives from, and only from experience (Garden-variety Empiricism) and in addition to a separation of the way nature is independently of our experience of it and that very experience, then it will follow that the best we can get is belief that our theories agree with our experience.

One could take this line of reasoning as grounds for pursuing a theory of truth besides correspondence, such as a coherence or pragmatic theory of truth. However, a correspondence theory of truth is apt for thinking about science that aims at understanding what the natural world is like as opposed to generating conceptual frameworks that successfully 'hang together' or 'work'. That a correspondence theory is the appropriate theory of truth in this context and that we will never be in a position to judge that our theories are in fact true in this sense are not inconsistent with one another. Moreover, proposing candidates for correspondence can be fruitful for amassing empirical constraints on any viable framework (and for generating frameworks that we choose to work within, while nevertheless abstaining from making judgments about their truth).

Another option would be to reject that there is anything beyond/underneath the appearances and take the appearances themselves to be in a sense all that is really there.[3] This latter move is not exactly to embrace idealism, or skepticism, but rather to insist that the appearances *are* the real. It is not entirely clear that such a view can be made coherent, and if so that it would be a desirable way to construe what is going on in scientific inquiry. Be this as it may, Garden-variety Empiricists who embrace the **appearance/world gap** and are interested in truth as correspondence with the world beyond its appearances will have to relinquish the idea that we are ever in a position to judge that we have obtained such truths. In other words, they will be lead to Full-bore Empiricism.

Let us focus on explicating the **necessity of experience** thesis of Garden-variety Empiricism. What does it mean that "we must submit our conceptions to the tribunal of experience"? Something like this thesis is, I think, a central component of the sort of empiricism that philosophers of science broadly want to endorse. Nevertheless there is a lot more that needs to be said in order to explicate this thesis.

2.2 EXPLICATING THE 'TRIBUNAL OF EXPERIENCE'

The term 'experience' connotes something at once familiar and vague. My experience of x encompasses my interactions with x from my perspective–the appearance of x in the guises in which I encounter it, for me. Experience construed broadly in the context of empiricism is something like the totality of appearances presented to relevant epistemic agents. To say this is not yet very helpful, especially when thinking about sophisticated contemporary science. What for instance is the relevant sense of 'appearances'? In the context of contemporary scientific practices it is useful to explicate the **necessity of experience** thesis by transitioning to speaking rather about empirical 'evidence'. One important reason to do this is to signal that the relevant sense of 'experience' goes beyond unaided human perception to include the mediated experience of the world that is gained through sophisticated instrumentation and technique.

[3] Nietzche does not advocate exactly this, but something nearby in his well-known aphorism 107 concluding Book Two of The Gay Science (2001, 104-105).

Another reason to speak of empirical evidence rather than experience is that the latter has a much broader connotation than the former. My experience broadly construed includes my experience of my dreams, hallucinations, imaginings, desires and so on. We ought to be very careful about how these aspects of experience are supposed to inform a properly empiricist position. They can in some contexts be understood as sources of empirical evidence (perhaps with respect to for example our conception of the functioning of the human mind or brain). In these cases in particular, thinking about evidence rather than experience is helpful because it makes plain the sense in which the epistemic utility of even our most immediate experiences (like desires) can be mediated.

Indeed, it has been a persistent mistake of empiricist positions to insist on a foundational role for a kind of *direct* experience (an idea that traces at least all the way back to Epicurus). Nothing important for Full-bore Empiricism is lost when we take the fuel that feeds our increasing knowledge of what the world is like to be mediated evidence rather than direct experience. This is because a Full-bore Empiricist already believes that there is a gap between our experience of the world and what the world is really like independently of our experience. Such an empiricist is already committed to thorough-going mediation of experience. For such an empiricist, giving up direct experience does not amount to giving up the important sense in which the **necessity of experience** thesis encodes the conduit through which world is to 'push back' on our understanding of it–namely our interaction with the world. This is because evidence is gathered by interaction with the world, mediated as that interaction may be.

Transitioning to speaking of evidence rather than experience may be more problematic for Garden-variety Empiricists who want to reject Full-bore Empiricism. Perhaps Garden-variety Empiricists with realist aspirations hope that the access to the world given by direct experience could ground our acceptance of our best scientific theories as *true*. There is an onus on such aspiring realists to defend the claim that we have direct experience of the world in the first place. In absence of such defense, we are stuck with Full-bore Empiricism.

How might the **necessity of experience** thesis be explicated in terms of empirical evidence? Consider the following: *either cherry-picking evidence is permissible or it is not.* Suppose cherry-picking is permissible. Then agents are free to engineer the adequacy of their theories as they please. Inconvenient evidence can simply be ignored. If anomalous results turn up in the laboratory, one need not worry, these can just be summarily swept under the rug.

To adopt the permissibility of cherry-picking evidence would clearly be antithetical to the enterprise of empirical science. It would make the viability of a theory a matter of pure whim and would sever the connection to nature that was supposed to be instantiated in the **necessity of experience** thesis. Perhaps evidence can be set aside (although we should be very careful to say what this means), but it ought not be set aside without good reason. That the evidence is inconvenient is certainly not a good reason, and there may be other bad reasons besides. Cherry-picking evidence is incompatible with Garden-variety Empiricism and with Full-bore Empiricism, this much is clear. But what are we committed to if cherry-picking is not permissible? What 'tribunal' precisely are we committed to saying that our theories must face?

I will advocate for the following explication of the **necessity of experience** thesis:

> **The Minimal Commitment of Empiricism** Good theories, whatever else they are, are empirically adequate: they are consistent with all of the available empirical evidence.

Supporting the Minimal Commitment of Empiricism The Minimal Commitment of Empiricism itself calls for explication, and the explication I will offer calls for defense. Let us first note that one might reasonably argue that much else besides the Minimal Commitment should be required of good scientific theories. However, to require any less of them would open the door to theories inconsistent with the evidence. But that would be to give up on the core distinguishing feature of science in comparison to other human pursuits. The impermissibility of cherry-picking evidence is encoded in the prescription that empirical adequacy is to obtain with respect to *all* of the available empirical evidence. To have any hope of learning about nature from experience at all, the world has to be able to push back on our understanding of it and the 'pushing' happens through the mechanism of maintaining empirical adequacy.

Demanding consistency with all of the empirical evidence expresses the heuristic that the more empirical constraints the better. This heuristic accords well with actual scientific practice. If new observational or experimental evidence can be garnered, scientists generally consider it epistemically desirable to do so. This is evidenced not only in the fact that scientists are perpetually appealing to funding agencies to build next generation experiments, but also in the fact that scientists opportunistically gather data when convenient circumstances present themselves. Similarly, if old evidence can be brought to bear on contemporary theorizing, that too is considered epistemically desirable.

Why should this be so? That is, why should having more empirical constraints be desirable? One plausible reason is that adopting the heuristic that more constraints are better serves as a mechanism for systematically dispensing with potential falsifying evidence. Unchecked evidence is an epistemic liability in the sense that it could turn out to be falsifying. Thus, one would like to examine the evidence if possible to remove the liability. Another reason is that empirical adequacy is cheap when there are few constraints with which to contend. That is, a theory from which no empirical consequences have been derived is empirically adequate in the cheapest sense. Similarly, a theory in a world where no empirical constraints have been generated, or all empirical constraints have been destroyed, is also cheaply adequate. In contrast, we can learn more about what the world is like from a theory that is consistent with many empirical constraints. Empirical constraints are the conduit through which we learn what the world is like; we learn that any viable theoretical framework must be consistent with the empirical evidence. Therefore the more evidence we have, the more we have learned about what the world is like.

Note that the way I have formulated the Minimal Commitment here is not how van Fraassen would do so. For van Fraassen, empirical adequacy is consistency with all of the evidence that there ever could be, whether anyone ever actually gets a hold of it in practice or not (see van Fraassen, 1980, 12-19). Thus, van Fraassen's take on empirical adequacy relies on the notion of what is observ*able* rather than what is as a matter of fact observed by any particular epistemic agent or agents. I have chosen not to follow van Fraassen on this point but to instead characterize empirical adequacy with respect to the *available* evidence. This has the consequence that an epistemic community could judge a theory to be empirically adequate with respect to the evidence available at some initial time and then come to judge that the same theory fails to be adequate with respect to the evidence available at some later time. In contrast, if a theory is empirically adequate in van Fraassen's sense at some time, it will always be so since the corpus of evidence on which such adequacy depends is static–it does not depend on the actual status of scientific research. I take this difference to count in favor of my formulation of the Minimal Commitment of Empiricism over that implicated by van Fraassen's view precisely because no one will ever actually be in a position to judge the adequacy of a theory with respect to all of the evidence that is observable in van Fraassen's sense. A notion of empirical adequacy relevant to science in practice should pick out something that real scientists are in fact in a position to adjudicate–it should be relevant to real epistemic agents.

We should note that adhering to the Minimal Commitment does not commit one to naïve falsificationism. In particular, that good theories need to be consistent with the available evidence does not mean that whenever a theory encounters anomalous evidence that it should be abandoned without further regard. The Minimal Commitment leaves open the possibility that 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 some relevant evidence, something has eventually got to give.

We should also note that not all scientific theories need be subject to the Minimal Commitment. Scientists sometimes investigate theories primarily for their intrinsic interest or instrumental value.[4] With this in mind, what makes a theory "good" in the farthest reaches of mathematical physics on one hand, or in applied and synthetic sciences on the other, need not necessarily involve empirical adequacy. The Minimal Commitment applies only to theories that are supposed to be theories of our actual world.[5]

[4]I have in mind a very broad sense of "theory" here, encompassing formal entities like sets of axioms and the models that satisfy them, as well as hypotheses expressed as propositions and even imprecise mixtures of mathematics and narrative components.

[5]Attending to the appropriate scope of the Minimal Commitment dispels the apparent tension between the present work and that of Bhakthavatsalam and Cartwright (2017), when they write "to *mandate* empirical adequacy as a minimum criterion for a scientific theory is entirely unreasonable and just wrong" [6, original emphasis]. Those authors are particularly interested in theories scientists use for "managing the world" [5]. Indeed, they state explicitly: "we have no quarrel with empirical adequacy as an indicator of theory acceptability when acceptability is to be judged in terms of truth: a theory with false implications cannot be true, whether its implications are about empirical phenomena or something else. But there are lots of other things one can intend by labeling a theory 'acceptable'" [3].

2.3 RE-CASTING FULL-BORE EMPIRICISM

With the explication of the **necessity of experience** thesis provided in the previous section in hand, we can return to the three empiricist theses introduced in the first section and see how they may be recast in light of it. I offer the following interpretation of the theses:

Garden-variety Empiricism Knowledge of nature derives from, and only from, experience. *That is, our knowledge about what the world is like derives from, and only from, empirical evidence.*

- **(Necessity of experience)** To learn about the natural world, we must submit our conceptions to the tribunal of experience. *That is, good theories, whatever else they are, are empirically adequate: they are consistent with all of the available empirical evidence (Minimal Commitment of Empiricism).*
- **(Opposition to rationalism)** We do not gain knowledge about nature by contemplation in absence of experience. *That is, we do not gain knowledge about nature by contemplation in absence of empirical evidence.*

[**(Appearance/world gap)** Experience is mediated and never gives us access to nature *directly* but always via 'appearances' that are contingent on our circumstances in the world, e.g. our limited and particular faculties, scale, and the part of the world that we occupy in space and time. *That is, empirical evidence is mediated and never gives us us access to nature* directly *but rather through results that are contingent on our circumstances in the world, e.g. our limited and particular faculties, instruments, techniques, scale, and the part of the world that we occupy in space and time.*]

Full-bore Empiricism

- **(Opposition to realism)** We ought not believe that our best ways of conceiving of the world are true, but only that they cohere with our experience. *That is, we ought not believe that our best theories are true, but only that they are consistent with the available evidence.*

Recasting the theses of empiricism in this manner pushes the task of their explication to the problem of specifying what the *empirical evidence* is with respect to which scientific theories should be adequate. Providing a characterization of empirical evidence appropriate for this role is the subject of chapter 3. First, however, we ought to inquire into what makes something distinctively *empirical* in the first place.

2.4 WHAT IS DISTINCTIVELY 'EMPIRICAL'?

An empiricist position ought to be able to distinguish the *empirical* from, for instance, the *virtual* if it is to remain faithful to the **Minimal Commitment of Empiricism**. The tribunal of experience ought to originate from the world lest empiricism lose what makes it distinctive from rationalism. In certain contrived circumstances it will be easy to discern the empirical from the non-empirical. Suppose a scientist forgets to record an entry in the laboratory notebook and to fill in the blank just makes up a number to record where the datum should be. For most epistemic contexts, this entry in the laboratory notebook would not be empirical in the relevant sense. Indeed, such a scientist could plausibly be accused of fraud. The provenance of the datum in question is the imagination of the scientist, not the worldly target of research. The exception of course would be if the worldly target happens to in fact be that scientist's imagination.

Consider another relatively straightforward case. Suppose a scientist designs a computer simulation in order to explore how the characteristics of a model respond to a variety inputs. Such a scientist could make records of the outputs of the simulation, and we might want to call these data, but we would not want to call them empirical data since their provenance is the simulation rather than the world. Again, there are special circumstances in which such data might be considered empirical data—as when the worldly target of research is the computational system itself rather that whatever the model was suppose to represent in the first place. In general though, we do well to distinguish between the *virtual* data produced from simulations and *empirical* data got by interaction in some manner with a worldly target of study.

There are, however, many circumstances for which it is not quite so straightforward to determine what is empirical and what is not. Two kinds cases are of particular interest. The first kind are cases in which the data is clearly empirical, but there is some question about how the worldly source of their provenance is related to the worldly target of interest. Consider an example that we will return to below: data collected from terrestrial fusion experiments that is purportedly relevant to the physical processes involved in distant supernovae.

The second kind of case that we ought to consider more carefully involves data that are sourced from the worldly target of interest, but for which the data collection process involves the outputs of simulations. An example of this second sort of case, which has been discussed extensively by Morrison (cf. 2009, 2015), is high energy particle physics data collected at the the Large Hadron Collider.

In the first kind of case, at first glance, the data appear to be empirical but in the wrong sort of way. But what precisely has gone wrong? In the second kind of case, the data appear to be hybrid in nature—partly empirical, partly simulated. Ought such hybrid data count as empirical in the relevant sense?

In this chapter, I will argue that *data are empirical with respect to target T when there is an interpretation of the provenance of those data using the resources of epistemic context C such that the data are products of causal interaction with T*. Thus, data can only be to judged to be empirical in a relational way. Data are empirical with respect to a target and an epistemic context. As will be readily apparent, this has the consequence that the very same data record can be empirical with respect to some target and context and simultaneously not empirical with respect to some other target and context. Explicating the *empirical* in this way will allow us to pronounce on the two kinds of troublesome cases introduced above.

This way of characterizing the distinctively empirical is not without precedent in the philosophy of science literature on measurement.[6] For instance, van Fraassen (2012) articulates the context sensitivity of the empirical nature of data as follows: "Whether a procedure counts as a measurement and, if so, what it measures are questions that have, in general, answers only relative to a theory" (774). Likewise, Parker (2017) emphasizes the causal production of data when she writes: "Measuring is an activity that involves, among other things, physical interaction with the system being measured" (285).

2.4.1 Data are empirical relative to a target and a context

Data are empirical relative to a target. Without specifying a target it is impossible to say whether some particular data are empirical or not. That this must be so is easy to see. Suppose a colleague hands you a table of values indicating the height of all of the high tides

[6] See also Tal (2013).

on a certain beach over the last year as she measured them. You wonder "are these *empirical* data?" The answer depends on the target we consider. When considered with respect to the tides, the data are indeed empirical. However, when considered with respect to any number of other targets—the stock market, for instance—these data are not empirical.

One might be tempted to say that if a data set is empirical with respect to *some* target, then that data set is empirical full stop. This would be ill-advised however, because it would be so permissive as to invite unnecessary confusion. For instance, it would then be permissible to say that simulation outputs, dream diaries, and characteristics of certain works of art are empirical data since there are indeed some targets with respect to which each of these records are empirical data. As mentioned above, simulation outputs could be empirical data with respect to the state of the computational system that produced them. Likewise, dream diaries could plausibly contain empirical data with respect to the sleep processes of the person who's dreams they record, and records of the characteristics of a works of art might plausibly constitute empirical data with respect to the evolution of the artist's style and technique. These are special circumstances. In general we would not want to say that simulation outputs, dream diaries, or the characteristics of artworks are empirical evidence because to do so would invite confusion. It would invite the mistake of construing these sort of records as empirical data with respect to other worldly targets.

Data are empirical relative to an epistemic context and the epistemic context supplies the resources with which the data are interpreted. Data never speaks for itself, but rather always requires interpretive resources. If a colleague hands you a table of values, without further context you will not be in a position to say whether those values are empirical data or not. You will need the resources of an epistemic context to help you interpret the what the values are values *of*, what their significance is, and to discern for what purposes the data may be used and whether and what further processing may be required for those purposes. In particular, you will need enough background theory to furnish a causal story connecting the worldly target of interest to the data collection and recording process. Insofar as many data collection processes involve detectors, measuring apparatuses, or other technical apparatuses, the resources in the epistemic context required to tell whether the data is empirical or not will likely include both theories and/or models of the target system, the technical equipment involved in data collection, and the intervening causal processes.

An epistemic context, in the sense I intend here, is just the collection of conceptual, theoretical and representational resources from the perspective of which the data is to be interpreted. It would be simpler to call the epistemic context "a theory", as van Fraassen does when he writes in the passage already quoted, that whether a procedure counts as a measurement has an answer "only relative to a theory". However, as this obscures the role that models, hypotheses, assumptions, definitions, and other such resources play in interpreting data I prefer to speak of "epistemic contexts", which I take to be inclusive enough to encompass these other resources that we might not want to call theories. This distinction allows those who so desire to reserve a more restrictive and formal definition for "scientific theory", as for example a set of axioms and the mathematical models that satisfy them.

The epistemic context need not be the original context of production. Data can be used for the specific purpose for which they were gathered and data can sometimes be used for purposes for which they were not initially intended. To take a simple example, data records of barometer readings could be used (as initially intended) to generate constraints on theorizing about atmospheric pressure and (opportunistically) to generate constraints on theorizing about the altitude from which the readings were collected using further background information about the relation between barometric pressure and altitude in the relevant context.

Since whether data are empirical or not depends on the epistemic context from which they are interpreted, it is possible to have data that were initially interpreted as empirical with respect to some target using the resources available at the time of their production, but that are later interpreted as *not* empirical with respect to that target using the resources of some subsequent epistemic context. For instance, the OPERA data that was initially used to support the existence of superluminal neutrinos, was subsequently reinterpreted as the epistemic context shifted in such a way that no longer afforded the right sort of causal connection between neutrinos and the data in question due to a faulty connection in the experimental apparatus.

Once data has been interpreted as empirical with respect to some target is it ever possible to reinterpret it in light of a subsequent epistemic context such that it ceases to be interpreted as empirical at all? Insofar as the data have been produced by *some* causal process, one could hope that there would always be some target with respect to which the data could be interpreted as empirical. Unfortunately, it could be the case that none of the available epistemic contexts can furnish a substantive enough causal story of the production of the data. That is, it could be that the causal processes involved in the production of the data in question are so poorly characterized by any available epistemic context, that no one is in a position to say what causally generated the data.

Under an interpretation, the provenance of empirical data involves causal interaction with the target. An important feature of the view of empirical data that I am defending is the causal production of data. To be properly empirical, data should have been produced by causal processes that connect the worldly target of research to the process of data collection and recording *from the perspective of the epistemic context in which the data are to be interpreted.* There is no perspective outside of an epistemic context from which the causal processes can be identified and traced. Indeed, there is no perspective outside of an epistemic context from which a worldly target can be identified in the first place. Yet, using the resources of an epistemic context, it can be possible to answer the question: were these data produced by causal interaction with the target?

Why is it important for their empirical character that data be connected by causal processes to worldly targets? One important reason in support of this approach is that if one does not require the causal production of data, one risks collapsing the distinction between empirical data and virtual data. We should like to be able to say that the outputs of a simulation of X are *of X* without having been causally produced by X. The products of models and simulations can stand relations to worldly targets, for instance representation relations, without having been produced by those targets.

To see why causal production matters, let us consider a characterization of what makes data empirical that eschews causal production and relies instead on ostension. I will suggest that this view suffers from a serious drawback—it is not easy to see how it could be applied to data sourced from far away worldly targets that no one can point to in an act of ostension.

The characterization I want to consider comes from an insightful and philosophically valuable paper by Matthias Kaiser (1991) titled "From Rocks to Graphs—The Shaping of Phenomena". In this paper, Kaiser develops an abstract characterization of the epistemological structure that carries scientific reasoning from what he calls "observable reality (data)" to "empirical phenomena" (121). As the title of the paper intimates, Kaiser traces out this structure for a concrete case study, from geological specimens gathered in the field (rocks), through data collected via the manipulation and transformation of these specimens in the laboratory (values of their magnetic declination and inclination), finally to claims about phenomena (continental drift). I will not review all of the details of Kaiser's epistemological picture. Suffice it to say that the basic elements of an instance of the structure are an *anchor-point* and *inference tickets*. Roughly, the anchor-point is the bit of the world that source the data, and the inference tickets transform the data by redescription and other operations licensed by theoretical resources (either via bootstrapping or the incorporation of accepted background theory), which might include incorporating data collected in other contexts (cf. ibid., 122-123). In Kaiser's geological case study, the anchor-point is just the "rocks *in situ* which can be pointed out by the paleomagnetist" (122).

For present purposes, what matters is that for Kaiser, the anchor-points are material objects picked out by ostension in what he calls "experiential anchoring" (125). He writes: "all data structures are 'anchored' in objects of immediate experience. One keeps track of these objects in the sense that the material basis of the layer-set of data can always be recovered" (122). Although Kaiser goes on to formalize these intuitions—"anchor-point" becomes "0-structure" and so on—the gist of the idea is as follows:

> [The 0-structure] consists of the objects of the scientific inquiry together with the simple act of presentation. My basic intuition is here that scientific data rest on raw material that can be presented, or reproduced, upon request. It is essential that this level is kept as simple as possible, e.g., "this rock here", or "these spots on the screen there", or — in social science — "these filled-in questionnaires", etc. The 0-structure contains all those things that are to be subjected to scientific scrutiny, i.e., that are to be measured, weighed, radiated, dissoluted, accelerated, etc. (125)

Thus, on Kaiser's view, what makes data *empirical* are these original (and reproducible) acts of ostension, like "this rock here". As I hinted above, I believe that Kaiser's view is ill-suited to far away targets in virtue of this feature of his account. On this account, when the material specimen is not present at hand, the scientist cannot anchor the data to it via acts of ostension like “this rock here”. Instead, on this account she is forced to anchor her data, not to the worldly target of interest, but to some intermediary as in “these spots on the screen there”. Let us consider an example of this sort. Figure 1 shows a visualization of data from the Arecibo telescope, displaying “spots” that can be interpreted as radio signals from distant galaxies.

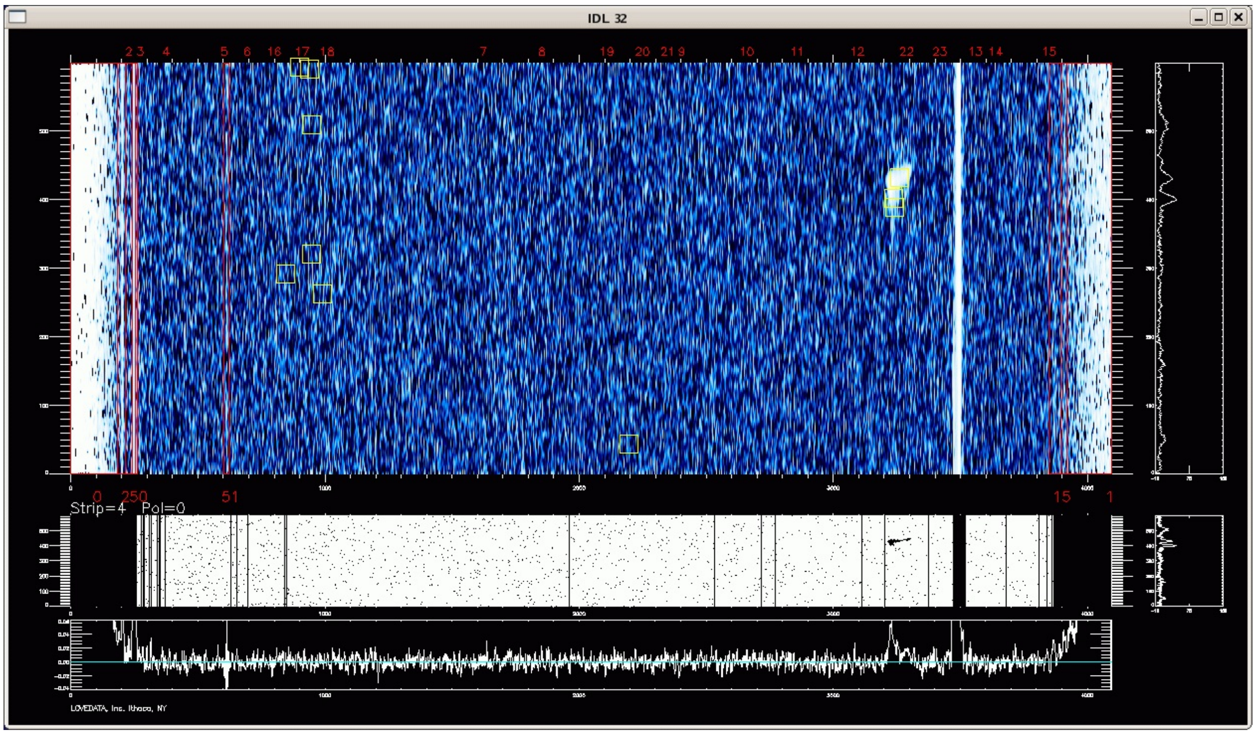

Figure 1: Data from the Arecibo radio telescope. Sourced from http://egg.astro.cornell.edu/alfalfa/ugradteam/hunt09/hunt2_09.htm. Thank you to Martha P. Haynes for permission to use this image.

On Kaiser’s account, what makes the data visualized in Figure 1 empirical is that they can be picked out by ostension, by some scientist gesturing to the image—“these spots on the screen there”. This approach locates the empirical at the site where some bit of the world is present to perception of a human inquirer. In this case, the spots in the image are what is present to the inquirer. But approach seems to miss the point. The data displayed in Figure 1 are empirical data with respect to galaxies—they are data encoding the radio frequency emissions of galaxies over time—in virtue of being produced by the interaction of electromagnetic radiation emitted by those galaxies traveling through space an interacting with the receiver of the Arecibo observatory telescope and subsequently the rest of the readout apparatus of that telescope. We should of course note that this causal interaction story is told using conceptual resources of a variety of theories that make up an epistemic context, including for this instance, theories regarding the composition and behavior of galaxies, electromagnetism, theories of electronics, and so on.

Moreover, Kaiser's approach has the consequence of making everything that can be picked out by ostension in this manner empirical. But this is just the problem that we have already discussed above, namely, it invites confusing virtual data (data generated by simulation or imagination) for empirical data.

In light of these difficulties, I suggest that we are better off giving up the impulse to find an experiential foundation for the empirical character of empirical data in the first place.

Let us consider another, more recent approach to defining data, that of Leonelli (2015). For Leonelli, data is roughly any product of scientific research that is packaged for dissemination for the purpose of serving as evidence. In particular, she defines data "as a relational category applied to research outputs that are taken, at specific moments of inquiry, to provide evidence for knowledge claims of interest to the researchers involved" (811). In particular, data is

> any product of research activities, ranging from artifacts such as photographs to symbols such as letters or numbers, which is collected, stored, and disseminated in order to be used as evidence for knowledge claims [...] what matters is that observations or measurements are collected with the expectation that they may be used as evidence for claims about the world in the future. Hence, any object can be considered as a datum as long as (1) it is treated as potential evidence for one or more claims about phenomena and (2) it is possible to circulate it among individuals. (817)

Note that Leonelli's definition of data is clearly relational—data is data in part in virtue of being potential evidence. Her view does not rely explicitly on an experiential foundation anchoring it to material objects present to the perception of a human inquirer. Yet, it is not clear that Leonelli's view sufficiently emphasizes the importance of the causal production of empirical data either. That data must be produced from physical interaction with the appropriate target system may be implicit in Leonelli's view. Plausibly (reasonable) scientists would not have the expectation that products of research activities could be used as evidence for claims about the world if they were not produced by physical interaction with the relevant parts of the world. But we do better to be explicit that causal production is a necessary condition for being empirical data.

Epistemic contexts also supply the resources for identifying and interpreting casual interactions. It will be noted that the view of what makes data empirical espoused here relies on the notion of a casual interaction. I am not committed to a special theory of causal interaction. In fact, I believe that this view of what makes data empirical is probably compatible with a variety of ways of explicating the notion of causal interaction. I take this flexibility as a virtue of the view because, with Norton (2003) I doubt that it will be possible to identify once and for all a universal causal principle that can capture all of the reasonable stories about causal interaction that scientists want to tell. In particular I agree with Norton that the notion of causality varies according to scientific domain:

> The sort of causation we recover in physical systems is not quite the same as the sort we recover in biological domains, for example. Finally, our notion of causation evolves in response to developments in the science. May causes act a distance? Is causation anything more than determinism? The answers depend on who you ask and when you ask; and those differences in part result from developments in the relevant science. (15)

Nevertheless, without some causal story furnished by an epistemic context, data will not be hooked up to the worldly target of interest in the right sort of way. Context sensitive as it may be, there must be some way of articulating how the target is the source of the data via interaction with the intervening objects and/or processes that result in the data records.

How detailed must the causal story connecting the target to the data be? Can one simply stipulate, drawing on the resources of one's epistemic context that the target causally produced the data, and leave it at that? For instance, suppose one's favorite theory of fundamental physics says that all physical objects are composed of strings. Could one then just say of any data, whatever it is, that it was causally produced by the interaction of the strings without supplying the intervening details?

There are three things I would like to say here. First, for a given context scientists may have good reason to pay close attention to the details of the intermediary processes that connect the target and the data. For instance, that some cosmic microwave background (CMB) photons interact with electrons in galaxy clusters (the Sunyaev-Zel'dovich effect) on their way from the last scattering surface to the site at which they interact with our detectors matters for interpreting the data thereby produced. In particular, it is in virtue of this effect that the CMB data can be used to constrain theories about galaxy clustering. Similarly, avoiding or subtracting artifacts or sources of systematic error in data often relies on characterizing such intermediary processes. In the case of the CMB, an example of this would be the effect of the interaction of the CMB photons with dust encountered along their journeys. In other words, in addition to a causal story connecting the target and the data, an epistemic context can also furnish details about the intervening processes that can be used in interpreting the data.

Second, scientists are not likely to be satisfied with the minimal causal story—that the target simply "causally produced the data". This is because filling in the causal details can be a promising way to tease out strategies for eliminating competition among theories. The naively minimal causal story, that the strings causally produced the data, can be told equally well for loops, causal sets, little Platonic solids, or whatever else you fancy. To break this sort of underdetermination, which is something that scientists are interested in doing where possible, one would want to see if distinguishing features might extracted from the competing alternatives by filling out the details and if differential empirical results could somehow be generated.

Third, regardless of whether the causal story connecting target and data is sketchy or flush with details, insofar as there is a causal story to be told at all, the data can be distinguished as *empirical* rather than *virtual* with respect to a specified target, and it is this distinction that will help us make sense of the troublesome cases to which we now turn.

2.4.2 Troublesome cases

Stars in jars Philosophers of science have claimed that there are no experiments in astrophysics. Hacking (1989) writes: "Galactic experimentation is science fiction, while extragalactic experimentation is a bad joke" (559). Morrison (2015), somewhat less polemically, writes:

> there are a variety of contexts where CSs [computer simulations] do take centre stage as the source of experimental knowledge simply because the systems we are interested in are

> inaccessible to us. An example is the evolution of spiral structure in galaxies. Because the typical time and distance scales are so vast, the only way experiments can be performed is by simulating the system on a computer and experimenting on the simulation [...] In the astrophysics case we may want to say that simulation is an acceptable source of experimental knowledge simply because we are unable to conduct materially based experiments in the way we can with other types of systems. (213-214)

In stark contrast to these philosophers, there are several research efforts that scientists themselves are happy to call astrophysical experiments. The Harvard-Smithsonian Center for Astrophysics boasts research under the heading "Laboratory Astrophysics", ADMX is the acronym for the Axion Dark Matter Experiment, and there is a Laboratory for Underground Nuclear Astrophysics at Gran Sasso. How should we understand the data that are being produced by these (terrestrial laboratory) "astrophysical experiments"?

To make the puzzle more concrete, let us consider a specific example of a laboratory bound astrophysics experiment. The National Ignition Facility (NIF) at Lawrence Livermore National Laboratory uses a 4 MJ laser facility to study matter in high-energy-density states, including nucleosynthesis in stars and supernovae, instabilities in supernovae, opacity of stars, black hole accretion, nuclear reactions in stars, and planetary interiors—in short: astrophysics.

The data gathered from NIF experiments is clearly empirical in that it is not the result of a computer simulation or imagined in an sense. But is it really astrophysical empirical data? After all, the experiments at NIF are not performed by intervening on stars themselves. Are the scientists just wrong to call such activities astrophysical experiments? Is this just a "bad joke"?

I suggest that the view of what makes data empirical articulated above helps to clarify this situation. NIF data are empirical with respect to high-energy-density states of matter and their behavior since there is an interpretation of the provenance of those data such that they are the products of the causal interaction of the matter energized and confined by the NIF lasers with the laboratory detectors systems. Insofar such high-energy-density states are instantiated in far away astrophysical systems also, the data gathered in NIF experiments can be used to constrain astrophysical theorizing.

Hybrid data Let us now turn to our second troublesome case, data that has been produced by a hybrid process involving causal interaction with some worldly target of interest and via the involvement of computer simulations. For concreteness, consider the following description of just such a case from Morrison (2015):

> Although it is obvious that no simulation can prove that the Higgs particle exists, experiments and equipment designed to discover the Higgs rely heavily on knowledge produced via simulation. Hence, to say that the discovery of the Higgs was only possible using simulation is by no means to overstate the case. Again, not only is simulation required to process the experimental or 'signal' data but simulation provides the foundation for the entire experiment. To put the point in a slightly more perspicuous way, simulation knowledge is what tells us where to look for a Higgs event, that a Higgs event has occurred, and that we can trust the overall capability of the collider itself. In that sense the mass measurement associated with the discovery is logically and causally dependent on simulation. (288)

Parker (2017) makes a similar claim:

> computer simulations on their own cannot be processes by which we measure the target systems being simulated, because they do not involve interaction (or even attempted interaction) with those target systems. Nevertheless, in principle, computer simulations can be embedded in studies that do involve this interaction and, indeed, can be embedded in them in ways such that results from simulations constitute raw instrument readings or even measurement outcomes. (289-290)

Indeed, Morrison goes on to claim that this example shows that "any sharp distinction between simulation and experiment practically meaningless" (289).[7] In light of this description, we might ask: is data generated at the Large Hadron Collider empirical?

Using the view of what makes data empirical articulated above, we can clearly say that the LHC data implicated in the Higgs discovery are empirical since there is an interpretation of the provenance of those data such that they are the products of causal processes connecting Higgs particles to the detector data. From within the framework of standard model particle physics, and with the help of many other resources besides, one can tell a (in this case quite complicated) causal story connecting the Higgs particles to the production of the data. Identifying the data as empirical in this sense does not at all diminish the importance of simulations and other resources for putting the data to use. Moreover, identifying LHC data as empirical with respect to the Higgs in this context preserves the distinction between virtual data from virtual Higgs decays simulated on a computer and the actual LHC data,

[7] Morrison's claim is not just that the distinction between simulation and experiment—between virtual data and empirical data—disappears in the particular context of high energy particle physics research at the LHC, but rather more generally insofar as LHC research is taken as paradigmatically experimental research.

which as Morrison admits could not have replaced the data produced at the accelerator for the purpose at hand.

In this chapter we have been concerned with the question: what makes data empirical as opposed to virtual? I have argued for a relational and context dependent view of what makes data empirical that emphasizes the causal production of data. In the following chapter I will present a view that connects empirical data to constraints on theorizing via processing and analysis. This view aims to supply an account of empirical evidence that both does what an empiricist wants it to do and at the same time respects the roles a variety of epistemic resources, including simulation outputs, assumptions, and models, play in interpreting and using the products of empirical research.[8]

[8]Ultimately I think Parker might well be sympathetic to this move since she explicitly states that readers who find her usage of "measurement" too permissive can treat her piece "as a discussion of how computer simulation might be embedded fruitfully in practices that aim to find out the values of target system parameters" 2017, 301.

3.0 EVIDENCE ENRICHED

Traditionally, empiricism has relied on the specialness of human observation, yet science is rife with sophisticated instrumentation and techniques. This chapter 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.

3.1 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 instrumentation eventually needs to make a trans-cranial 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" (ibid., 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. (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, rather than conformity with

[1] Van Fraassen himself begins to address this problem in his work on measurement and measuring instruments (van Fraassen, 2008). I will have more to say about the views of 21st century van Fraassen below in section 3.3.

[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.

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.

3.2 ENRICHED EVIDENCE

The history of philosophy exhibits a variety of empiricisms (cf. Lipton, Peter., 2015). The shared gestalt is that knowledge of nature derives from, and only from, experience. Of course "experience" is vague, and what gives 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 possibly 'push back,' by which the 'tribunal' could possibly judge. Note that requiring that theories be consistent with the evidence does not commit one to naïve 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 that 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 over 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 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 (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. Perovic (2017) argues for a "relaxed stance" towards calibration procedures that incorporate past empirical results, theory, and the outcomes of the very experiments under consideration that is compatible with empiricism broadly construed (cf. his section 6). 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 is collected and recorded. Sometimes the data collected is 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, data is often, especially in contemporary science, 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 exposure to a common coordinate system, 5) co-adding exposures, etc. [3]

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

[3]Partial list from Neilsen's Notes on the Essentials of Astronomy Data: http://home.fnal.gov/~neilsen/notebook/astroImagingDataReduction/astroImagingDataReduction.html

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 mal-adapted. 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 mal-adapted to that theory. Consider ancient Chinese records of astronomical events. These observations were recorded using categories quite different than 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, mal-adapted 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 results 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 mal-adapted 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 mal-adapted 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 only sensitive 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* when all of the presuppositions that have been incorporated into it throughout the course of data collection and processing are either 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.[4] 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 as well as 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) captures this idea succinctly: "it is a fact of life in empirical science that experiments often give discordant results" (35). 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 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;

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

Planck Collaboration, 2016a). 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.[5] 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 is auxillary 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, GIS 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 et al., 2011, 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, formula for calculating time since the rock cooled from quantities of isotopes in the sample, and a variety of assumptions including lack of contamination from non-radiogenic ^{40}Ar (cf. McDougall and Harrison, 1999).[6]

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 cli-

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

[6]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).

mate science to molecular biology to particle physics.[7] 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 Railton's notion of an ideal explanatory text.[8] 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" (ibid., 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 Sabina Leonelli has done on data, database curation, and traveling facts (cf. Leonelli, 2009, 2013, 2015, 2016).[9] 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 re-used and re-purposed. Indeed, I think that focusing on understanding such successful

[7]cf. `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. Perovic (2017) for a philosophically-informed discussion of calibration at the LHC, especially sections 3-5.

[8]Chris Smeenk and Porter Williams independently suggested this analogy to me.

[9]See also: Howlett and Morgan (2010).

transfer across contexts gets at issues of interest to many philosophers of science, for instance those interested in epistemic progress and theory change after **?**, generalization, 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, section 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" (ibid., 30). In Leonelli (2009), she discusses this move as the "liberation" of data from the details of their provenance (746). 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 providence 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 David Colaço's dissertation *An Investigation of Scientific Phenomena*, which engages with these topics in the context of biology, pscychology, and neuroscience and Dana Matthiessen's manuscript "The Role of Local Knowledge in Mobilizing Data", which discusses how theoretical and practical knowledge support repurposing data accross diverse epistemic contexts.

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 its associated metadata—data can never be permanently decoupled from its 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.

3.3 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 like conservatism or simplicity à la Duhem and Quine. In the remainder of this paper, 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 mal-adapted 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 re-process 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 the generation of that result.

Recall the ancient Chinese astronomical observation 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" (cf. 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 fact a record of a nova. There are 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 were 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 (ibid., 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 have allowed researchers to precisely date the end of the visibility of the event. In particular, by carefully interpreting a Chinese observation record, Green (2015) extracts the date of April 6, 1056 (97).[11]

Will it always be possible to adapt initially mal-adapted results to the context of interest? Unfortunately not. Consider a data record that is mal-adapted to some epistemic context.

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

One can come to know that the record is mal-adapted in the first place by having access to the associated provenance metadata that includes information about in what way the record is mal-adapted. This very information would tell us that it will be impossible in practice to generate a useful constraint on theory from that data. In a sense this means that the evidence associated with that 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 un-enriched 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 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 re-purpose 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 66.93 ± 0.062 *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) (and discussed by Galison (1987))—the experiment that eventually discovered the existence of weak neutral currents:

> When the experiment was initailly 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 event 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. (159)

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 experimental design, such results would have still 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, i.e. 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, that results used in joint constraints could interact in epistemically problematic ways.

Consider the multi-probe 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 datasets 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 et al., 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 afterwards. 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" (ibid., 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 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 the 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 upon 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 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 are 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 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 GeV* or *13.8 billion years*, a plot, a photograph, etc.) 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* (cf. 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" (ibid.).

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" (fn 26). For van Fraassen, the requisite link between a data model and reality crucially involves locating the representation user, as in: a theory is empirical adequate to the phenomenon as represented *by us* (cf. 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 deer population (for us). This collapse is supposed to be facilitated by the role of the representation user.

However, I think van Fraassen mis-emphasizes what it is that makes results relevant and that consequently his view is unnecessarily restricted. van Fraassen's 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 has 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 not only allows many agents to appreciate the relevance of data models so produced to the theory or theories for which the data was 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 mal-adapted 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.

3.4 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 will all the available information about how the data was 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 mal-adapted results 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.

4.0 EMPIRICAL ADEQUACY

The apparent simplicity of characterizing empirical adequacy as ‘saving the phenomena’ belies the contortions required to bring theory and phenomena together and the conditions under which this is possible at all. I offer a new characterization of empirical adequacy that illuminates this fine-grained structure, explore the conditions under which empirical adequacy in the proposed sense can be adjudicated, and discuss implications for data stewardship. It seems that scientists rarely have the information that they would need to responsibly adjudicate empirical adequacy, but this situation could be dramatically improved with increased documentation and preservation of empirical results and the processes that generate them.

4.1 INTRODUCTION

In 1977 Mike Mandel and Larry Sultan displayed a series of photographs at the San Francisco Museum of Modern Art, calling the work *Evidence*. All but one of the pieces had been gleaned from files at institutions like the Jet Propulsion Laboratories, the General Atomic Company, and the Stanford Linear Accelerator Center. The photographs were displayed without identifying text, as a “poetic exploration upon the restructuring of imagery” (quoted in Phillips, 2003). One depicted figures climbing on a tree surrounded by a translucent box, backlighting them like Indonesian shadow puppets. In another, a group of men in hardhats progressively disappear into a field of white foam. What do these photographs depict? Why were they made? The viewer was left to speculate, conjuring possible narratives for these uncanny artifacts. Mandel and Sultan had bet the museum’s Curator of Photography John Humphrey a good bottle of whiskey that he would not be able to pick out an inauthentic piece that they included in the show (a photograph of an array of posts on a flat foreground with hills in the distance) guessing that without identifying information, even a keen eye could not discern the impostor (ibid.).

Mandel and Sultan's *Evidence* vividly shows that record detached from its context is utterly useless as evidence. Without sufficient metadata records of empirical research can at best serve as sources of "poetic exploration". This is precisely the point for which I will argue in what follows. When scientists want their theories to be empirically adequate, they must check them against the available evidence. But getting something to serve as evidence requires substantive information about its provenance.

Despite its crucial role in any empiricist philosophy of science, the notion of empirical adequacy is murky. An adequate theory is one that 'saves the phenomena'. But what precisely does it mean to save the phenomena? And how are scientists themselves to go about doing this?

In a 2017 paper titled "What's so special about empirical adequacy?", Bhakthavatsalam and Cartwright offer the following:

> It is good to start with a definition of "empirical adequacy", but it turns out to be difficult to find one in the philosophical literature. We think that what is usually intended is something like this: a theory (or model or set of scientific claims) is empirically adequate when the claims it makes about empirical phenomena – or at least the bulk of these claims, or the central ones – are correct, or approximately correct enough. We shall assume this usage and leave aside the question of what counts as empirical and what as theoretical (2, footnote omitted)

Boiling this definition down somewhat, Bhakthavatsalam and Cartwright are suggesting that theories are empirically adequate when the claims they make about empirical phenomena are correct. This characterization leaves room for much more to be said about "empirical phenomena" and how these are to be compared to claims that theories make. Many phenomena of interest to scientists are not studied directly but rather via sophisticated application of instrumentation and techniques that elicit data from which empirical results can be derived. In such cases, the propositions derived from the theory side are compared to the empirical results. It would be desirable to have a more precise characterization of empirical adequacy than that offered by Bhakthavatsalam and Cartwright (2017) in order to see in more fine-grained detail how the empirical content of theories runs up against the world–what is compared to what, and what conditions are required for that comparison?

The present chapter aims at two advances. The first is to give a specific characterization of empirical adequacy. The characterization I offer has the advantage that it makes transparent the sense in which the epistemic utility of empirical results as constraints on theories is dependent on the manner in which those results were generated and the presuppositions that have been incorporated into them along the way. In other words, the epistemic utility of empirical results depends crucially on the details of their provenance. The second is to show how once we conceive of empirical adequacy in this way, then the importance of good data stewardship becomes obvious and pressing in a way that it might not otherwise have been.

Suppose, as I have argued in Chapter 3, that we conceive of the evidence with respect to which empirical adequacy is to be adjudicated to be the evidential corpus consisting of all available enriched lines of evidence. Consider a single *enriched line of evidence* in the evidential corpus, which traces from records of data collection through a variety of intermediary empirical results to terminate in an empirical constraint on some theory *and* all of the enriching information associated with that line of evidence (provenance and workflow metadata). What would it take for a theory to be consistent with this enriched line of evidence? If the empirical constraint at the end of the line of evidence were *well-adapted* to the theory to be constrained, then to be consistent with that enriched line of evidence the theory need simply be consistent with the constraint at its terminus.

However, if the empirical constraint is not well-adapted to the theory to be constrained, then it could still be possible for that theory to be consistent with the enriched line of evidence that produced the mal-adapted constraint. In particular, this will be the case if there is some other intermediary result upstream in the line of evidence that produced the constraint that is adaptable to that theory. In that case, to demonstrate the consistency of the theory with the enriched line of evidence a second line of evidence will have to be generated stemming from the adaptable upstream result. The empirical constraint at the end of the new line should then be well-adapted to the theory to be constrained. If the theory is consistent with this new constraint, then (I suggest) it is empirically adequate with respect to the evidence at hand. Note that the same possibilities apply if we consider any empirical result, for instance and intermediary one, instead of the one at the end of a line. With respect to a given theory, any empirical result will either be well-adapted to it or not. If it is not well-adapted, then demonstrating adequacy will require identifying some other result upstream from it that is adapt*able*.

If we allow the term "accommodate" to stand in for these two ways in which a theory can be adequate with respect to an enriched line of evidence, then we can define empirical adequacy compactly as follows:

> **Empirical Adequacy** A theory is empirically adequate when it can accommodate every available enriched line of evidence in the evidential corpus.

In other words, a theory is empirically adequate when for every result in the evidential corpus, the line of evidence that produced that result shares an empirical result with some line of evidence, the final result of which is both well-adapted to and consistent with that theory.

The definition of empirical adequacy given above relies on the notions of *the evidential corpus*, *lines of evidence*, and the *well-adaptedness* of results, which I have already explicated in Chapter 3. In light of this, I will take the first two of these notions on board without further comment, and although I will recapitulate discussion of the third notion, I will focus on specifying the conditions under which adequacy can be adjudicated (Section 4.2) and showing how the above characterization is reflected in scientific practice (Section 4.3). Finally, Section 4.4 discusses consequences of the foregoing for the documentation of evidence and the stewardship of data and its metadata.

4.2 ADJUDICATING EMPIRICAL ADEQUACY

How exactly is the empirical adequacy of a theory checked with respect to some evidence? And what are the conditions required for that checking? To begin with, being able to judge whether a result falls within the scope of a theory is a necessary condition for being able to use that result as constraint on that theory in an epistemically responsible way. Whether a result falls within the scope of a theory depends crucially on the provenance of the data from which it derives. Without metadata on data collection, one cannot judge on which theories empirical results derived from that data bear. Such results are like Mandel and Sultan's *Evidence*–destined to be sources of "poetic exploration" not empirical constraints on theorizing.

Obviously *some* record, either the original data records or some subsequent records of processed data, are necessary for being able to derive an empirical constraint. If the constraint is to be derived from subsequent records of processed data (i.e. intermediary empirical results), then metadata on that processing is also necessary for being able to generate a constraint. To use an empirical result responsibly, one needs to know whether presuppositions have been incorporated into that result in the course of processing that will cause trouble for the constraint one would like to generate down the line. In particular, one wants to know that the result is well-adapted or adaptable to the theory one would like to constrain.

We will say that a data product is *well-adapted* when all of the presuppositions that have been incorporated into it throughout the course of data collection and processing are either 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. 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. We will say that an empirical result is *adaptable* when it can be further processed in order to generate a well-adapted constraint.

There are two main strategies for successful adjudication of empirical adequacy:

Conditions Requisite for Adjudicating Empirical Adequacy Where sufficient metadata is available, the adequacy of a theory with respect to some result may be adjudicated only if it is possible to either

1. (**forward direction**) use an extant empirical result, by either (i) using the original data records to generate an empirical constraint via data processing, (ii) using an extant intermediary result to generate an empirical constraint via data processing, or (iii) using an existing well-adapted empirical constraint, or

2. (**reverse direction**) use workflow metadata to undo incompatible data processing to reconstruct an adaptable empirical result and then generate a well-adapted empirical constraint from it via new data processing.

In other words, the resources that one needs in order to adjudicate the adequacy of a theory with respect to some evidence depends on empirical results available. In any case one needs provenance metadata for the records of data collection. Some cases will additionally require metadata regarding the subsequent data processing workflow, and some will require undoing the processing initially done. Such resources are of course not *sufficient* for adjudicating the adequacy of a theory–one needs much else besides. For instance, one certainly needs a theory to be adjudicated in the first place, one needs to be able to derive from that theory something that can be checked against the constraint produced at the end of the line of evidence, and one needs standards of constraint. There is much more to say about these further conditions that is beyond the scope of the present task. However, I can provide somewhat more in the way of support for the characterization of empirical adequacy just articulated by exhibiting cases where the importance of metadata for adjudicating adequacy is clear.

Option 1.iii above is in some sense the easiest route to adjudicating adequacy. If one is already in possession of a well-adapted result, then no further data processing is required to generate a useful constraint on theory. In another sense this route is quite onerous. Having good reasons to think that one is in possession of a well-adapted result requires knowing a lot about both the provenance of the data from which that result derives and its subsequent processing. In comparison, starting with original data records (option 1.i) will often require further processing to be carried out, however less metadata need be known at the outset. Since one is considering data records before subsequent processing, metadata regarding such processing is obviously not relevant. Once further processing has been carried out, the metadata associated with it will be relevant to the epistemic utility of the result thereby generated but the researcher(s) doing the processing could plan to document that metadata as they work.

The variety of strategies enumerated above furnish hints about what kind of resources are needed for adjudicating empirical adequacy. With the exception of 1.iii, the case in which one already has a well-adapted empirical constraint, these strategies are heuristics for how to salvage empirical constraints from evidence that would otherwise be inapplicable to the theory whose adequacy one would like to adjudicate. As I will demonstrate in the following section, these other strategies show how the adequacy of theories can be adjudicated with respect to evidence originally produced in significantly different epistemic contexts—if the right resources are available. The viability of salvaging evidence across epistemic contexts is something that should be of significant interest to philosophers of science since it serves as a mechanism for maintaining the continuity of empirical evidence over the history of science and across competing theoretical frameworks.

4.3 SALVAGING EVIDENCE

Demonstrating the adequacy of a theory with respect to some evidence often involves re-interpreting the the empirical result originally produced. Such re-interpretation sometimes significantly changes the epistemic significance of the evidence in question. Researchers at an Australian telescope picked up novel signals, initially thought to be extragalactic in origin, and dubbed them "perytons" after one of Borges' monsters. The signals were eventually tracked back to occasions on which the door to the observatory microwave oven was prematurely opened (Burke-Spolaor et al., 2011; Petroff et al., 2015). Allan Franklin has extensively treated the case of the "disappearing" 17-keV neutrino, which turned out to have been an apparition caused by the narrow energy window used in data analysis (Franklin, 1995, 2002). The discovery of superluminal neutrinos by the OPERA collaboration was later retracted, the signal being explained as resulting from a poor fiber optic cable connection (Reich, Eugenie Samuel, 2012).

If it is possible to accommodate evidence by re-interpretation it is epistemically better to do so than to abandon that evidence. The reason is that evidence that is not accommodated stands as a prima facie anomaly, threatening the empirical adequacy of the theory in question. I discuss two cases wherein evidence is salvaged so as to facilitate accommodation with respect to contemporary theorizing. These cases show the nature and extent of resources that can be needed to salvage evidence.

4.3.1 Forward direction

Use data records and their provenance metadata (1.i) The first case is one in which scientists generate a constraint from old records of data collection by utilizing enough metadata regarding the provenance of those records.

Days on Earth are getting longer. That is, the length of the average solar day is gradually elongating as the rotation of the Earth gently slows by losing angular momentum to the Moon through the tides.[1] There are two timescales at play here: universal time (UT), which measures the (variable) length of the average solar day, and terrestrial time (TT), which is a uniform timescale determinable by atomic clocks. Recent calculations of the change in the length of the day suggest that the mechanism of tidal-breaking is not enough to account for Earth's slowing spin. Other factors such as the coupling between the Earth's core and mantel and changes in the shape of the Earth after the retreat of the glaciers could account for the difference. Thus, careful determination of the changes in the length of the day over long timescales can yield constraints on geophysical modeling Stephenson et al. (2016a, 24).

To calculate changes in the length of the day one can attempt to use historical observations of astronomical events such as eclipses. This is no easy task since such observations (from ancient and medieval Babylon, China, Europe and Asia) must be carefully translated and interpreted. In order to be useful for this purpose, a historical observation must meet certain requirements (following Stephenson and Morrison, 1995, 171):

1. it must be possible to determine the geographical location from which the observation was made
2. the observation must be of an event in the solar system so that it is possible to calculate the timing of the event in TT from the applicable dynamical equations
3. it must be possible to determine the exact date of the observed event
4. it must be possible to determine the UT of the event[2]

[1] Changes in the average length of the solar day are not due to the usual seasonal variation accounted for by the tilt of the Earth's axis, but are rather the remaining changes when the seasonal variation has already been accounted for.

[2] For exceptions see Stephenson and Morrison (1995, 171)

Extracting eclipse observations from Babylonian records for instance requires deciphering cuneiform script on broken (and rare) clay tablets and then translating the inscriptions to extract the desired timing information. Figure 2 depicts and example of a Babylonian record of table of lunar eclipses for at least 609-447 BC (number 32234 in the collection of the British Museum).

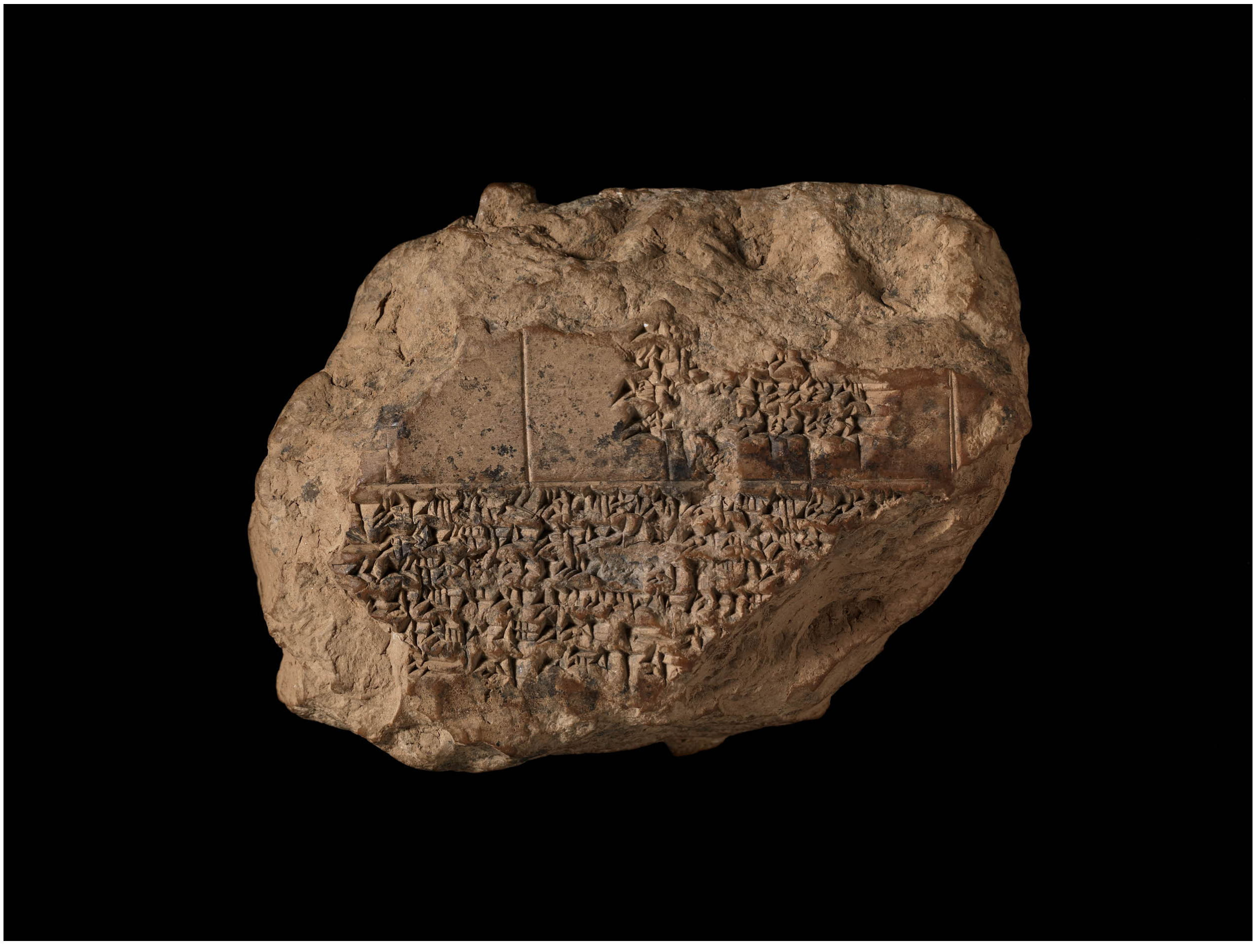

Figure 2: Babylonian table of lunar eclipses ©Trustees of the British Museum

For instance, the unit of time used is the *us*, corresponding to the interval of time in which the celestial sphere turns through 1 degree (or 4 minutes), which are thought to have been measured using a clepsydra (water clock) (Stephenson and Morrison, 1995, 174). The magnitude of an eclipse was given in *si* (fingers), where 12 fingers spans the diameter of the disk of eclipsed body (sun or moon). The timing of eclipses were given with reference to time since or time to sunrise or sunset.[3]

There are many interesting challenges in deciphering, interpreting and using Babylonian eclipse records to constrain our understanding of the evolution of the Earth's day. For instance, there are difficulties in determining whether a record is indeed a record of an eclipse observation or rather if it is a record of a prediction for an eclipse. Purported observation records may actually be predictions. Contemporary researchers suspect that an alleged observation of the lunar eclipse of 522 BC is rather best interpreted as a prediction of the eclipse of 522 BC "made by quoting an actual observation of 54 years earlier" (Huber and De Meis, 2004, 7).

Moreover, damage to the tablets themselves can "obliterate the distinction between 40 and 50, or between 4, 5, 6, 7, and 8", thereby frustrating the project of extracting good timing data from these records (ibid., 8). The particularities of the Babylonian epistemic context also obviously influence what is recorded on the tablets. For instance, the day began with sunset for a Babylonian, and the year with the vernal equinox (ibid., 10). Fascinatingly, the sign for WATER+EYE (meaning "weeping, lamentation") is to be interpreted as the time of maximal phase of an eclipse (ibid., 9, see also 14). Huber and De Meis note: "these texts are written in a peculiar shorthand notation, using a combination of (pseudo-)Sumerian ideograms and phonetically written Akkadian words. The latter sometimes are abbreviated to their first syllable. Sumerian and Akkadian lexica offer little help, the meaning of the texts must be deciphered from the astronomical context" (9).

Consider a further challenge: the sources of error relevant to these data. The accuracy of the Babylonian timing data is probably not great, perhaps admitting random errors of about 12 minutes for short time intervals and of about 15 percent of the interval recorded for large intervals (ibid., 19). In addition to these random errors, there seem to also be sources

[3] Thus (just to give a flavor of the task at hand) a typical record would be translated as follows:

> Year 168 (Arsacid), that is year 232 (Seleucid), Arsaces, king or kings, which is in the time of king Orodes (I), month I, night of the 13th. . . 5°before μ Her culminated, lunar eclipse, beginning on the south-east side. In 20°of night it made 6 fingers. 7°of night duration of maximal phase, until it began to become bright. In 13°from south-east to north-west, 4 fingers lacking to brightness, it set [. . .] (Began) at 40°before sunrise (see Stephenson and Morrison, 1995, 175)

of systematic error in the Babylonian records, which have still not been fully characterized:

> First, Babylonian measurements are affected by elusive systematic errors of 2 to 5 percent. In addition, there are systematic discrepancies between the ancient naked-eye observations of the eclipse phases and modern calculations geared towards telescopic observations. For example, observed Babylonian lunar eclipses appear to last about 12 minutes longer than calculated, while solar eclipses are shorter by about the same amount. Unfortunately, most Babylonian eclipse timings are relative to the same phase, namely time intervals from sunset or sunrise to the onset of the eclipse, so systematic effects will enter most time differences in the same way. Since they are larger than the hoped-for accuracy of the estimate, we must be extremely careful, lest they vitiate the latter. (19)

Additionally, the observation records do not always explicitly provide all of the details of the eclipse observation that contemporary researchers need in order to transform it into a useful constraint on length of day theorizing. When the desired information is not immediately given in the inscription, it is sometimes possible to recover it by engaging in what we might call "evidential forensics". For instance, one can use strategies from historiography and archeology to try to reconstruct the conditions under which the original observation was made. Mention of certain kings or cities can provide clues as to the date of an eclipse. So can more subtle orthographic conventions.

Reasoning through the circumstances of the observation can help too. Stephenson et al. (2016b) determine a constraint from a Babylonian record for 694 BC that states the Moon set while eclipsed. They argue:

> Assuming an observer at an elevation of 10 to 15 m above the ground (the height of the walls of Babylon), and horizontal refraction as 34', the true lunar altitude, corrected for parallax, would need to be -0.°4 for the whole Moon to be visible (3)

From the inferred timing of the moonset they derive the difference between UT and TT on that date. The use of the assumed height of the walls of Babylon to make this calculation is both remarkable and very clever!

The use of Babylonian observation records to constrain geophysical theories regarding the elongation of the solar day is an example of generating empirical constraints from records of data collection with the help of sufficient provenance metadata. Once the information regarding the location, date, and timing of the eclipse has been extracted from the observation record, that data can be processed to estimate the difference between UT and TT at that date. The historical trend of such differences can be compared to that predicted by models of mechanisms through which the angular momentum of the turning Earth might plausibly dissipate. Thus, the use of Babylonian clay tablets as a source for generating useful empirical constraints on contemporary geophysical theorizing is an example of salvaging evidence from a distant epistemic context.

Use an extant result and its metadata (1.ii) The second case we will consider demonstrates how a data product that has already been significantly processed can be appraised for use as an empirical constraint, provided there is enough accessible metatdata about how the data product was generated.

A fantastic example of this sort of resourceful salvaging of processed data is Astronomy Rewind.[4] Astronomy Rewind is a citizen-science project in which users add key metadata to otherwise mysterious scientific results by investigating the context in which those results were published. The publicized goal of the project is to "rescue tens of thousands of potentially valuable cosmic images that are mostly dead to science and bring them fully back to life" (American Astronomical Society, 2017). In an online interface, users are shown pictures of figures extracted from papers printed in astronomy journals like *The Astrophysical Journal* before the 1990s when publication went digital. In the first stage of the interface, users are asked to identify what type of image they are looking at (e.g. diagram or image of the sky). Further stages aim to extract specific metadata about scale, orientation, position on the sky, and astronomical bodies pictured from figure labels, surrounding text, and comparison to other catalogs of astronomical images.

The explicit aim of Astronomy Rewind is to make available the metadata that is needed in order to use these graphical results that were published in the past:

> Modern electronic astronomical images often include information about where they fit on the sky, along with which telescope and camera were used and many other details. But such "metadata" are useful to researchers only if the original image files are published along with the journal articles in which they're analyzed and interpreted. This isn't always the case (ibid.)

Astronomy Rewind is premised on the idea that such metadata can be recovered from careful scrutiny of the context in which a figure was published on the page, thereby rescuing it for use as an empirical result relevant for constraining theorizing. This project is very new

[4]See https://www.zooniverse.org/projects/zooniverse/astronomy-rewind

and so we will have to wait some time to see the fruits it will bear. However, astronomers anticipate that the enriched results produced by Astronomy Rewind could include valuable resources for constraining cosmology (Gibney, 2017). For instance, digging through these archival results could lead to the identification of more recurrent novae (ibid.). Studying these objects is presently a matter of intense interest since such research may have important implications for ΛCDM cosmology. The inference to the existence of dark energy relies in no small measure on observations of Type Ia supernovae, under the assumption that they are practically identical and can therefore be used as standard candles. Yet it is not entirely clear that all Type Ia supernovae are the same, indeed there might be two dominant sub-classes that differ in their progenitors (Schaefer, 2010, 276).[5] Recurrent nova are candidate Type Ia supernova progenitors and archival records are crucial for learning about these objects. According to Schaefer "Archival data is the *only* way to see the long-term behavior of stars, unless you want to keep watch nightly for the next century, and this is central to many front-line astronomy questions"(quoted in LSUNews, LSUNews). The hope is that results on recurrent nova (and many other phenomena) will be gleaned from the records enriched by Astronomy Rewind participants.

[5] According to Schaefer:

> supernova cosmology is based on the idea that the luminosity-decline relation has no evolution with redshift [...] The question of evolution is critical as the amount of evolution at $z \sim 1$ is comparable to the difference between cosmologies with and without Dark Energy. Without knowing the identity of the progenitor, evolution calculations are not possible and the effect can significantly change the shape of the Hubble diagram. So, in principle, the progenitor problem is critical for the entire supernova cosmology enterprise

4.3.2 Reverse direction

One could also envision salvaging a useful empirical constraint from a mal-adapted one by taking the mal-adapted result and backtracking through the data processing to reconstruct an earlier adaptable result to be re-processed into a well-adapted constraint. Doing this would require not only provenance and workflow metadata, but the amenability of the result to un-processing. It would be analogous to solving a kind of inverse problem. The task would be to reconstruct upstream intermediary results starting from a downstream result and then generate a new constraint starting from the reconstructed intermediary.

Despite arduous hunting I have found it extraordinarily difficult to find examples of salvaging useful empirical constraints by this sort of un-processing strategy. The most promising case that I have found again involves historical astronomical observations. An observation of a dark spot on the face of the Sun in A.D. 807 during the reign of Charlemagne was interpreted as the transit of Mercury (van Helden, 1976, 3).[6] However, it is impossible to see the transit of Mercury without a telescope, and so this early observation was probably of a sunspot instead. Recognizing the observation as of a sunspot however, would have been difficult for inquirers deeply embedded in an Aristotelian view of the cosmos in which the Sun and indeed all of the heavens were thought to be perfect and unchanging. Interestingly, the mistake has been made in the other direction as well. In 1631 Pierre Gassendi (with the help of an assistant) observed the Sun using a Galilean telescope projected onto paper. Gassendi observed a spot on the sun but interpreted it as a sunspot when it now seems clear that he was observing the transit of Mercury on the Sun (ibid., 4). In fact, Gassendi realized the error himself during the course of his observations (ibid., 4-5).

The reinterpretation of observations of Mercury transits as observations of sunspots, and vice versa, seem to me to illustrate the reverse direction of salvaging evidence. Pre-telescope observation records reporting transits of Mercury have to be stripped of that interpretation and recast as something else. However, this case is not an ideal exemplar because the "un-processing" required is quite minimal. For instance with regard to the A.D. 807 observation, it seems that one simply swaps in *sunspot* for *transit of Mercury* in the interpretation of the

[6] See also Stephenson and Clark (1978, 90) and Fitzgerald (1953).

record. In addition to such straightforward swaps it would interesting to see an example of the deliberate undoing of more complicated data processing to recover an earlier adaptable result for use in a novel epistemic context.

There may be other examples that I do not know about, and if there are it would be worthwhile to study them. However, the fact that such examples are not forthcoming is perhaps not so surprising. Solving inverse problems is difficult, especially if one wants to avoid massive underdetermination. It would make sense if scientists prefer to start with an extant upstream result and process anew (as in the Babylonian tablets example) rather than try to reconstruct those upstream results in the first place.

In cases where the upstream results are inaccessible then scientists would be forced to reconstruct them if they are to salvage the evidence in question. But does this ever actually happen? It seems that in practice scientists may just give up on salvaging such evidence. On my view, whether or not giving up is epistemically responsible or not will depend on if the reconstruction project in question is practically feasible or not. If enough information is available to carry out the reconstruction, then it ought to be done. Having more tightly constrained empirically adequate theories means learning more about what the world is like, and learning more about what the world is like is desirable. However, in a given case it may not be possible in practice to reconstruct intermediary results from which a well-adapted constraint can be derived. Thus the strategy of generating a new constraint by reprocessing may largely be relegated to the realm of theoretical possibility.

4.4 DATA STEWARDSHIP

If one adopts the view of empirical adequacy espoused above, there is good reason to be seriously concerned with careful documentation and data stewardship, and worried about the lack of access that scientists in fact have to data and metadata.

Consider again historical astronomical observations. Huber and De Meis lament that even with widely available and trustworthy transliterations and translations of Babylonian eclipse records, one still wants access to the original clay tablets:

> It is not easy to work directly with the Babylonian source texts—broken eclipse records are hidden in them like needles in a haystack—and one needs a reference extract not only for easier access, but also for fixing the readings, datings and interpretations upon which one is basing the analysis. This still holds now, when most of the texts are accessible in authoritative transliterations and translations. Eclipse records are still hard to locate—actually, like a human face in a crowd, the tell-take graphic patterns contained in them may jump into one's eye more readily from the cuneiform text than from a transliteration or translation. It goes without saying that an extract cannot give the full context, an therefore does not replace access to the original sources: photos, hand-copies and text editions such as those by Sachs and Hunger. (Huber and De Meis, 2004, v)

However, as we have seen, just taking good care of the original data records and to making them accessible to researchers is not enough to make those data records useful for generating empirical constraints. In addition, researchers need details about the epistemic context in which those records were made—they need metadata too.

Information scientists and database curators are among the most sensitive to the importance of preserving and stewarding metadata. Goodman et al. (2014) express the problem with terrifying lucidly:

> the amount of real data and data description in modern publications is almost never sufficient to repeat or even statistically verify a study being presented. Worse, researchers wishing to build upon and extend work presented in the literature often have trouble recovering data associated with an article after it has been published. More often than scientists would like to admit, they cannot even recover the data associated with their own published works (1)

Especially in light of the actual paucity of metadata available to scientists in practice, there are two objections we ought to consider. First, perhaps losing results over the history of science is not so bad after all—perhaps we should just let them die. Second, perhaps metadata is not really necessary for generating an empirical constraint—perhaps, for instance, the records of data collection alone are all that are really needed.

What's so bad about loss? Only about 10% of the Babylonian clay tablets inscribed with astronomical records have survived to the present day (Stephenson et al., 2016a, 3). Valuable records of civilizations periodically go up in flames. Moreover, researchers intentionally throw out data constantly. At the very early stages of data processing, outlying results are often simply excluded. For instance, as Franklin (1981) reports, Millikan's well-known results on the fundamental electric charge from his oil-drop experiments relied on 58 drops while excluding 49. At CERN's Large Hadron Collider, data from around 600 million collision events per second is pared down to about 100 or 200 events per second identified as physically interesting candidates before being recorded and stored.[7] The rest is lost to the aether.[8]

Given that the loss of data and metadata is so prevalent, is salvaging evidence really as desirable as I have suggested? One reason for thinking that salvaging evidence is not so important is that it is sometimes possible to re-do a relevantly similar observation or experiment to replace lost data. In light of this, perhaps there is no problem with throwing out the vast majority of collision events from the LHC since the relevant experimental run can simply be repeated. Rather than trying to decipher the results reported in records of the original alchemical experiments, perhaps one should reenact the experiments and interpret the newly minted results afresh.

I think this response is fair enough as far as it goes—that is, for observations and experiments that can in practice be repeated. There are unfortunately some that cannot. Astronomical events of the sort we have been considering are precisely of this nature. It is impossible to recreate a historical supernova or eclipse, and yet these occurrences harbor material from which valuable constraints on theorizing can be wrought. Historical records furnish empirical constraints on the slow dissipation of the Earth's angular momentum over the past few thousand years in a way simply not afforded by contemporary (or future) observations.

In the context of geoscience research data, the National Research Council has grappled with this very issue—whether and when it is expedient and epistemically desirable to replace

[7] https://home.cern/about/computing/processing-what-record

[8] See Leonelli (2016, 163-4):

> The exclusion of old data (sometimes called "legacy data") is perfectly understandable on practical grounds, given the difficulties involved in accessing and assembling such results, their dependence on obsolete technologies and media, the enormous variety in their formats, and the concerns surrounding their usefulness, which is often questioned given the ever-changing research contexts in which data are obtained. This compares to similar situations in other areas, most notably high- energy physics where data from particle accelerators that have been discontinued are no longer available in usable formats (e.g., the data accumulated from the predecessor to the Large Hadron Collider at CERN are kept on floppy disks and thus rarely consulted).

data rather than preserve it (Committee on the Preservation of Geoscience Data and Collections, 2002). In geoscience, it is often desirable to retain the physical samples collected from the field such as rocks, cores, and fossils. Preservation of this sort of data requires lots of space, large and well-organized storehouses. Given limited real estate, difficult decisions must be made about what is worth keeping and what may be reasonably discarded. Mistakes have already been made. Apparently the deepest well cored in the United States, the replacement of which would cost an estimated $12.3-16.4 million, has been lost (ibid., 2). To guide preservation decisions, the Council recommends prioritizing data that are effectively irreplaceable: "The committee recommends that the highest priority for retention and preservation be directed toward data and collections that are well documented and impossible or extremely difficult to replace" (ibid., 3).[9] In addition to these two factors they recommend taking into consideration potential applications of the data, its accuracy, its quality/completeness, and the viability of replicating it (ibid., see Table ES-1).

Supposing we set clearly irreplaceable cases aside, I think that there is still reason to worry about the loss of data and metadata. In losing data scientists lose opportunities to constrain their theories. Every bit of lost data can be seen as an un-checked liability that could potentially serve to undermine the empirical adequacy of one's favorite theory were it to successfully seed a well-adapted empirical constraint. Scientists should be interested in any potential threat to the empirical adequacy of their theories since whatever else they are, good theories should be empirically adequate. To ignore potential falsifiers is just to stick one's head in the sand. It does not make them go away. Scientists should be interested in salvaging old data *and* in gathering new data, since both activities contribute to the evidential corpus and thus the accumulation of constraints on empirically viable theorizing.

Moreover, we ought to proceed cautiously in how we construe throwing out data, i.e. making data "cuts". Millikan had reasons for throwing out those 49 drops and the high energy physics collaborations using LHC data also have reasons for structuring their triggers

[9] The authors note the crucial role of good documentation for the epistemic utility of these data: "All collections must be well documented before any other assessment of their utility and future can be done. Indeed, whether or not a rock, fossil, core, or other item is replaceable is completely unknown in the absence of adequate documentation to access uniqueness" (Committee on the Preservation of Geoscience Data and Collections, 2002, 3). In other words, one would not even know whether to keep or discard a sample without access to good provenance metadata.

as they do. The reasons that researchers have for throwing out data can be codified amongst the many presuppositions that enter into an enriched line of evidence, alongside, say, the presuppositions involved in preparing the experimental apparatus or in conducting the observation that produces the data in the first place. Indeed there can be very good reasons for setting data aside. Recall the discussion of ancient Chinese astronomical observations used for constraining contemporary theorizing about supernovae from Chapter 3. Clark and Stephenson found a Chinese record of a ko-hsing ("visiting star") from 902 AD that clearly describes it as having moved, suggesting that it was a tail-less comet, but which also claims the star was visible for a whole year, suggesting that it was not a comet (ibid., 45). The authors speculate that perhaps the word for "year" was mistakenly inserted for what should have been "day", or that perhaps there were two different new stars, but the matter has not be satisfactorily settled and they conclude: "Regretfully we must abandon the AD 902 star as of uncertain nature" (ibid.). This example shows that for some lines of evidence, the earliest recoverable result is not adaptable to contemporary theory. The observational record of the AD 902 event cannot be used to constrain theories of the mechanisms involved in supernova explosions given the limited available documentation, it has to be left aside.

What one wants to avoid is cherry-picking data by throwing out the pieces that seed constraints inconsistent with one's theory for that very reason. Thinking in this way has the consequence that it is not permissible to exclude outliers or anomalies simply because they are outliers or anomalies. If these are to be excluded, there should be some rationale for doing so. Perhaps the apparatus was not calibrated properly, perhaps the data reflects unaccounted for backgrounds, perhaps someone made a mistake...whatever it is, something or other ought to be blamed when data is excluded.[10]

There is a sense in which data that are excluded with rationale are not really lost. There are two sort of cases to consider. In the first sort, lack of metadata makes the constraint that one can generate from the data less crisp than it would otherwise be. For the observation record of the ko-hsing of AD 902, if more metadata were available, it might be possible to settle whether the object in question should be classed as a supernova or as a comet.

[10] Millikan it seems was guilty of the bad sort of data exclusion in that he excluded some drops simply because the implied value for the charge of the electron disagreed markedly with his best value Franklin (1981, 195).

If the verdict fell to supernova, then the observation record in question could be used as the basis for a constraint on contemporary theorizing about supernovae. Yet in absence of such metadata, the record can at best serve as the basis for a *softened* constraint. Our astronomical theories still need to be consistent with constraints derived from it, but those constraints will evidently not discern between supernovae and comets.

The second sort of case are those in which new presuppositions serve to show that any constraints derived from the data would fall outside of the scope of the theory to be constrained. For instance, it might turn out that the particular triggers used to cut collision events at the LHC render the constraints derived from the recorded data mal-adapted to some future epistemic context. If that turns out to be the case then new data, new processing or both will be desirable. In such cases, the data might very well still be useful for generating empirical constraints on theorizing—these just might not be the constraints initially anticipated. Recall the case of the "perytons" mentioned at the beginning of Section 4.3. Once the signal picked up by the radio telescope had been correlated with premature microwave oven door openings, the characteristics of that signal could no longer be used to constrain theories about exotic extragalactic astronomical phenomena as had originally been hoped. The characteristics of the signal could perhaps still be used as empirical constraints on theories of something *else* (microwave oven emissions? coffee drinking habits of radio astronomers?). The data are not lost exactly. Furthermore, the characteristics of the signal interpreted as originating from the observatory kitchen are *not inconsistent* with theories of extragalactic phenomena. Rather, they are consistent by default in virtue of falling outside of the scope of those theories. Thus, when the peryton signals were attributed to the microwave oven, they shifted outside of the scope of astronomical theories without disrupting the empirical adequacy of those theories and without being lost entirely.

I have suggested that there is indeed reason to worry about loss of evidence because lost evidence piles up as un-checked epistemic liabilities and because throwing out evidence (rather than softening it or shifting the scope under which it falls) is just anti-empiricist cherry-picking. There is another reason to worry about loss of evidence. If, over the history of science, the corpus of evidence is not cumulative then it will be difficult to makes sense of how we have learned more about they way that the world is through scientific inquiry. If the evidence of the Babylonians, or of pre-relativistic physics, is not *our* evidence in what sense could we see our own theories as accounting for more phenomena, more experience, more observations, than our predecessors? Without cummulativity, accounting for *our* evidence is just accounting for *different* evidence than our predecessors, not more. All this is to say that loss of evidence due to sub-optimal data stewardship really is a problem that ought to be addressed with full vigor.

What's so important about metadata? I have argued that adjudicating the empirical adequacy of a theory with respect to some evidence always requires access to some metadata—at least metadata associated with the provenance of the original records of data collection, and sometimes much more metadata about the data processing workflow and the presuppositions that have been incorporated throughout it. But is metadata really necessary? For instance, in the examples of historical astronomical observations discussed above, it might look like the records of data collection are all that are required to generate empirical constraints. Aren't the constraints generated from the Babylonian clay tablets themselves? Similarly, in the case of generating constraints from published results in the Astronomy Rewind project, aren't the constraints being generated from the published results themselves?

In both cases I think that metadata is still necessary for generating empirical constraints with any epistemic utility. The Babylonian clay tablets by themselves are utterly useless. As we have already seen, one needs extensive background knowledge about the cultural circumstances in which these records were produced in order to translate them into something useful for contemporary astronomers. One needs to know a lot about cuneiform script, a lot about the metrical units and calendars to which the inscriptions refer, and it helps if one knows lots else besides—like the height of the walls of Babylon. Without this enriching information about the provenance of the records, they cannot serve as the basis for useful empirical constraints.

In the case of Astronomy Rewind the very purpose of the project is to make explicit metadata that is contained in text and other circumstantial features of the published results. By attending to the axes labels, captions, context on the page etc., it is hoped that useful information can be extracted from plots, pictures, and diagrams that would otherwise be as useless for constraining theory as scraps of unidentified detritus in laboratory drawers.

That *some* metadata is necessary to generate empirical constraints with any epistemic utility does not settle the question of how much metadata is required. Is there some minimum amount of metadata that one can get away with in order to generate a useful constraint?

Although I have largely drawn examples from astronomy, the case for good stewardship of data and metadata is also obvious in the context of archiving environmental (especially climate) data. Thus the National Oceanic and Atmospheric Administration (NOAA) recommends: "Metadata that adequately document and describe each archived data set should be created and preserved to ensure the enhancement of knowledge for scientific and societal benefit" (Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA, 2007, 43). In the context of NOAA data, they provide a useful precisification of what counts as minimally "adequate" metadata:

> Metadata are all the pieces of information necessary for data to be independently understood by users, to ensure proper stewardship of the data, and to allow for future discovery. This information should include, at a minimum: a thorough description of each data set, including its spatial and temporal resolution; the time and location of each measurement, and how the data were originally collected or produced; and a thorough documentation of how the data have been managed and processed, including information about any media and format migrations, the accessibility of the data, and the algorithms or procedures used for any reprocessing, revisions, or error corrections. Collectively, these pieces of information are what make the data in an archive useful. (ibid.)

The sort of metadata (and how much of it) will be required in a given epistemic context will largely be determined by the nature of the theory to be constrained and the manner in which it is to be constrained. Therefore it will be easier to specify more concretely what sort of enriching information will be required to constrain the theory at hand from the vantage point of a particular epistemic context. Nevertheless, I think that something like the minimal requisite metadata specified by NOAA in the quote above could serve as a useful guide for making decisions about documentation in many contexts in such a way that anticipates a wide variety of subsequent applications.[11] Metadata documentation should include thorough

[11]The NOAA recommendation is similar to that implied by Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009): "raw data are typically of use only to the research group that generated them. To be useful to others, data must be accompanied by metadata that describe the content, structure, processing, accesses conditions, and source of the data [...] For observational data, the documentation of the hardware, instrumental calibrations, preprocessing of data, and other circumstances of the observation are generally essential for using data" (106).

information on data provenance and processing. Determining what counts as sufficiently thorough will depend on the nature of the data in question and the epistemic pull to include ever more metadata will be tempered in practice by the availability of resources to do so and the ability of those making documentation decisions to discern what metadata could be important for the subsequent utility of the data.

In addition to the minimally adequate metadata documentation, NOAA also specifies what documentation would "ideally" include:

> Ideally, metadata should also describe appropriate applications of the data, the relationship between the data and other data both within and outside of the archive, and enough high-level information to allow different types of users to find and understand the data. Adding these additional pieces of information would help support the discovery and integration of data across different archives and disciplines. (ibid.)

This characterization seems roughly generalizable as well. Ideally, in general, it would be desirable for information about data relationships and applicability to be included in the enriching information associated with a line of evidence. Of course, such information can only be included in metadata documentation if it has already been determined, but in general we cannot expect epistemic agents to be able to anticipate all of the contraindications and applications of any given data. Sometimes it only becomes clear in hindsight what sort of metadata it would have been desirable to have documented thoroughly.

One might be tempted to ask: ideally, should *all* data and metadata be preserved and made accessible? However, this question invites relatively unhelpful speculation since it will never be possible to accomplish this ideal in practice. There will always be context-dependent practical trade-offs to be made in the task of data stewardship. Nevertheless, two lessons from the account of empirical adequacy proposed here may be emphasized. First, *all* of the enriching information of an enriched line of evidence is implicated in the epistemic utility of the empirical result in which that line culminates, even if no one in practice ever makes all of it explicit. Second, as I have already stated, at the very least *some* enriching information about the provenance of the records of data collection will be required. The provenance of the records of data collection will determine to which theories results subsequently generated from those records can by applied as constraints, and which they will fall outside of the scope of.

Responsible adjudication of empirical adequacy relies on the accessibility of evidential resources—data records, empirical results, and metadata regarding the manner in which these were generated. Unfortunately, these very resources are often discarded in practice. This has consequences: our favorite scientific theories are not as tightly constrained as they might be if we took better care of our evidential resources. Furthermore, the paucity of records of data and metadata might make us nervous about whether or not the theories that we currently take to be empirically adequate have in fact been adjudicated in an epistemically responsible way. Fortunately the way forward is clear—ordinary good, thorough, data stewardship and the utilization of the stewarded resources in the adjudication of empirical adequacy.

4.5 CONCLUDING REMARKS

Explicating the notion of empirical adequacy reveals the under-appreciated extent of auxiliary information required to constrain theory in an epistemically responsible way. Data processing often introduces substantive presuppositions on which results are conditioned. I have presented a characterization of empirical adequacy and outlined the conditions under which such adjudication is possible. Repurposing, replicating, and even judging the relevance of empirical results implicates a rich reservoir of metadata about how those results were generated. The good news is that on the view presented here, we have the resources to appreciate the continuity of evidence across epistemic contexts by seeing the ways in which even heavily conditioned evidence can be salvaged. The bad news is that scientists rarely seem to have the information required to judge the relevance of results to the theories they would like to constrain, let alone actually produce the desired empirical constraint. Appreciating the fine-grained structure of the epistemology of empirical adequacy renders the need for concerted efforts to preserve, curate, and disseminate evidential resources strikingly apparent.

5.0 THE VARIETIES OF EMPIRICAL CONSTRAINT

5.1 INTRODUCTION

Scientists are equipped with a wide variety of strategies for generating empirical constraints on theorizing. The variety in strategies engenders a corresponding variety in the nature of the constraints they produce. In this chapter, I focus on characterizing one strategy for generating constraints that has not yet received due attention in philosophy of science. I argue that *putting bounds on a parameter* is sufficiently distinct to constitute an interesting genus in its own right. In part, my aim is to contribute to a more accurate and nuanced description of scientific methodology. In addition, I suggest that explicating the variety of strategies for generating empirical constraints furnishes the resources for certain prescriptions. Different strategies are appropriate for different epistemic contexts, and awareness of the variety of strategic options can help to guard against abandoning the hope of generating constraints too early. Some cosmologists have recently been tempted by this latter possibility in the context of theories of the early universe. I conclude this chapter by using the resources developed here to show why such temptation ought to be resisted.

5.1.1 An epistemic shift

In December 2014, something rare happened: two distinguished cosmologists addressed philosophers in print. The cosmologists, George Ellis and Joe Silk, published their provocatively titled article "Scientific method: Defend the integrity of physics" in the Comment section of the journal *Nature*. The authors were apparently incited to write the piece out of exasperation with methodological attitudes in the physics community itself.[1] The article begins,

> This year, debates in physics circles took a worrying turn. Faced with difficulties in applying fundamental theories to the observed Universe, some researchers called for a change in how theoretical physics is done. They began to argue explicitly that if a theory is sufficiently elegant and explanatory, it need not be tested experimentally, breaking with centuries of philosophical tradition of defining scientific knowledge as empirical. We disagree. As the philosopher of science Karl Popper argued: a theory must be falsifiable to be scientific.

The model of the scientific method that these physicists have in mind seems to be traditional hypothesis testing wherein for an hypothesis H and test implication of the hypothesis I (borrowed from Hempel, 1965, 7):

> If H is true, then so is I.
> But (as the evidence shows) I is not true.
> ______________________________
> H is not true

Consider for instance the hypothesis of the 19th century physician Ignaz Semmelweis, which Hempel discusses, investigating the cause of childbed fever in Vienna's General Hospital: that the appearance of the priest (preceded by an attendant ringing a bell) so terrified patients that they became more susceptible to sickness. If this hypothesis were true, then if the priest were to take a roundabout route without the bell (rather than walk through five wards on his way to the sickroom) the mortality rate in the First Division should have decreased. It did not. Therefore the hypothesis was to be rejected. This method, Ellis and Silk implied, is how science ought to be done—and cosmology is no exception.

In addition to calling out the physicists, they mention philosopher Richard Dawid by name, attributing "a philosophical case to weaken the testability requirement for fundamental physics" to him.[2] Ellis and Silk reacted to Dawid's philosophy with a call for more and deeper interdisciplinary dialog:

[1] The severity of the situation was corroborated by another article, a *New York Times* opinion piece by Adam Frank and Marcelo Gleiser, both scientists, in June of 2015 titled "A Crisis at the Edge of Physics".

[2] Note that testability and falsifiability are used interchangeably in the article. See Kragh (2014) for a list exhibiting a variety of notions that physicists may have in mind when speaking of "testability". I think traditional hypothesis testing is what Dawid considers the "canonical view" that his non-empirical confirmation is supposed to amend (cf. 2016, 191).

> We applaud the fact that Dawid, Carroll and other physicists have brought the problem out into the open. But the drastic step that they are advocating needs careful debate. This battle for the heart and soul of physics is opening up at a time when scientific results – in topics from climate change to the theory of evolution – are being questioned by some politicians and religious fundamentalists. Potential damage to public confidence in science and to the nature of fundamental physics needs to be contained by deeper dialogue between scientists and philosophers.

In fact, they concluded the article by calling for a conference to be convened in the following year to begin to address what they saw as the pressing action item: rigorous philosophical engagement on the connection between the scientific method and empirical testability. Dawid took them up on the offer and convened a conference in Munich in December of 2015 with the explicit mission of having physicists and philosophers hash out the issue together.

This sort of dialog is so unusual that it warrants further attention. Physicists rarely take any concerted interest in the work of philosophers. If they engage with any philosophy of science at all it is almost certainly with a stance on the importance of falsifiability that they attribute to Karl Popper (Ellis and Silk are evidently no exception). Physicists typically have no understanding of the state of contemporary advances in philosophy of science, nor even a good grip on the sort of projects and questions that philosophers of science would find engaging today. The fact that the result of this recent encounter seems to have been to mainline Dawid's particular non-empirical application of Bayesian confirmation theory is, I think, cause for some concern. Concern is warranted in part because there are indeed real and pressing methodological and epistemological questions arising in contemporary physics.

For instance, scientists are hotly debating amongst themselves how theories of the early universe are to be constrained. Kragh (2014) has raised the possibility that debates such as this one may be signs of a full scale "epistemic shift" in cosmology—a shift away from presuming a tight connection between empirical testability and scientific credibility. And interestingly, as Ellis and Silk intimate, the physicists are having this debate in public. For instance, in a 2017 *Scientific American* article, physicists Ijjas, Steinhardt, and Loeb, conclude "inflationary cosmology, as we currently understand it, cannot be evaluated using the scientific method" and suggest that by nevertheless refusing to abandon the inflationary paradigm, some scientists "have proposed that [...] science must change by discarding one of its defining properties: empirical testability" thereby "promoting the idea of some kind of nonempirical science".[3]

One fairly clear case of cosmologists instantiating a shift towards non-empirical methods is the defense of the eternal cyclic universe cosmology offered by proponents Paul Steinhardt and Neil Turok. According to that theory our universe lives on a surface in a higher-dimensional space (a "brane") and very close by in this space there is another such brane that periodically collides with our own, thereby causing successive "big bangs" on the order of every trillion years. The basic scenario is as follows. The branes collide as a result of a springlike force (the "interbrane force") between them (Steinhardt and Turok, 2002, 1437). The potential energy density associated with this force is positive when the branes are far apart after a rebound, which causes the expansion of the branes to accelerate. Because of the nature of the interbrane force, the energy density then passes through zero and becomes negative, which eventually causes the branes to collide again.

Note that an eternal cyclic model is indistinguishable in principle from a finite cyclic model. The claim that cycles are similar, periodic, and occur eternally cannot be supported by observational or experimental evidence even in principle. Instead of empirical support, Steinhardt and Turok offer non-empirical support by claiming that the eternal cyclic model has significant explanatory power, namely that it explains why there is a dark energy component in the energy density of the universe.

In particular, they suggest that the eternal cyclic model "naturally" provides dark energy a key role in cosmic history (ibid., 1439). The best reconstruction that I can muster of the explanation Steinhardt and Turok offer is a teleological explanation whose telos is the eternal nature of the universe. Consider the following quotes in which Turok and Steinhardt invoke their eternal cyclic scenario to explain the presence of dark energy:

> [T]he cyclic model leads naturally to the prediction of quintessence and cosmic acceleration, explaining them as essential elements of an eternally repeating universe. (ibid., 1439)

> Each cycle may be almost identical to the one before it, and the presence of the cosmological constant allows for a stably periodic solution in which the cycles continue forever. (Turok, 2003, 786)

[3] In response to the Ijjas et al. article, 33 disgruntled scientists wrote a collective piece that reacts to their conclusions with the flat response: "We have no idea what scientists they are referring to." This suggests taking the "crisis" type claims with a grain of salt.

> If a periodic cyclic solution is to be an attractor, it is essential that positive dark energy be present to redshift away the density inhomogeneities present form the previous cycle so they do not accumulate and make the universe more and more inhomogeneous with every cycle. (ibid., 798)

> [D]ark energy is just what is needed to restore the branes to a flat, parallel state, thereby allowing the collisions to repeat in a regular manner. (Steinhardt and Turok, 2007, 168)

In other words: if we are to have the eternal cyclic solution (which is conceptually desirable for other reasons) then our theory must have some component that blocks inhomogeneities from accumulating across successive cycles because if the theory predicts that they do accumulate then it is falsified by the apparent homogeneity of our observable universe on large scales. In other words, if dark energy is present then a cyclic universe is possible. The power of this teleological explanation is apparently supposed to provide support for the eternal cyclic model.

I take it that this is an example of the sort of reasoning that Ellis and Silk think needs to be deflected if the integrity of science is to be maintained. But Ellis and Silk make use of precious few philosophical resources in their defense. Engaging almost exclusively with a caricature of Popperian falsifiability straitjackets the scientists who are trying to get clear on the methods appropriate for their own disciplines. It corners them into a kind of false dilemma: either speculative theorizing is not scientific (because not falsifiable), or is indeed scientific, but because empirical evidence is not crucial for theory choice after all. The dilemma is false of course, because there are plenty of ways to retain a crucial role for empirical evidence in theory choice without demanding that all science model Semmelweis-style rejection of hypotheses.

5.1.2 Resisting the shift

It would be good to be able to meet the epistemological and methodological questions that the physicists themselves are raising with the resources of a robust empiricism from the philosophical side of the aisle. Instead, what seems to have happened so far is that Richard Dawid's minority viewpoint has gotten disproportionate coverage. If Ellis and Silk are right that this is a "battle for the heart and soul of physics", then Dawid's philosophical campaign has been conspicuously free of challengers.

According to Dawid, a major point in favor of a role for non-empirical confirmation is that in historical sciences "the general character of scientific hypotheses in those fields often makes it difficult to extract specific and quantitative predications from them" (Dawid, 2016, 192). He thinks that this feature, in conjunction with the fact that "those scientific fields often deal with empirical situations where most of the empirical record has been irretrievably lost to natural decay or destruction" renders empirical confirmation "patchy" and incapable of supporting trust in the theory in question (ibid.)[4] For Dawid, this lack of support constitutes a lacuna for non-empirical confirmation to fill.

A great example of precisely what Dawid refers to—a field in which "specific and quantitative predictions" are difficult to extract—is indeed contemporary physical cosmology. In this context, traditional hypothesis testing is not always readily applicable due to the relatively underdeveloped status of theorizing in the field. The nature of dark energy is as yet so mysterious that concrete physically plausible proposals are just not available. Observations of Type Ia supernovae imply the accelerated expansion of the universe and "dark energy" names whatever it is that is responsible for that acceleration. One cannot test a hypothesis that one has not formulated in the first place. Happily, hypothesis testing (à la Semmelweis) certainly does not exhaust the variety of strategies that scientists have for generating empirical constraints on theorizing.

In fields where theory is relatively underdeveloped, it is sometimes desirable to constrain the space of empirically viable theories without actually specifying them individually. In such cases, the following strategy can be useful for constraining theorizing instead of traditional hypothesis testing:

[4] The meaning of "trust" in this context is not totally clear. Later in the piece Dawid writes:

> we understand trust in a theory in terms of the theory's empirical predictions rather than in terms of truth. If a scientist trusts a theory, she believes that the theories predictions in its characteristic regime, if tested, will get empirically confirmed. If a theory's predictions in its characteristic regime are indeed in agreement with all possible data, the theory shall be called empirically viable. (ibid., 194)

This characterization supports reading Dawid's notion of trust as roughly van Fraassen's version of acceptance, namely, belief that a theory is empirically adequate, where adequacy is determined with respect to all possible observations.

Putting bounds on a parameter Model the phenomenon of interest with a generic parameter that refrains from committing to unmotivated assumptions about its nature. Construct an inference chain that connects empirical data to that parameter.

The requisite inference chain may have to proceed in many steps via intermediary parameters that represent other phenomena that are presumed to physically mediate between the sources from which the data are collected and the phenomenon of primary interest. Recent efforts to constrain theories of dark energy illustrate this strategy. Within the now-standard framework of ΛCDM cosmology, dark energy is modeled as one among a handful of ingredients contributing to the total energy density of the universe. Some of the most basic questions that one can ask about dark energy are unanswered at present. For instance, does dark energy function as a cosmological constant (contributing the same energy density to the total budget over all cosmic epochs) or does its contribution vary in time? In an attempt to answer this question, cosmologists represent whole families of different models of dark energy. Making some assumptions about the nature of dark energy, it can be characterized by an equation of state parameter w, the ratio of the "pressure" and energy density associated with it. A cosmological constant would correspond to $w = -1$, and so the question of the time variability of dark energy becomes a question about whether the value of w departs from -1 or not. Cosmologists then use "probes" such as supernovae, gravitational lensing, galaxy clusters, and Baryon Acoustic Oscillations to constrain the phenomenology of dark energy via w. Posed thus, empirical constraints can be placed on the equation of state parameter w thereby whittling away the space of empirically viable models of dark energy (generically construed). All this is accomplished without the traditional formula of specifying a particular theory, identifying a particular hypothesis, deriving a concrete prediction from that hypothesis, and testing the prediction against empirical results whose relevance is determined by the prediction.

In the specific case of dark energy, the inference chain required to generate constraints on dark energy theorizing connects data collected on the probes (like supernovae) to the generic parameter w by passing through intermediary parameters like apparent magnitude and luminosity distance.

Although it will take further work (see section 5.3.1) to argue that this strategy is not well-characterized as traditional hypothesis testing after all, we can already note a conspicuous difference. A particular hypothesis is simply not specified in the former—and for good reason. Efforts to understand mysterious phenomena are not always aided by specific hypothesis about them. This might suggest that, supposing my argument below for the dissimilarity with respect to hypothesis testing is compelling, this strategy is better countenanced as some kind of exploratory experimentation. After all, exploratory experimentation too is motivated by epistemic contexts where little is known about the phenomenon of interest and thus where hypotheses regarding the nature of the phenomenon have not yet been specified. In section 5.3.2 I will present an extant account of exploratory experimentation (as systematic parameter variation) and argue that this suggestion does not pan out. Finally, in section 5.4 I will argue the efficacy of the strategy elucidated in section 5.2 shows how epistemic progress can indeed be made in contexts where it can be difficult to make "specific and quantitative predictions", and that therefore we have no need for recourse to the non-empirical.

Before I present these arguments, I illustrate the proposed strategy in more detail by exhibiting the inference chain for constraining w in slightly higher resolution. The purpose of this illustration is to both document an actual instance of the strategy in scientific practice in order to demonstrate that its characterization is not speculative and also to furnish details that will be useful in making the two arguments just mentioned.

5.2 PUTTING BOUNDS ON THE DARK ENERGY EQUATION OF STATE PARAMETER

Contemporary cosmologists model the universe as a solution to Einstein's field equations by making the idealizing assumption that all of the matter/energy in it is distributed homogeneously and isotropically.[5] The dynamics of the universe can then be characterized by an equation relating the components that contribute to the energy density of the universe to

[5]The exposition here follows section 2.2. of Weinberg et al. (2013). Note that others make slightly different choices in notation, for instance writing ϵ for energy density rather that u. Besides notational choices there is nothing non-standard about the formalism and modeling explicated here.

the evolution of a scale factor that tracks distances in spacetime, the Friedmann equation[6]

$$\frac{H^2(z)}{H_0^2} = \Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\phi \frac{u_\phi(z)}{u_\phi(z=0)} \tag{5.1}$$

The unknown dark energy component is modeled as an ideal fluid which, as I said above, one can associate with the equation of state parameter w equal to its "pressure" p_ϕ over its energy density, the u_ϕ in equation 5.1, i.e.

$$w(z) = p_\phi(z)/u_\phi(z) \tag{5.2}$$

For constant w, the expression in the far right term of equation 5.1 would be:

$$\frac{u_\phi(z)}{u_\phi(z=0)} = (1+z)^{3(1+w)} \tag{5.3}$$

A true cosmological constant, that is a dark energy component whose corresponding energy density does not vary with time/redshift, corresponds to $w = -1$.

A strategy, exemplified by the Dark Energy Task Force (DETF) for making headway on the nature of dark energy is to

1. determine as well as possible whether the accelerating expansion is consistent with a cosmological constant, i.e., unevolving dark-energy density
2. measure as well as possible any time evolution of the dark energy density (Albrecht, Amendola, Bernstein, Clowe, Eisenstein, Guzzo, Hirata, Huterer, Kolb, and Nichol, Albrecht et al., section D)

To accomplish this, cosmologists need to get empirical evidence to hook up with the parameter w. They want to determine whether w deviates from -1 (addressing point 1 above) and if it does, to map the evolution over time (point 2). The next step in hooking up theorizing about dark energy to empirical evidence involves specifying a parameterization of w. There

[6]The Ωs are the energy densities of the different components (m for matter, r for radiation, k for curvature, and ϕ for the unknown dark energy component) normalized with respect to the total energy density that would produce flat spatial geometry. z is redshift, and H is the Hubble parameter which is just the time derivative of the scale factor (usual a) over the scale factor. H_0 is the Hubble parameter today.

are choices about how to do this.[7] The DETF, for instance, uses a two-parameter model:

$$w(a) = w_0 + w_a(1 - a) \tag{5.6}$$

Figure 3 represents empirical constraints on theorizing about dark energy via the dark energy equation of state parameter.[8]

So much for preliminaries. At this point we have not yet done much to elucidate how *exactly* empirical constraints on theorizing about dark energy work. For instance, we have not yet exposed how the various datasets referenced in Figure 3 are cajoled into pronouncing upon the empirical adequacy of various possible values of w_a and w_0. In order to get at these details, we will have to peel back another layer.

5.2.1 Observables

At a coarse descriptive grain, energy density in the form of dark energy contributes to the total energy density and thereby to both the geometry of the universe and the evolution of the scale factor. Moreover, accelerated expansion driven by dark energy serves to slow the gravitational collapse of overdense regions, i.e. slow the growth of structure in the universe. Thus, one can hope to learn about the nature of dark energy by understanding the evolution of distances and material structures in the universe.

These physical associations between the nature of dark energy on one hand and distances

[7] Other options include a simple Taylor expansion

$$w(z) = w_0 + w'z + \dots \tag{5.4}$$

or

$$w(a) = w_p + w_a(a_p - a) \tag{5.5}$$

where a_p is a "pivot" value of the scale factor chosen to minimize the correlation between errors associated with w_p and with w_a. Another approach is to approximate w(z) with a stepwise-constant function defined in discrete bins over some range of allowed values.

[8] The shaded areas delimit the region that the parameter values associated with the true model can be expected to lie within with 95% confidence (light) and 68% (dark), given the specified datasets. The datasets considered here are denoted by TT (the Planck 2015 cosmic microwave background temperature data), lowP (low-l polarization), ext (BAO, JLA, H_0) and WL. JLA stands for Joint Light-curve Analysis. It involves supernovae data from the SuperNova Legacy Survey (SNLS) and the Sloan Digital Sky Survey (SDSS) (for references see Planck Collaboration, 2016a, 25). BAO (baryon acoustic oscillations) data come from SDSS and from the Baryon Oscillation Spectroscopic Survey (BOSS), (ibid. 24). The WL (weak lensing) data comes from CFHTLenS (ibid., 28).

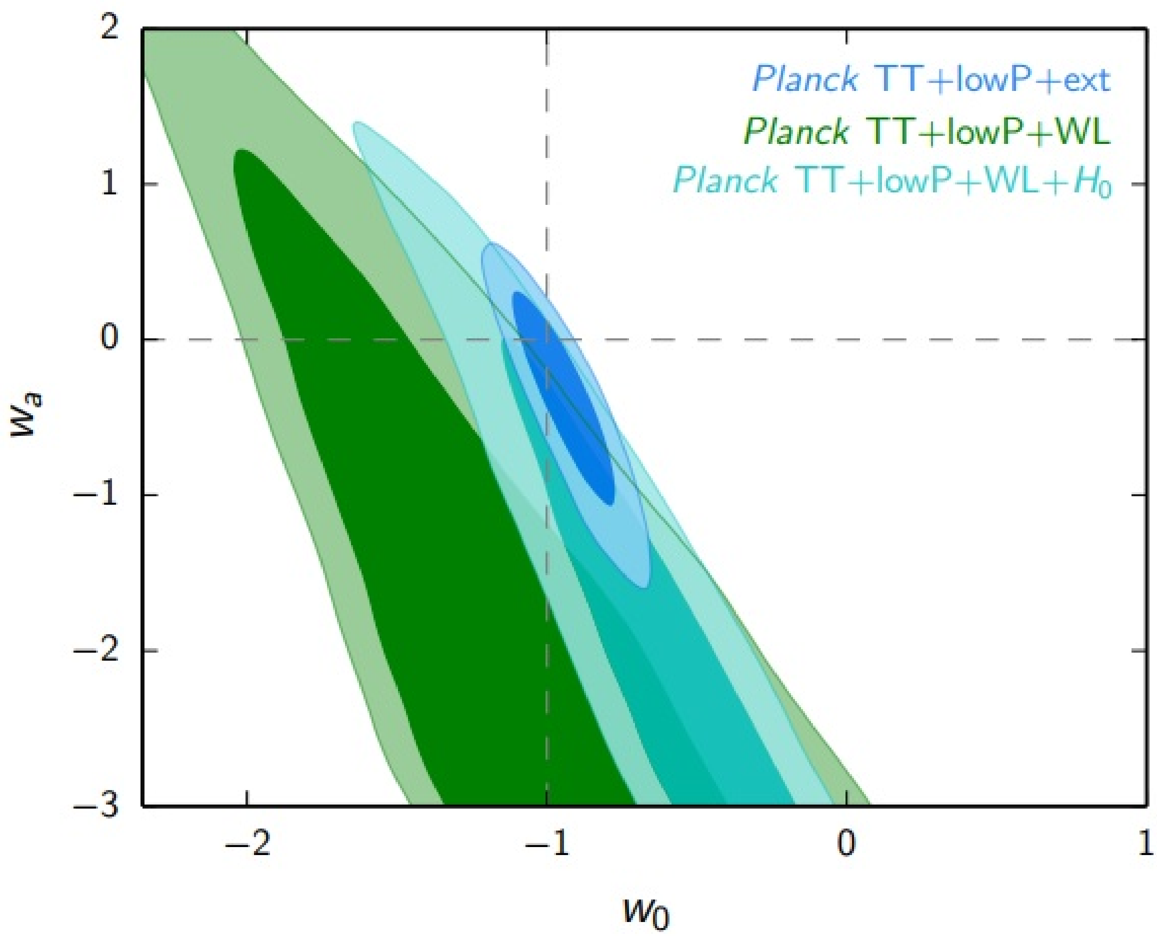

Figure 3: Constraints on dark energy equation of state parameters, from Planck Collaboration (2016a, 40)

and structures on the other hand are represented in the formalism by way of intermediary "observables". These are certainly not observables in the typical philosopher's sense meaning observable by human eyes. They are better construed as intermediary parameters, i.e. parameters that serve some mediating role between data and the parameter on which bounds are eventually sought. In this case, relevant observables are $H(z)$ the Hubble parameter, $D(z)$ (either luminosity or angular diameter distance), and $G(z)$ the growth function (cf. Weinberg et al., 2013, 98). We have already encountered $H(z)$ above in the Friedmann equation, but the other two parameters warrant some further discussion.

There are many operationalizations of “distance” in astronomy. If two objects of the same intrinsic luminosity are placed at different distances from the observer, the farther one will appear less luminous. If two objects of the same intrinsic diameter are placed at different distances from the observer, the farther one will appear smaller. Thus both apparent luminosity and apparent diameter can serve as proxies for transverse distance. Expressions for luminosity distance and angular diameter distance are given in Figure 4.

measurable	Definition
proper distance	$D(z)=\int_0^z \frac{dz'}{H(z')}=\begin{cases} \lvert k\rvert^{-1/2}\sin^{-1}\left[\lvert k\rvert^{1/2}r(z)\right] & k>0 \\ r(z) & k=0 \\ \lvert k\rvert^{-1/2}\sinh^{-1}\left[\lvert k\rvert^{1/2}r(z)\right] & k<0 \end{cases}$
luminosity distance	$d_L(z)=r(z)(1+z)$
angular diameter distance	$d_A(z)=r(z)/(1+z)$
volume element	$dV=\frac{r^2(z)}{\sqrt{1-kr^2(z)}}dr\,d\Omega$

Figure 4: Intermediary parameters, from Albrecht, Amendola, Bernstein, Clowe, Eisenstein, Guzzo, Hirata, Huterer, Kolb, and Nichol (Albrecht et al., 29)

The growth function $G(z)$ represents the evolution of the density of matter over cosmic history. Consider an initial distribution of matter/energy in the universe that is relatively uniform and then consider fluctuations added to that distribution that render some places in the universe more dense than average. Assuming the matter to be pressureless dark matter subject to gravity, on large scales the evolution of density fluctuations follows linear perturbation theory. In the context of general relativity (see Weinberg et al., 2013, 97) the linear growth function $G(t)$[9] obeys

$$\ddot{G}_{GR} + 2H(z)\dot{G}_{GR} - \frac{3}{2}\Omega_m H_0^2(1+z)^3 G_{GR} = 0 \tag{5.7}$$

Solutions can only be written for particular forms of $H(z)$, which would require specifying a particular dark energy model by furnishing its energy density $u_\phi(z)$. To avoid this one can consider an approximation of the logarithmic growth rate of the perturbations

$$f_{GR}(z) \equiv \frac{d\ln G_{GR}}{d\ln a} \approx [\Omega_m(z)]^\gamma \tag{5.8}$$

which can then be integrated to yield

$$\frac{G_{GR}(z)}{G_{GR}(z=0)} \approx \exp\left[-\int_0^z \frac{dz'}{1+z'}[\Omega_m(z')]^\gamma\right] \tag{5.9}$$

with

$$\gamma = 0.55 + 0.05[a + w(z=1)] \tag{5.10}$$

which exhibits the connection between $G(z)$ and the dark energy equation of state parameter w.

Thus, one step in generating a constraint on w is to identify connections in the representational formalism between w and intermediary "observables". As Weinberg et al. (2013) note, the connection of the intermediary observables to the representation of dark energy in the formalism goes through the Friedmann equation:

> The properties of dark energy influence the observables – $H(z)$, $D(z)$ (either luminosity or angular diameter distance), and $G(z)$ – through the history of $\frac{u_\phi(z)}{u_{\phi,0}}$ in the Friedmann equation. (98)

This connection is part of what affords the strategy for generating empirical constraints on w:

> The above considerations lead to the following general strategy for probing the physics of cosmic acceleration: use the observations to constrain the functions $H(z)$, $D(z)$, and $G(z)$, and use these constraints in turn to constrain the history of $w(z)$ for dark energy models (ibid., 99)

[9]The variable t is for our purposes functionally equivalent to the variable z (redshift) used above.

However, this step—connecting w to observables—is not yet enough of the story for us to clearly see how empirical constraints on theorizing about dark energy are generated, and therefore not enough to clearly see what the nature of these constraints are. A second major step is required to get from the observables to actual empirical results.

5.2.2 Hooking up the observables

Supernovae, gravitational lensing, galaxy clusters, and Baryon Acoustic Oscillations are probes of observables relevant to dark energy. To take just one of these as an example, measuring the luminosity curves of Type Ia supernovae (SNe Ia) can be used to constrain the intermediary parameter ("observable") $H(z)$. SNe Ia are stars that explode in a characteristic way such that the evolution of the light that they output over the course of the explosion follows nearly the same form. Recording the apparent magnitude of SNe Ia over time yields data like those plotted in the top part of Figure 5. The similarity between the output of these different supernovae can be made more obvious by applying a stretching transformation to yield the plot on the bottom of the figure. Since the similarity is thought to be intrinsic to the SNe Ia, the apparent dissimilarities between various light curves are attributed to their different distances from the observer, and can therefore be used to estimate those distances. To see concretely how data from SNe Ia can be used to constrain observables like $H(z)$, let us consider in particular the photometric distances estimated from the 3rd year Supernova Legacy Survey (SNLS).[10]

SNe Ia candidates were identified in an extensive photometric survey using the Canada-France-Hawaii Telescope (CFHT). The data initially collected are optical images recorded using the CFHT digital camera Megacam using four filters. While we do not need all of the details, it is worth a few stages of the subsequent data processing steps (Guy et al., 2011):

1. Preliminary image processing: standard image processing is performed including bias

[10] The SNLS astronomers talk about the relationship between the observations and constraints on dark energy in ways that are surprising to a philosopher's ear. They claim for instance that "SNe Ia observations are currently the most sensitive technique to study dark energy or its alternatives, since they can be used to *directly measure* the history of the expansion of the Universe" and "SNe samples [...] give consistent *measurements* of the effective equation of state parameter of dark energy (w, the ration of pressure over density)" (Guy et al., 2011, 2, emphasis added). These remarks are surprising because the steps involved in connecting the data collected at the CFHT to the parameter w are pretty intricate.

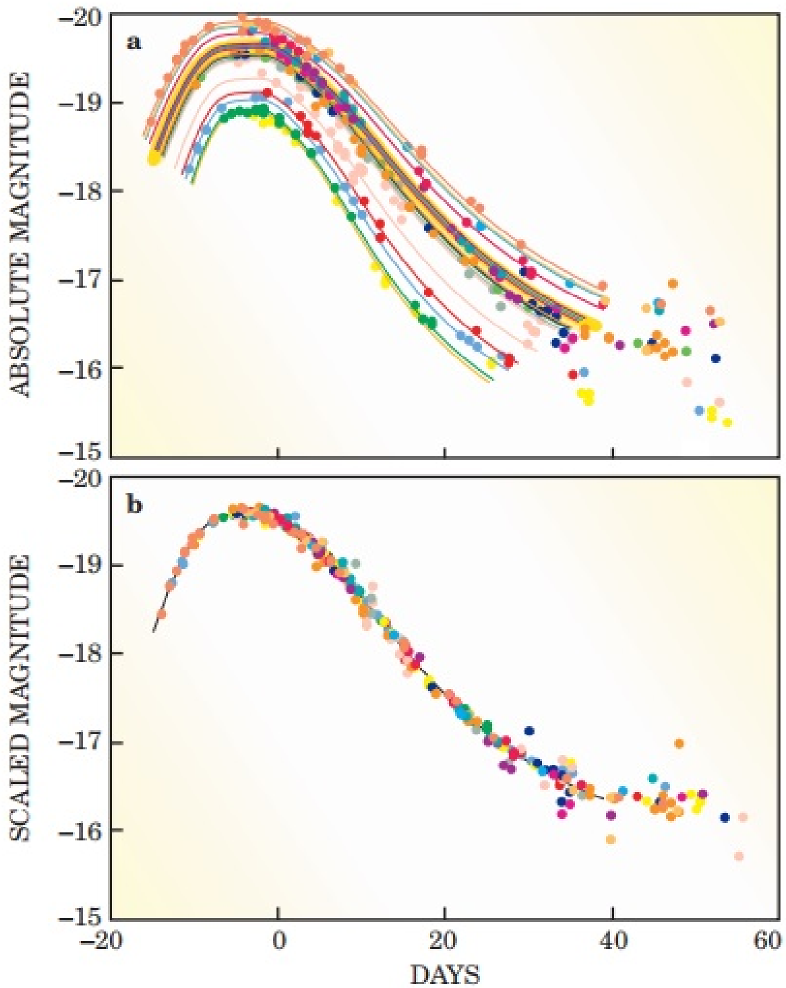

Figure 5: SNe Ia light curves, from Perlmutter (2003, 54)

subtraction, flat-field correction, and fringe removal

2. Further image processing: sky-background subtraction, astrometry, and photometric correction
3. Fitting the supernova fluxes
4. Calibration of fluxes using the magnitude of the star BD + 174708 (chosen by convention)
5. Fitting the flux data to a model of supernovae light curves and extracting parameters including the rest-frame magnitude, shape, and color for each supernova
6. Calculating the distance modulus

The last step is accomplished via the following equation:

$$\mu = m_B^* - M + \alpha \times shape - \beta \times C \tag{5.11}$$

where m_B^* is the rest-frame magnitude in the B band, and C is the color parameter, in this case the $(B - V)$ color at maximum derived from a weighted combination of various properties of the light curve (for details see ibid., 12). Here B and V refer to two frequency ranges of light observed. The particular *shape* parameter depends on the fitter used in step 5. For the fitter SiFTO for instance, it is $(1 - s)$, where s is the stretch factor (related to the kind of stretching transformation referenced above regarding Figure 5). M is the absolute magnitude, which is fit along with the two linear coefficients α and β (ibid., 15-16). Note that the distance modulus μ is related to the distance r in parsecs according to:

$$\mu = 5 \log_{10}(\frac{r}{10}) \tag{5.12}$$

To summarize this section so far, the empirical data (optical images of supernovae) are hooked up to the intermediary parameter distance via further intermediaries, in this case: model parameters extracted from light curve fits such as rest-frame magnitude, shape, and color. We have seen that empirical constraints on theorizing about dark energy can be generated by, for example, connecting up digital images recorded from a telescope to a generic parameter characterizing a whole space of possible dark energy theories, *without ever specifying those theories or particular hypotheses derived from them.*

I suggest that this example demonstrates how epistemic progress can be made in a circumstance where it is "difficult to extract specific and quantitative predictions" without relinquishing the uniquely important role of the empirical. I will argue below that this example provides reason to resist the move that Dawid endorses and that Steinhardt and Turok exemplify (non-empirical confirmation). First I want to argue that the strategy we have encountered in this section is interestingly distinct from other more familiar strategies in science. I do so in order to offer a kind of error theory: the distinctiveness of this strategy helps to explain why the shift towards the non-empirical seems appealing—the inapplicability of familiar strategies to cases like dark energy make it seem like we are out of options, but this appearance only reflects the failure to consider less familiar (but no less legitimately empirical) strategies.

5.3 THE DISTINCTIVENESS OF THE STRATEGY

5.3.1 Against construing putting bounds on a parameter as traditional hypothesis testing

The strategy of *putting bounds on a parameter* is not well-characterized as traditional hypothesis testing. On the traditional model of hypothesis testing, some particular implication of a hypothesis specified in advance is tested against an empirical result. In contrast, putting bounds on a parameter allows one to constrain the empirically allowed values of that parameter, even without specifying a particular model that implies some particular parameter value. In essence, one can get a constraint without making a prediction. This is a particularly useful approach when theoretically well-motivated proposals are hard to come by.

Are there really no hypotheses and implications being specified when empirical constrain proceeds through putting bounds on a parameter? One could perhaps think of the whole w_a-w_0 plane as a very long conjunction of specific implications of different hypotheses. Each point in the plane corresponds to a tuple that could in turn be associated with some model of dark energy. Considered in this light, one could see putting bounds on parameters as a particularly efficient, but not particularly novel, version of hypothesis testing.

While there is some sense to this proposal, I think it would be ham-fisted. It is not the case that cosmologists sat around before the Planck 2015 data was analyzed and specified a very long list of tuples for values of w_0 and w_a. The whole point of parameterization is to be able to constrain a swath of values without having to articulate each individual possibility. Moreover, I think it is useful to maintain a distinction between cases where theoretically motivated hypotheses are articulated and used to derive particular implications and the (ham-fisted) sense in which a plane in parameter space can be construed as a specification of a long list of fine-grained implications. When I said above in the Introduction to this chapter that "concrete physically plausible proposals are just not available", I was being a little misleading. There *are* some concrete theoretical proposals regarding the nature of dark energy, see for instance Planck Collaboration et al. (2015) (although the extent to which these are *plausible* is subject to debate and these more specific proposals are often parameterized themselves). However, even if it were possible to test the implications of these more specific proposals, this is not what is happening in the case considered above. Parameterizing the dark energy equation of state allows researchers to make progress constraining theorizing, even when the shape that such theorizing will take is for the time being relatively opaque.

Even if we maintain that there is an important difference between the plane of parameter values and specifying test implications, it seems that there is at least one specific test implication in play in the dark energy equation of state example, namely $w = -1$. This particular prediction is implied by the hypothesis that dark energy is a true cosmological constant. However, attending to this particular prediction alone would be to ignore the other constraints that are generated simultaneously (i.e. those on other values for the parameter) and thereby fails to capture what is so powerful about this strategy for generating empirical constraints. The prediction $w = -1$ is constrained, but so too are other values for the equation of state parameter.

If putting bounds on a parameter is unlike traditional hypothesis testing, perhaps there are yet other extant methodological frameworks that would suit it more aptly? The dark energy case fails to fit the traditional model of hypothesis testing in large part because particular hypotheses and predictions are not specified in advance. That is, this case exemplifies an epistemic context in which theories of the phenomenon under investigation are not (or have not yet been) specified. Since this is also true of exploratory experimentation, perhaps that would be a better fit.

5.3.2 Exploratory experimentation

Some attention has been paid in the philosophy of science to exploratory experiments and the significance of such research for questions regarding the theory-ladenness or autonomy of empirical results, conceptual change, progress in science, and realism (see e.g. Galison, 1987; Steinle, 1996, 2002; Franklin, 2005; Marcum, 2010; Stojanović, 2013; Karaca, 2013). Philosophers of science have addressed the issue of scientific research context where theoretical proposals are sparse in the literature on exploratory experimentation. Exploratory experimentation is contrasted with “theory-driven” experimentation, and the extant literature is concerned in large measure to show how experimental research proceeds even in contexts where theories of the subject of investigation are either in development or as yet absent. I will argue that the strategy exemplified in the dark energy case is not well represented by at least one of the most developed and prominent accounts of exploratory experimentation, that of Steinle (1996).[11]

Steinle characterizes exploratory experimentation as:

> driven by the elementary desire to obtain empirical regularities and to find out proper concepts and classifications by means of which those regularities can be formulated. It typically takes place in those periods of scientific development in which–for whatever reasons–no well-formed theory or even no conceptual framework is available or regarded as reliable. Despite its independence from specific theories, the experimental activity may well be highly systematic and driven by typical guidelines. They are of a general methodological type. Here is a list of the most important ones:
>
> - Varying a large number of different experimental parameters,
> - determining which of the different experimental conditions are indispensable, which are only modifying,
> - looking for stable empirical rules,
> - finding appropriate representations by means of which those rules can be formulated,
> - forming experimental arrangements which involve only the indispensable conditions, thus presenting the rule in particular clarity. Those experiments are typically characterized as "simple," "elementary," or "pure" cases.
>
> (S70)

Steinle takes Ampère's experiments with a magnetic needle and electrical wire as a paradigm example (S66-S67). By varying the relative position between the needle and the wire, Ampère discovered that the needle would always move to be perpendicular to the wire. It is important for Steinle's characterization that this experimentation was conducted in the absence of any well-articulated theory of the subject matter under investigation: "No specific theories of electricity and magnetism played a role. As a result of the experimentation, a general rule emerged" (S67). Ampère was induced to experiment in the exploratory mode because he did not have a concrete theory of electromagnetism with which to make predictions that could then be constrained by a well determined empirical test. Rather, Ampère had to

[11] Colaço's dissertation *An Investigation of Scientific Phenomena* discusses an alternative mode of exploratory experimentation that does not obviously involve systematic parameter variation. His account highlights the manner in which experimentalists engage in exploratory investigation of a phenomenon in the course of identifying and characterizing the phenomenon in the first place. Although I have no reason to expect conflict, exploring the extent to which Colaço's framework maps onto the strategy discussed here is a task for another occasion.

systemically tinker in order try to determine the nature of the very phenomena with which he was tinkering.

I take it that Steinle is interested in characterizing methodological steps involved in what he identifies as exploratory experimentation, which he stresses can be "highly systematic". Thus, Steinle (2002) writes:

> Far from being a mindless playing around with an apparatus, exploratory experimentation may well be characterized by definite guidelines and epistemic goals. The most prominent characteristic of the experimental procedure is the systematic variation of experimental parameters. The first aim here is to find out which of the various parameters affect the effect in question, and which of them are essential. Closely connected, there is the central goal of formulating empirical regularities about these dependencies and correlations. Typically they have the form of "if–then" propositions, where both the if- and the then-clauses refer to the empirical level. (419)

Steinle evidently has in mind here propositions such as those Ampère might have formulated in his investigations with the magnetic needle, perhaps of the form: If (under certain conditions) the polarity of the battery is *such-and-such*, then the orientation of the needle is *thus* (cf. 413). The generic procedure of exploratory experimentation then, according to Steinle, is to systematically vary the parameters of the experiment to try to elicit regularities, by for instance attending to which parameters are essential for the effect of interest and how different arrangements of the parameters affect its production (ibid., 419). Putting bounds on a parameter is importantly dissimilar from this sort of exploratory experimentation, or so I will now argue.

5.3.3 Against construing putting bounds on a parameter as systematic parameter variation

In the dark energy case, constraints on theorizing are achieved by leveraging data from observables thought to be relevant to the phenomenon in question against a generalized representation of some basic characteristics of the nature of that phenomenon. Are there elements of this case that could be construed as analogous to the systematic variation of the circumstances of an experimental set-up? To accomplish this one would presumably have to recast the notion of parameter variation as something besides the activity of an agent like Ampère systematically altering an experimental setup. We would need an expanded account of exploratory science that included exploratory *observations* in addition to exploratory experimentation via manipulation and intervention. Perhaps one could construe naturally occurring variation in systems of interest as exhibiting analogous parameter variation. We already have the notion of a "natural experiment" wherein conditions found in nature are composed in such a manner that they can be treated as if they had been prepared experimentally (cf. Morgan, 2013). Perhaps we could also recognize "natural exploratory experimentation" wherein natural circumstances have produced systematic variation in the parameters relevant to some effect of interest such that an observational scientists might gain access to the relevant empirical results without intervening on an artificially prepared system.

The circumstances under which dark energy exerts its influence do change in significant and potentially informative ways over the natural course of cosmic evolution. Dark energy is thought to be present all along, but the characteristics of the universe vary immensely. Until about 300,000 years after the big bang, all the matter and radiation in the universe is in the form of an undifferentiated plasma pervading space nearly uniformly. As the universe expands and cools, radiation decouples from matter, both dilute progressively, and eventually the nested structures of starts, galaxies, clusters, and superclusters cohere to form the nodes and filaments of the cosmic web.

Thus, by probing cosmic history, we are able to gain information about the interaction of dark energy in significantly distant contexts. For instance, we are able to investigate its effects when the energy density of the universe was matter dominated and compare that to effects when the energy density is dominated by dark energy itself (our epoch). To go this route would be to construe characteristics of the universe such as composition of the energy density indexed to cosmic time as analogous to "experimental parameters" and to think of the natural evolution of the cosmos as the mechanism that varies such parameters rather than a human scientist in the laboratory. Perhaps we will not be able to identify "a large number" of such parameters (as would be fit an approach closely analogous to Steinle's) but it seems plausible that the more such parameters we could identify, the more we stand to learn about the nature of the phenomenon of interest.

What is there to be gained by construing the strategy of putting bounds on a parameter as analogous to systematic experimental parameter variation? One benefit might be to illuminate the similarities between generating empirical constraints in experimental and in observational sciences. There is a persistent intuition that the epistemology of experimental sciences and that of the (merely) observational sciences is different in a way that makes a difference for the sort of knowledge that we can attain through each (cf. Hacking, 1989). But does this intuition really bear out? Investigating parallels in the manner suggested above could help to address this question.

However, even if drawing parallels with parameter variation turns out to be fruitful, construing the strategy of putting bounds on a parameter as exploring parameter variation would stop short of capturing the distinctive feature of this strategy for generating empirical constraints: efficiently representing families of models within one framework. Indeed there are other clear examples of parameterized representations and family resemblance between these instances is much stronger than that between the dark energy case and Ampère's exploratory experimentation. For instance, the parameterized post-Newtonian formalism (PPN) for representing gravitational theories is another good example that belongs in the family (see Will, 2014, especially sections 3 and 4). Using the PPN formalism, empirical results can be used to put bounds on a collection of parameters ($\gamma, \beta, \xi, \alpha_1, \alpha_2, \alpha_3, \zeta_1, \zeta_2, \zeta_3$, and ζ_4) thereby constraining diverse gravitational theories from general relativity, to Brans-Dicke theory, to $f(R)$ theories, and so on, at the same time (see Figure 6).

Inspired by PPN formalism, cosmologists have recently introduced a parameterized post-Friedmann framework (PPF) for representing gravitational theories relevant for cosmological scales. Baker et al. (2013) express the motivation for such a framework as follows:

> Constraining modified theories on an individual basis is likely to be an infinite process, unless our ingenuity at constructing new theories wanes. We need a fast way to test and rule out theories if we are to drive their population into decline. (1, reference omitted)

The authors stress that "PPF can be used to make statements about unknown regions of theory space in addition to the testing of known theories. Such statements could be of use in guiding model builders to the most relevant regions of theory space" (ibid., 1-2). The formalism allows researchers to put constraints on gravitational theories even when the

Parameter	Effect	Limit	Remarks
$\gamma - 1$	time delay	2.3×10^{-5}	Cassini tracking
	light deflection	2×10^{-4}	VLBI
$\beta - 1$	perihelion shift	8×10^{-5}	$J_{2\odot} = (2.2 \pm 0.1) \times 10^{-7}$
	Nordtvedt effect	2.3×10^{-4}	$\eta_N = 4\beta - \gamma - 3$ assumed
ξ	spin precession	4×10^{-9}	millisecond pulsars
α_1	orbital polarization	10^{-4}	Lunar laser ranging
		4×10^{-5}	PSR J1738+0333
α_2	spin precession	2×10^{-9}	millisecond pulsars
α_3	pulsar acceleration	4×10^{-20}	pulsar $\dot{P}$ statistics
ζ_1	—	2×10^{-2}	combined PPN bounds
ζ_2	binary acceleration	4×10^{-5}	$\ddot{P}_p$ for PSR 1913+16
ζ_3	Newton's 3rd law	10^{-8}	lunar acceleration
ζ_4	—	—	not independent [see Eq. (73)]

Figure 6: Current limits on the PPN parameters, Table 4 from Will (2014, 46)

action corresponding to that theory has not been explicitly written down by anyone. That is, "the PPF framework systematically accounts for allowable extensions to the Einstein field equations, while remaining agnostic about their precise form" (ibid., 3). The authors suggest that the PPF formalism can be used in two "modes", one in which multiple theories that have been specified before had can be constrained simultaneously, and a second in which "one can use the framework as an exploratory tool for model building" (ibid., 20).

Another example from cosmology is the parametrization of models of cosmological inflation. In the very early universe it is thought that spacetime underwent a brief period of exponential expansion. The physics driving this period of expansion is represented in a very general manner, as an unknown scalar field ϕ that evolves in time according to a potential $V(\phi)$. Different models of inflation correspond to different shapes this potential could have. A useful parametrization can be constructed from the potential V and its first and second derivatives with respect to ϕ (written as V_ϕ and $V_{\phi\phi}$ respectively, see Planck Collaboration

(2016b, 14) and references therein):

$$\epsilon_V = \frac{V_\phi^2 M_{Pl}^2}{2V^2} = \epsilon \frac{\left(1 - \frac{\epsilon_1}{3} + \frac{\epsilon_2}{6}\right)^2}{\left(1 - \frac{\epsilon_1}{3}\right)^2} \tag{5.13}$$

$$\eta_V = \frac{V_{\phi\phi} M_{Pl}^2}{V} = \frac{2\epsilon_1 - \frac{\epsilon_2}{2} - \frac{2\epsilon_1^2}{3} + \frac{5\epsilon_1\epsilon_2}{6} - \frac{\epsilon_2^2}{12} - \frac{\epsilon_2\epsilon_3}{6}}{1 - \frac{\epsilon_1}{3}} \tag{5.14}$$

where M_{Pl} is the reduced Planck mass. Figure 7 shows empirical constraints on these "slow-roll" parameters, depicting 68% and 95% confidence regions from data from the Planck satellite as of 2015 and other data sets. Again, by constraining these parameters, cosmologists need not specify particular models of inflation but can rather proceed in the exploratory mode, chipping away at a vast space of viable theories.

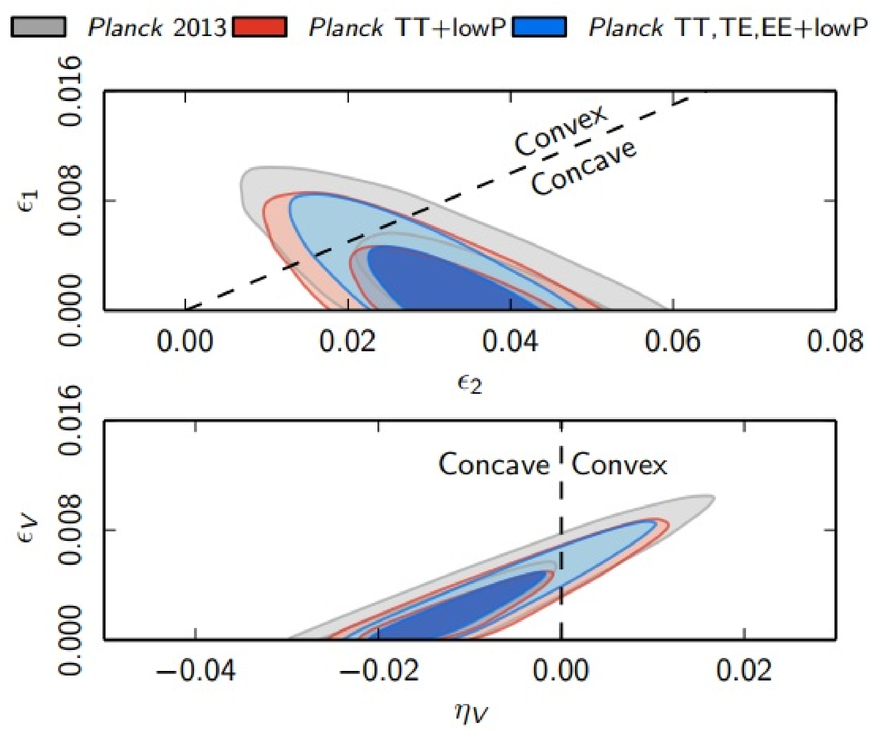

Figure 7: Constraints on slow-roll parameters, Figure 10 from Planck Collaboration (2016b, 14)

Each of the parameterizations that I have discussed in this section—the dark energy equation of state parameter, the PPN and PPF formalisms, and the slow-roll parameterization of models of cosmic inflation—are all examples of species in what I take to be the same genus: putting bounds on a parameter. The strategy is as follows: construct a representation that is generic enough to efficiently cover a range of possible theories and then use empirical results to constrain that space. This approach has the benefit of allowing researchers to constrain theories *even without specifying them*. As a result, this approach can be particularly useful in epistemic contexts where the theoretical resources available regarding the phenomenon of interest do not provide much guidance (yet) about which among the possible theories might be worth pursuing.

5.4 CONCLUDING REMARKS

Different epistemic contexts call for different research strategies. In cases where concrete models or hypotheses can be used to generate specific predictions, effective observational campaigns attempt to measure or rule them out. In cases where no concrete model or hypothesis has been generated for some laboratory phenomenon, research can nevertheless proceed by systematic experimentation. But besides these two there are plenty of contexts for which other strategies are appropriate. In this chapter I have discussed another context, one in which researchers seek to understand some natural phenomenon about which theorizing is still relatively under-developed. In such contexts a useful strategy is to constrain theorizing via generic parameterizations that serve to represent whole families of models. Such parameterized models can be compared with empirical results, scientists can seek to place bounds on those parameters to create a "model in relief" fenced in by a perimeter of empirical results.

There are sure to be many further ways that empirical results can be made to constrain theorizing beyond those I have discussed here. There are at least two other strategies that strike me as worth investigating in further detail, which are be exemplified by the following vignettes:

> **What is dark matter?**
> Dark matter is another relatively mysterious aspect of the contemporary cosmological picture. Again, the name refers to an as yet unknown something, which was originally inferred from aberrant galactic rotation curves suggesting that the orbits of stars at the out reaches of galaxies could not be accounted for by the luminous matter contained in those systems alone. Since the earliest days of dark matter research there have always been several proposals for what dark matter could be, although the specifics have changed over time. At first it was plausible that ordinary dim matter–such as hydrogen gas, dust, or burnt out stars–might account for the anomalous rotation curves. However the contribution of these ordinary denizens was not enough to fully account for the effect and more exotic possibilities began to emerge including modifications to gravitational theory and a slew of weakly interacting massive particles concocted in extensions to the standard model of particle physics. Further empirical research elicited different constraints on the nature of dark matter beyond the original galactic rotation curves at a range of scales, e.g. the motion of galaxy clusters, cosmic mass density maps inferred from surveys of weak gravitation lensing fields, and the Bullet Cluster (the remnant of two colliding clusters of galaxies which appear to have passed through each other, leaving the interacting luminous matter clumped in between while taking their relatively non-interacting dark matter halos with them). The multiplicity of empirical results provide a variety of constraints on the nature

of dark matter–whatever it is, it will have to be consistent with each of them.

The Hubble Deep Field
In 1995, Robert Williams allocated some of his Director's Discretionary Time on the Hubble Space Telescope to observations of a patch of sky relatively free of foreground objects but otherwise thought to be typical. That is, rather than targeting known astronomical objects, the strategy was to literally look where nothing in particular had previously been observed. The researchers did have a broad aim—they were interested generally in the formation and evolution of galaxies—and they wanted to use Hubble to image these processes in the early universe. However there was no specific hypothesis driving this observational research. Rather, they were seeking general results that could be used to constrain theories of galaxy evolution. The long exposures made of this region revealed images of the young universe, depicting an astounding array of ancient galaxies which were too faint to have been observed previously. These early galaxies were found to be more irregular than galaxies are now in our own neighborhood, supporting the idea that galaxy mergers were more frequent in the early universe than today. The success of this strategy lead Hubble researchers to perform several subsequent observations in a similar spirit resulting in the Hubble Deep Field South, the Hubble Ultra-Deep Field, and the eXtreme Deep Field.

In the case of dark matter, more concrete proposals are available but there are a number of rather different options on the table that are underdetermined by the evidence so far. In making observations of different physical systems thought to involve dark matter, researchers can accumulate desiderata to which the nature of any empirically viable dark matter candidate must conform, i.e. more phenomena that proposals must save. The impetus for the imaging the Hubble Deep Field was also born out of a relatively open epistemic context—lots of unanswered questions about the formation and evolution of galaxies over cosmic history. We might tentatively call the two strategies exemplified here *characterizing a phenomenon* and *surveying*. It would be useful and interesting to explore these strategies and others further, to investigate their variety and applicability across the empirical sciences–in experimental as well as observational contexts, in the life sciences as well as the physical sciences.

If a comprehensive enough collection can be articulated, then it may be possible to shift from the primarily descriptive mode to the normative. That is, it may be possible to recommend strategies for generating certain types of constraints given a particular epistemic context. Perhaps philosophy of science could systematize a methodological palette, accompanied by conditional recommendations for the application of its various elements. Of course one might object that scientists do very well on their own without any such suggestions from a philosophical peanut gallery. Most of the time this may well be the case. However, as the recent public debate between scientists around the appropriate methods for scientific investigation of the early universe show, there can come a point at which having the sort of sophisticated philosophical resources advocated here would sure come in handy.

We now have the resources to say something helpful about the sticky situation in early universe cosmology. The epistemic context in inflationary cosmology is much like that in theorizing about dark energy. As mentioned above, a whole family of models of the potential of the "inflaton" field can be characterized by just a few parameters, which can then be constrained by empirical results. This is a way to make progress in constraining theorizing about inflation without abandoning the unique role for empirical data. Given the existence of such a strategy, the move to non-empirical confirmation in early universe theorizing looses its motivation. Even Dawid (2016) admits that when empirical data is available it ought to trump non-empirical considerations:

> eventual empirical testing is always in the background and, once conclusive empirical testing can be achieved, will in all cases make a stronger case for the theory's viability than non-empirical confirmation ever could. (204)

Perhaps one could object that the shift to non-empirical confirmation allows much *more* or *better* progress than that afforded by putting bounds on generic parameters characterizing possible inflaton potentials. Indeed, constraining the shape of the inflaton potential provides precious little insight into the nature of inflation (cf. Baytaş et al., 2015). A committed empiricist will obviously deny that more or better progress could be had with the non-empirical shift. However, I hope to have shown that such a shift would be premature in cosmology at least—there are still empirical constraints to be had in this field.

6.0 CONCLUSIONS: EPISTEMIC ATTITUDES AND PROGRESS

Appreciating the nature of empirical constraints yields a picture of epistemic progress in science. I argue that scientists accumulate empirical evidence and that they also gain knowledge that viable theorizing will be consistent with the evidential corpus, suitably interpreted. In other words, on my view scientific knowledge is not to be understood as knowledge that any particular theory is true, that any particular model will continue to be fruitful, or even knowledge that the truth is somewhere in the landscape of viable theoretical approaches even if it cannot be singled out. Instead, I propose a shift in the target of our usual epistemic commitment away from theories towards empirical evidence and to the constraints that they pose on viable scientific proposals.

Anything realistically recognizable as the result of a modern experiment or observation is mediated by layer upon layer of interpretation. Peeling back these successive layers does not reveal some immutable core that constitutes the pure experiential foundation of the result. This does not mean that there is nothing that is distinctively empirical, but it does mean that whether a result is empirical or not already involves interpretive resources. In cases where aspects of the epistemic resources to be employed in interpreting results change, the relevant layers of an existing result will have to be revisited and and my have to be reinterpreted or reprocessed in order to be put to use. Particularly important among the interpretive layers are those by dint of which the result is understood to be influenced by the target of study in nature since being produced via causal interaction with the target is what differentiates the result as empirical.

The layers I have in mind are associated with significant stages in the actual production of the result in the course of the experiment or observation, the subsequent data analysis, and presentation of the final data products. For instance, there are typically various assumptions about how the experimental or observational apparatus works, informed cuts and other processing that are performed on the raw data, as well as choices regarding how the data should be expressed, depicted, or otherwise packaged. A bare number or a curve printed on a page is useless as a source of empirical content without rich interpretive resources introduced at these stages. Even data records themselves are expressed within particular interpretive frameworks. In unfortunate situations, scientists hoping to re-appraise some data record may not have access to the background information that they would need in order to reinterpret it.

Nevertheless, empirical evidence is durable in the sense that there is an onus on any viable framework to furnish a way of understanding the evidence in a consistent manner. Enriched evidence accumulates over time and is relevant across multiple successive and co-contemporaneous contexts. The accumulating evidential corpus provides continuity between researchers such that even old evidence bears on contemporary models. As the corpus grows, the interpreting scientists have to carefully wend their way through an ever-thickening labyrinth of evidence.

Can the view presented above prove fruitful for characterizing a non-internalist account of scientific progress? In this final chapter, I present the outline of such an account, which crucially involves accumulation of empirical evidence.

Scientific theorizing on a topic consists in a landscape of (usually multiple) currently viable theoretical approaches, or interpretive frameworks, for understanding their subject matter. Often, various details of the viable alternatives will be incompatible with one another while at the same time being consistent with presently available evidence. The boundaries of this landscape at any given time are constrained by the body of empirical evidence that has been accumulated. Progress is driven by the influx of empirical constraints, which change the boundaries of the landscape over time. As new constraints are uncovered, interpretive approaches are rejected, modified and added in light of the growing body of evidence.

I want to be clear that I am not claiming that the accumulation of empirical evidence alone accounts for much of the scientific progress that has been made. Constraining particular models or families of models requires that those models actually be developed, which is the activity of theorists and modelers. However, I do think that amassing evidence does play a particularly important role in scientific progress. Unchecked by empirical constraints, scientists can continue to expand the space of viable theoretical alternatives in a given field. However, without adding further input from nature, they would have no assurance that such advancements would bring them increasing knowledge about what the world is actually like. In contrast, exploratory empirical research does generate knowledge about what the world is like insofar as such research yields new empirical evidence.

It is useful to distinguish between evolving theoretical landscapes and individual modeling lineages. Within a landscape of viable theoretical alternatives, particular scientists or research groups often work on one approach in particular. Guided by a particular approach, scientists develop models with the aim of representing and understanding their subject matter. The constraining influence of empirical evidence on the level of individual proposals can be traced by appreciating the alterations made to viable models over time. These modeling lineages—successions of models developed according to a particular approach and altered in response to new empirical evidence—can be thought of as particular trajectories through the theoretical landscape.

In a diverse epistemic environment, when there are multiple viable modeling approaches, scientists attempt to constrict the landscape by extracting distinguishing features from the competing alternatives, collecting differential empirical evidence and evaluating model features in light of that evidence. In cases where particular models have not been developed, scientists try to construct models that are consistent with known empirical evidence and that could be used to generate further testable inferences. In contrast, exploratory observational research such as the Hubble Deep Field, does not require motivation from any particular modeling or theorizing trajectory, although novel evidence collected via that research could certainly inspire new proposals. Empirical constraints on models add up to empirical constraints on the broader landscape (see Morgan and Morrison, 1999).

Philosophers of science have extensively considered the underdetermination of theories by evidence. However, typically philosophical discussions of underdetermination have been guided by—and perceived of as having consequences for—worries about which particular scientific theories (if any) we should believe to be true. Underdetermination is perceived as a problem standing in the way of picking out the one true theory from a crowd of empirically equivalent doppelgangers. This focus has obscured the fact that the existence of multiple viable theoretical alternatives at any given point in time is absolutely germane to scientific practice.

Scientists typically adopt a particular theoretical approach from among the viable alternatives, which they see as promising. Students being trained in a research group will often be schooled in the favorite approach of the group, which has significant effects on the particular types of research activities with which they engage. However, I think that it is important to distinguish this kind of tentative adoption of frameworks from belief in a particular theory. The epistemic attitudes that scientists have towards their working theoretical approaches are less committal and more subtle than belief. These attitudes leave room for the reassessment of alternatives by stepping back to re-engage with the larger landscape of viable alternatives. Moreover, the tentative epistemic attitudes leave room for the modification of the working framework in light of new evidence.

Let us consider the epistemic attitudes appropriate to models, modeling lineages, and whole theoretical landscapes. Scientists are not particularly epistemically attached to individual models—and for good reason. Individual models are often altered in response to empirical constraints, sometimes forming long-term modeling lineages, sometimes being abandoned altogether. Models are adopted tentatively, and modeling lineages are pursued (or at least kept alive in the background) until empirical constraints are so severe as to render them untenable. That is to say, the epistemic attitude appropriate to particular theoretical approaches (the MACHO dark matter hypothesis say) is something short of belief.

It seems to me that the epistemic attitude appropriate to live theoretical options is captured by regarding those options as empirically viable. This attitude is perhaps best described as tolerance. It would be a mistake to wed ourselves to one of our current theories. We have every reason to believe that we have not exhaustively collected all of the empirical evidence that the world has to offer and it may turn out that such evidence will constrain our present theories in unanticipated ways. Indeed, if we were to abandon the kind of epistemic caution that I am advocating and endorse our current best theories, it would be difficult to make any sense of the motivation for cases in which research aims to build up new theoretical resources and to put them in contact with new empirical constraints in order to learn about some part of nature that is not yet well understood.

Of course the non-epistemic attitudes that scientists have towards theories can be quite diverse and very preferential. For instance, a scientist may favor a particular approach due to its novelty, elegance, facility, familiarity, etc. Indeed, some scientists have such strong preferences for particular approaches that they spend their entire careers developing and defending them. However, these factors are irrelevant to epistemic commitment. Whether an approach should be tolerated or not is determined by it compatibility with the available evidential corpus.

Do (or should) scientists somehow place their belief in the entire landscape instead? I do not think so. Scientists certainly do not believe all of the various approaches in a landscape simultaneously. The landscape is the repository of approaches which are still deemed appropriate for further work. Rather, what comes to be known is the evidence—the dynamic boundary molding and delimiting viable interpretations, which has been gathered from observations and experiments. Beyond the mere accumulation of empirical evidence, scientists also come to know that any realistic understanding of the part of the world being investigated will have to be consistent with that evidence. Thus, rather than coming to know specific theories, scientists come to know the outlines of possible theories. Nevertheless, in filling in the outlines, scientists continually learn about what the natural world is really like.

Objection: empiricism is unproductive One might worry that adopting the relatively conservative epistemic attitude of tolerance towards our most successful scientific theories is misguided since taking those theories seriously is a productive way to encourage theoretical development. My response to this objection is that taking the content of our best theories seriously as a heuristic for developing new theories is perfectly compatible with my view. In particular, my view is very liberal with respect to the reasons particular scientists (or scientific communities) have for pursuing and developing one theory over another. Consider dark matter research. A particular group (ADMX) takes the proposal that that the theoretical particle the axion could be galactic dark matter so seriously that they built an apparatus to try to detect such axions interacting with photons of the field generated by a 10T superconducting solenoid. The energy ranges they look at, and the required sensitivity of the experiment is determined by what they would expect to observe if the galactic dark matter were axions. Results of this experiment will be significant for the development of particle physics and theories of dark matter, insofar as they will provide new empirical constraints. I could readily tell a similar story about gravity experiments, neutrino experiments, etc. These research programs are predicated on taking the theoretical proposals seriously as viable options. Similarly, the conceptual elaboration of these proposals, and the development of their attending models is also predicated on taking them seriously as viable candidates. However, "taking the proposals seriously" for the purpose of developing them and testing them, need not (and indeed should not) amount to believing them to be true. One can perfectly well pursue a theory without endorsing it with your epistemic commitment.

To summarize: progress consists in the accumulation of empirical evidence, but this evidence does not have a single fixed interpretation. Furthermore, the accumulation of evidence in this sense furnishes knowledge about what the natural world is like in the sense that any viable theory of the natural world will have to accommodate the evidential corpus.

This view is not a variety of instrumentalism or of foundationalist empiricism. In particular, my view is different from an instrumentalist who thinks that scientific theories only aim to match predictions and empirical results—by whatever means. For such an instrumentalist, a parameterized model whose components have no suspected physical correlates would be just as acceptable as a model whose components are taken to represent objects and processes in nature. I think this view is both descriptively inadequate with respect to scientific practice and ill-advised as a method for learning about nature. In contrast, scientists in fact try to propose physically plausible theories, they aim at representational fidelity.

I think that one could reasonably be an instrumentalist in the following sense: certain "patches" are introduced in the course of science with the aim of expediently generating predictions, which scientists know full well are not intended to be representationally faithful. However, it is important that these patches be carefully documented and kept track of, so as not to be mistaken for physically plausible representations. Similarly, I think that there are perfectly good uses of instrumental models in many domains of science (especially the applied sciences) such as for making weather predictions. However, in the context of basic science research, theories are intended as candidate representations.

Furthermore, my account of empirical evidence is far more permissive than that associated with traditional empiricists. In particular, my view is different from that of an empiricist who thinks that we can only have knowledge about that which we can observe with our unaided senses. It is clear that empirical constraints are interpreted using the conceptual resources of theory. However I want to stress that this fact does not prevent the accumulation of empirical evidence across theory change. Indeed, as I have already said, when scientists have access to information about the methods by which empirical constraints were generated, they can often repurpose old results in the context of new theoretical resources.

It may be useful to contrast the present view with those espoused by Larry Laudan and Bas van Fraassen. Laudan (1977) discusses different types of epistemic commitment that researchers can have to their models, identifying what he calls the context of acceptance and the context of pursuit (108-109). He suggests that scientists should accept theories (treat them as if they were true) if they have the highest problem-solving adequacy. For Laudan, progress in science is made by choosing theories or research traditions that are better solvers of empirical and conceptual problems. He also suggests that scientists may rationally choose to pursue a theory that they would not accept, if the theory has a higher rate of increasing problem-solving adequacy than its competitors.

My methodological bar is somewhat lower than Laudan's in that on my account, scientists may rationally choose to work on a theoretical approach that is viable given the current evidential corpus—or even that they suspect can be developed to become viable via physically plausible modifications. In addition, on my view progress in science does not consist in moving from acceptance of one theory to acceptance of another theory that solves more empirical and conceptual problems. Rather than understanding progress in terms of the particular (viable) frameworks that scientists choose to work within, I suggest that progress is better understood in terms of the accumulation of empirical evidence. My view is not internalist in the manner that Laudan's view is, because I think that the growing stockpile of empirical constraints is relevant to scientists working in different research traditions. Even if a particular observation solves a local empirical problem for a scientist working within a particular research tradition, that is not the full extent of its epistemic import. The epistemic impact of the evidence lingers long after individual research traditions are buried.

My view shares more ground with van Fraassen's constructive empiricism. Like van

Fraassen, I do not think that scientists are ever in a position to believe a particular theory as "the one true theory". Rather, at any given time, scientists have a repository of theoretical resources that they judge to be viable relative to the available evidential corpus. The appropriate epistemic attitude towards such viable theories is not belief.

However, my view differs from van Fraassen's in two important respects. First, my account of empirical accessibility is much more permissive than his. van Fraassen was wrong to cash out empirical access in terms of what is observable via the unaided human senses. According to van Fraassen (1980), "X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it" (16). Much of the subject matter of modern science is not observable in this sense (see Churchland and Hooker, 1985; Hacking, 1985; Teller, 2001). However, an empiricist need not maintain that the only knowledge that can be gained is about objects and events that are observed by unaided human senses. We have an understanding of empirical access that encompasses sophisticated detection methods and instruments such as bolometers and gravitational wave interferometers. I think that we can learn about (for example) what happens at the center of the sun, where presumably van Frassen's definition of "observable" fails. Scientists have gathered empirical evidence (e.g. from solar neutrinos) that constrains viable theorizing about the center of the sun and the processes and substances therein.

Second, on my view the commitment that scientists have to the landscape of viable theoretical approaches is not as strong as van Fraassen's. For van Fraassen, scientists accept theories they believe to be empirically adequate, i.e. theories that they believe are consistent with currently available evidence and will continue to be adequate to all future evidence (ibid., 12). This notion of acceptance is so strong that I doubt that very many scientists hold this sort of attitude towards even their most favorite theories. In contrast, when scientists do manage to construct long-lasting theories, I think we should expect that they are likely to engage in exploratory research to see if a research front can be re-ignited. On my view, scientists are committed to the durability of the empirical evidence accumulated so far, even if the interpretation of that evidence will have to be modified in the future.

My approach also shares some similarity with Popper's falsificationism, in the sense that the landscape of viable theories is pruned by empirical evidence and agreement of some theory with the evidential corpus does not warrant belief that it is true. One important sense in which my view differs from Popper's however, is that I do not think that scientific methodology properly consists in bold conjectures handed from theorists to experimentalists bent to try to falsify them. Indeed, there are episodes that contribute to scientific progress in which surprising empirical evidence has motivated new theoretical approaches (e.g. supernovae observations motivating theorizing about the accelerating expansion of the universe). In fact, some exploratory research can be characterized as open-ended investigation employed for the purpose of generating new empirical constraints on viable theorizing.

In this dissertation I have laid the groundwork for an empiricist epistemology of science that is applicable to science in practice. I have articulated the core commitments of empiricism in philosophy of science, presented an argument for what makes data distinctively empirical, introduced an account of empirical evidence that embraces theory-ladenness, and I have explored how this account of evidence affects what it means to adjudicate empirical adequacy and what resources are required for this task. In addition, I have begun to address the nature and variety of empirical constraints on theorizing, discussing a type of exploratory research that is importantly dissimilar to hypothesis testing, in which empirical constraints are generated in a context where available theoretical resources yield little insight into the specific nature of the worldly target of interest. That is, I have discussed a manner in which theorizing can be constrained by empirical evidence when the theory to be constrained is very much under development. Taken together, these arguments trace out a story about the epistemology of science that flows from the natural world, through empirical data and its processing, to arrive at the use and reuse of data products for generating constraints on theorizing.

There are two senses in which this work is about scientific progress at the boundaries of experience. First, I hope to have shown how scientific progress happens by adding new evidence to the evidential corpus—that is, how progress happens by enlarging the boundaries of "experience" via increasing epistemic access to nature. Second, in developing the components of this empiricist philosophy of science, I have hoped to crystallize the resources that an empiricist needs to say how science functions well in the far reaches of our experience. How is it that human inquirers make genuine epistemic progress in learning about parts of the cosmos distant in space and time? My answer to this question resists the temptation of a drastic epistemic shift in scientific methodology that purports to take scientific inquiry beyond the boundaries of experience by utilizing non-empirical virtues like parsimony and explanatory power to guide theory choice once empirical access runs dry. Instead, my stance is that we make genuine epistemic progress at the boundaries of experience in the ordinary way: by adding to the evidential corpus. This progress happens at the boundaries of experience to be sure, but not beyond them.

APPENDIX

ENRICHED EVIDENCE FROM THE HULSE-TAYLOR PULSAR

In this appendix I will illustrate my account of enriched lines of evidence with a concrete example. This example exhibits the different components of enriched lines of evidence and highlights why the presuppositions incorporated throughout data collection and processing determine the epistemic utility of the empirical results thereby generated. The generic structure to be made concrete is depicted in Figure 8. Data collection generates data records, which are transformed by data processing into (often a series of) processed data/data products, and finally to an empirical constraint tailored to some particular theoretical context. We can refer to the data records, processed data, and empirical constraints all as "empirical results".

To achieve their celebrated confirmation of general relativity, Taylor and Weisberg (1982) had to introduce an extensive series of data processing and analysis stages in order to transform the receiver signal from the Arecibo telescope into a estimate of the parameter value of interest (in this case $\dot{P}_b$ the orbital rate of decay of the astronomical object PSR 1913+16) and then to the claim that gravitational radiation exists. Without belaboring the point, it is worthwhile to examine this case in slightly more detail since it displays two widespread (although not universal) features of data processing well: 1) the intricate confluence of presuppositions (both empirical and theoretical) required to produce a final result and 2) the necessity of invoking the theory to be tested in the course of data processing.[1]

Note that the empirical value of $\dot{P}_b$ that Taylor and Weisberg (1982) compare to the

[1] cf. Glymour (1975)

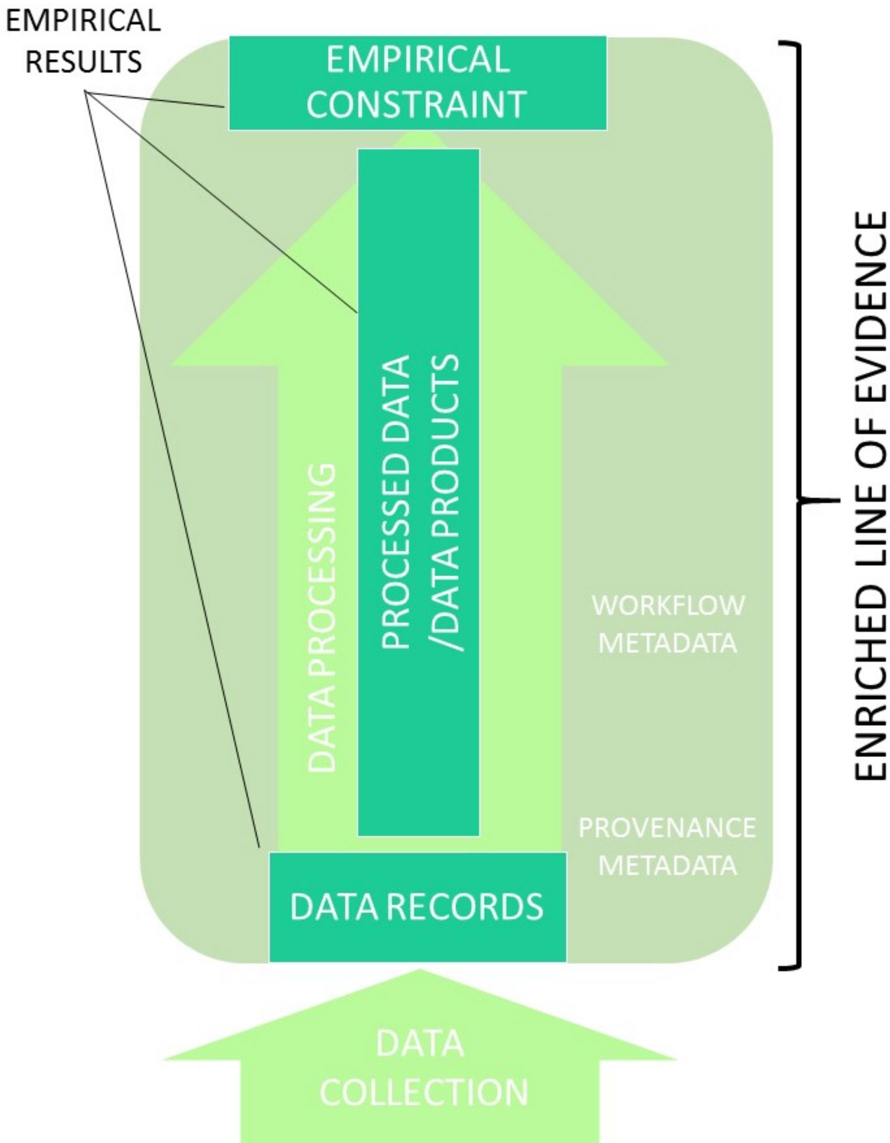

Figure 8: Elements of an enriched line of evidence

theoretical value is the result of a calculation that presupposes the general theory of relativity. In particular, the model of the stellar system to which the empirical pulse arrival times are fit is built on the assumption that general relativity is the "correct" theory of gravity (Taylor and Weisberg, 1982, 911). Yet Taylor and Weisberg were interested in using the

Hulse-Taylor pulsar results to constrain theories of gravitation besides general relativity including those of Brans-Dicke, Rosen, Ni, and Lightman-Lee [916, 917]. Parameterized versions of these alternate theories were used to generate predictions regarding the value of $\dot{P}_b$. However, Taylor and Weisberg did *not* calculate the masses of the binary pulsar system (which are requisite intermediary results on the way to the final estimate for the orbital decay rate parameter) assuming each different theory independently, but rather just used those calculated assuming GR. One might well wonder: is this epistemically admissible? Is it reasonable to assume GR for the purposes of testing theoretical competitors in this context?

If assuming GR is not epistemically admissible, then the pedantic work of calculating the masses of the binary system objects actually needs to be carried out for each alternative theory. In fact, since the Blandford-Teukolsky-Epstein model of the pulsar system assumes GR, and since relativistic corrections are employed in transforming the pulse times to the pulsar frame, one would have to start all the way back at the pulse times as recorded in the receiver frame and re-analyze these using the infrastructure of each theory under consideration. Whether or not such reanalysis is possible in practice will depend on the availability of the original data records–which could amount to a significant logistical problem.

Taylor and Weisberg (1982) provide enough information in their publication for us to be able to outline their data processing workflow beginning with the data records of pulses received by the radio telescope (cf. Taylor and Weisberg, 1982, 911). The pulses picked up by the radio telescope were so weak that in order to obtain a high enough signal-to-noise ratio, for some datasets the effect of dispersion (due to the interaction of signal with free electrons on its long journey from the pulsar to the telescope) was compensated for in hardware and about 5000 pulses averaged together in a block, all before the data are ever recorded Taylor and Weisberg (1982, 908-909). The recorded arrival times of the (averaged) pulses from ten different data sets from 1974-1981 were transformed from the receiver frame of reference to that of the solar system barycenter using planetary positions estimated from eighth-order interpolation of Lincoln Laboratory ephemeris. These times were then transformed to the pulsar frame, including second-order relativistic terms and gravitational propagation delay due to the presence of the companion star using estimates made in earlier research. The transformed pulse arrival times were used to calculate values for the pulse phase ϕ. A particular relativistic model of the pulsar system, the Blandford-Teukolsky-Epstein model (which assumes that the pulsar is an intrinsically accurate clock and that the two stars can be modeled dynamically as point masses), was fit to seven of the data sets using the least squares method with the residuals calculated for the pulse phase. In light of that fit a set of values for the model parameters were estimated, including $\dot{P}_b$, the rate of change of the orbital period of the pulsar. Using a subset of the parameters, values for the masses of the astronomical objects were computed. These masses, together with some of the other parameter values, were used to calculate the expected (theoretical) value, $(-2.403 \pm 0.005) \times 10^{-12}$, of $\dot{P}_b$ via the general relativistic quadrupole formula. This value was then compared to what the authors call the the "measured" (i.e. estimated empirical) value for $\dot{P}_b$: $(-2.30 \pm 0.22) \times 10^{-12}$. As one can see, the empirical value includes the expected value within its error bars.

BIBLIOGRAPHY

Albrecht, A., L. Amendola, G. Bernstein, D. Clowe, D. Eisenstein, L. Guzzo, C. Hirata, D. Huterer, E. Kolb, and R. Nichol. Findings of the Joint Dark Energy Mission Figure of Merit Science Working Group. https://arxiv.org/abs/0901.0721.

American Astronomical Society (2017, March 22). With Astronomy Rewind, Citizen Scientists Bring Zombie Astrophotos Back to Life (Press Release). https://aas.org/media/press-releases/astronomy-rewind.

Anderl, S. (2016). Astronomy and Astrophysics. In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science*, pp. 652–670. Oxford University Press.

Baker, T., P. G. Ferreira, and C. Skordis (2013). The parameterized post-Friedmann framework for theories of modified gravity: Concepts, formalism, and examples. *Physical Review D 87*(2), 1–24.

Baytaş, B., A. Kesavan, E. Nelson, S. Park, and S. Shandera (2015, Apr). Nonlocal bispectra from super cosmic variance. *Physical Review D 91*, 083518.

Bhakthavatsalam, S. and N. Cartwright (2017). What's so special about empirical adequacy? *European Journal for Philosophy of Science 7*, 445–465.

Bird, A. (2007, mar). What Is Scientific Progress? *Nous 41*(1), 64–89.

Bird, A. (2008, jun). Scientific progress as accumulation of knowledge: a reply to Rowbottom. *Studies in History and Philosophy of Science Part A 39*(2), 279–281.

Bogen, J. and J. Woodward (2005). Evading the IRS. In M. R. Jones and N. Cartwright (Eds.), *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences (Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 86)*, pp. 233–267. Amsterdam/New York: Rodopi.

Burke-Spolaor, S., M. Bailes, R. Ekers, J.-P. Macquart, and F. Crawford III (2011, jan). Radio Bursts with Extragalactic Spectral Characteristics Show Terrestrial Origins. *The Astrophysical Journal 727*(1), 18.

Chang, H. and G. Fisher (2011). What the Ravens Really Teach Us: the Intrinsic Contextuality of Evidence. In P. Dawid, W. Twining, and M. Vasilaki (Eds.), *Proceedings of the*

British Academy 171: Evidence, Inference and Enquiry, pp. 345–370. Oxford University Press.

Churchland, P. M. and C. A. Hooker (Eds.) (1985). *Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen*. University of Chicago PressChicago, IL.

Clark, D. H. and F. R. Stephenson (1977). *The Historical Supernovae*. Oxford: Pergamon Press.

Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA (2007). *Environmental Data Management at NOAA: Archiving, Stewardship, and Access*. Washington, D.C.: National Academies Press.

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington, D.C.: The National Academies Press.

Committee on the Preservation of Geoscience Data and Collections (2002). *Geoscience Data and Collections: National Resources in Peril*. Washington, D.C.: The National Academies Press.

Dawid, R. (2016). Modelling Non-empirical Confirmation. In E. Ippoliti, R. Sterpetti, and T. Nickles (Eds.), *Models and Inferences in Science: Studies in Applied Philosophy, Epistemology and Rational Ethics 25*, pp. 191–205. Springer.

Duhem, P. (1954/1974). *The Aim and Structure of Physical Theory*. Princeton University Press.

Fitzgerald, A. P. (1953). Transits of Mercury. *Irish Astronomical Journal 2*(7), 203–209.

Franklin, A. (1995, apr). The appearance and disappearance of the 17-keV neutrino. *Reviews of Modern Physics 67*(2), 457–490.

Franklin, A. (2002). *Selectivity and Discord*. University of Pittsburgh Press.

Franklin, A. (2015). The Theory-Ladenness of Experiment. *Journal for General Philosophy of Science 46*(1), 155–166.

Franklin, A. D. (1981). Millikan's Published and Unpublished Data on Oil Drops. *Historical Studies in the Physical Sciences 11*(2), 185–201.

Franklin, L. R. (2005). Exploratory Experiments. *Philosophy of Science Proceedings of the 2004 Biennial Meeting of The Philosophy of Science Association Part I: Contributed Papers 72*(5), 88–899.

Galison, P. (1987). *How Experiments End*. University of California Press.

Gibney, E. (2017, March 24). Citizen scientists to rescue 150 years of cosmic images: Long-lost images could offer insight into rare and moving stars. http://www.nature.com/news/citizen-scientists-to-rescue-150-years-of-cosmic-images-1.21702.

Glymour, C. (1975). Relevant Evidence. *The Journal of Philosophy 72*(14), 403–426.

Goodman, A., A. Pepe, A. W. Blocker, C. L. Borgman, K. Cranmer, M. Crosas, R. Di Stefano, Y. Gil, P. Groth, M. Hedstrom, D. W. Hogg, V. Kashyap, A. Mahabal, A. Siemiginowska, and A. Slavkovic (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Computational Biology 10*(4).

Green, D. A. (2015). Historical Supernova Explosions in Our Galaxy and Their Remnants. In W. Orchiston, D. A. Green, and R. Strom (Eds.), *New Insights From Recent Studies in Historical Astronomy: Following in the Footsteps of F. Richard Stephenson, Astrophysics and Space Science Proceedings 43*. Springer International Publishing.

Guy, J., M. Sullivan, A. Conley, N. Regnault, P. Astier, C. Balland, S. Basa, R. Carlberg, D. Fouchez, D. Hardin, I. Hook, D. Howell, R. Pain, N. Palanque-Delabrouille, K. Perrett, C. Pritchet, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. Ellis, S. Fabbro, H. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. Graham, E. Hsiao, T. Kronborg, C. Lidman, A. Mourao, S. Perlmutter, P. Ripoche, N. Suzuki, and E. Walker (2011). The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *Astronomy and Astrophysics 523*, A7.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.

Hacking, I. (1985). Do We See through a Microscope? In P. M. Churchland and C. A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen*, pp. 132–152. University of Chicago PressChicago, IL.

Hacking, I. (1989). Extragalactic Reality: The Case of Gravitational Lensing. *Philosophy of Science 56*(4), 555–581.

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

Howlett, P. and M. S. Morgan (Eds.) (2010). *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*. Cambridge University Press.

Hubble, E. (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), 168–173.

Huber, P. J. and S. De Meis (2004). *Babylonian Eclipse Observations From 750 BC to 1 BC*. Associazione Culturale Mimesis.

Ijjas, A., P. J. Steinhardt, and A. Loeb (2017). POP Goes the Universe. *Scientific American 316*(2), 32–39.

Jenni, P., M. Nessi, M. Nordberg, and K. Smith (2003). *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN.

Kaiser, M. (1991). From Rocks to Graphs–The Shaping of Phenomena. *Synthese 89*, 111–133.

Karaca, K. (2013). The Strong and Weak Senses of Theory-Ladenness of Experimentation: Theory-Driven versus Exploratory Experiments in the History of High-Energy Particle Physics. *Science in Context 26*(01), 93–136.

Kragh, H. (2014, may). Testability and epistemic shifts in modern cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 46*, 48–56.

Krause, E. et al. (2017). Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses. *Submitted to: Phys. Rev. D*.

Kuhn, T. S. (1975). *The Structure of Scientific Revolutions* (Fourth ed.). University of Chicago PressChicago, IL.

Kukla, A. (1990). Ten Types of Scientific Progress. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association One: Contr*, 457–466.

Laudan, L. (1977). *Progress and Its Problems: Towards a Thoery of Scientific Growth.* University of California Press.

Laudan, L. (1996). *Beyond Positivism and Relativism: Theory, Method and Evidence*. Westview Press.

Laymon, R. (1988). The Michelson-Morley Experiment and the Appraisal of Theories. In A. Donovan, L. Laudan, and R. Laudan (Eds.), *Scrutinizing Science: Empirical Studies of Scientific Change*, pp. 245–266. Baltimore and London: The Johns Hopkins University Press.

Leonelli, S. (2009). On the Locality of Data and Claims about Phenomena. *Philosophy of Science* (5), 737–749.

Leonelli, S. (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), 503–514.

Leonelli, S. (2014). Data Interpretation in the Digital Age. *Perspectives on Science 22*(3), 397–417.

Leonelli, S. (2015). What Counts as Scientific Data? A Relational Framework. *Philosophy of Science 82*, 810–821.

Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Case Study*. The University of Chicago Press.

Lipton, Peter. (2015). Empiricism, history of. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition ed.)., pp. 567 – 70. Oxford: Elsevier.

LSUNews. Front-line Astronomy from Century-old Archives (Press Release). http://www.phys.lsu.edu/recurrentnova/RNpressrelease.pdf.

Marcum, J. A. (2010, jun). Horizon for Scientific Practice: Scientific Discovery and Progress. *International Studies in the Philosophy of Science 24*(2), 187–215.

McDougall, I. and T. M. Harrison (1999). *Geochronology and Thermochronology by the 40Ar/39Ar Method* (2nd ed.). Oxford University Press.

Miller, M. (2016). Mathematical Structure and Empirical Content. http://philsci-archive.pitt.edu/12678/.

Mizrahi, M. (2013, nov). What is Scientific Progress? Lessons from Scientific Practice. *Journal for General Philosophy of Science 44*(2), 375–390.

Mizrahi, M. and W. Buckwalter (2014, jan). The Role of Justification in the Ordinary Concept of Scientific Progress. *Journal for General Philosophy of Science 45*(1), 151–166.

Morgan, M. S. (2013). Nature's Experiments and Natural Experiments in the Social Sciences. *Philosophy of the Social Sciences 43*(3), 341–357.

Morgan, M. S. and M. Morrison (Eds.) (1999). *Models as Mediators*. Cambridge University Press.

Morrison, M. (2009, jan). Models, measurement and computer simulation: the changing face of experimentation. *Philosophical Studies 143*(1), 33–57.

Morrison, M. (2015). *Reconstructing Reality: Models, Mathematics, and Simulations*. Oxford University Press.

Nietzche, F. (1882/2001). *The Gay Science: With a Prelude in German Rhymes and an Appendix of Songs*. Cambridge University Press.

Niiniluoto, I. (2014, jun). Scientific progress as increasing verisimilitude. *Studies in History and Philosophy of Science Part A 46*, 73–77.

Norton, J. D. (2003). Causation as Folk Science. *Philosophers' Imprint 3*(4).

Palmer, C. L., N. M. Weber, and M. H. Cragin (2011). Analytic potential of data: assessing reuse value. *Proceedings of the 11th annual international {ACM}/{IEEE} joint conference on Digital libraries*, 425–426.

Parker, W. S. (2017). Computer Simulation, Measurement, and Data Assimilation. *British Journal for the Philosophy of Science 68*, 273–304.

Perlmutter, S. (2003). Supernovae, Dark Energy, and the Accelerating Universe. *Physics Today April*, 53–60.

Perovic, S. (2017). Experimenter's regress argument, empiricism, and the calibration of the large hadron collider. *Synthese 194*(2), 313–332.

Petroff, E. et al. (2015, Jun). Identifying the source of perytons at the Parkes radio telescope. *Monthly Notices of the Royal Astronomical Society 451*(4), 3933–3940.

Phillips, S. S. (2003). A History of the Evidence. In *Evidence*. New York: Distributed Art Publishers, Inc.

Planck Collaboration (2016a). Planck 2015 results XIII. Cosmological parameters. *Astronomy & Astrophysics 594*(A13), 1–63.

Planck Collaboration (2016b). Planck 2015 results. XX. Constraints on inflation. *Astronomy and Astrophysics 594*(A20).

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, H. C. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, J. M. Diego, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, M. Giard, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, F. K. Hansen, D. Hanson, D. L. Harrison, A. Heavens, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, J. Knoche, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, N. Mandolesi, A. Mangilli, A. Marchini, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. Mazzotta, P. McGehee, P. R. Meinhold, A. Melchiorri, L. Mendes, A. Mennella, M. Migliaccio, S. Mitra, M. A.

Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, A. Narimani, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, V. Salvatelli, M. Sandri, D. Santos, M. Savelainen, G. Savini, B. M. Schaefer, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, L. Valenziano, J. Valiviita, B. Van Tent, M. Viel, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, D. Yvon, A. Zacchei, and A. Zonca (2015). Planck 2015 results. XIV. Dark energy and modified gravity. *Astronomy & Astrophysics 594*(A14), 1–31.

Popper, K. (1959). *The Logic of Scientific Discovery.* London and New York: Routledge Classics.

Quine, W. V. (1951). Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review 60*(1), 20–43.

Railton, P. (1981). Probability , Explanation , and information. *Synthese 48*(2), 233–256.

Reich, Eugenie Samuel (2012, April 2). Embattled neutrino project leaders step down: No-confidence vote follows confirmation of faults in experiment's cable and clock. https://www.nature.com/news/embattled-neutrino-project-leaders-step-down-1.10371.

Rowbottom, D. P. (2008, jun). N-rays and the semantic view of scientific progress. *Studies in History and Philosophy of Science Part A 39*(2), 277–278.

Rowbottom, D. P. (2015, mar). Scientific progress without increasing verisimilitude: In response to Niiniluoto. *Studies in History and Philosophy of Science Part A 51*, 100–104.

Schaefer, B. E. (2010). Comprehensive Photometric Histories of All Known Galactic Recurrent Novae. *The Astrophysical Journal Supplement Series 187*(2), 275–373.

Shapere, D. (1984). *Reason and the Search for Knowledge: Investigations in the Philosophy of Science.* D. Reidel Publishing Company.

Steinhardt, P. and N. Turok (2002). A Cyclic Model of the Universe. *Science 296*, 1436–1439.

Steinhardt, P. J. and N. Turok (2007). *Endless Universe: Beyond the Big Bang.* New York: Doubleday.

Steinle, F. (1996). Entering New Fields: Exploratory Uses of Experimentation. *Source: Philosophy of Science Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers 64*(May), 65–74.

Steinle, F. (2002). Experiments in History and Philosophy of Science. *Perspectives on Science 10*(4), 408–432.

Stephenson, F. R. and D. H. Clark (1978). *Monographs on Astronomical Subjects: 4, Applications of Early Astronomical Records*. New York: Oxford University Press.

Stephenson, F. R. and D. A. Green (2002). *Historical Supernovae and Their Remnants*. Oxford: Clarendon Press.

Stephenson, F. R. and L. V. Morrison (1995). Long-Term Fluctuations in the Earth's Rotation: 700 BC to AD 1990. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 351*(1695), 165–202.

Stephenson, F. R., L. V. Morrison, and C. Y. Hohenkerk (2016a). Measurement of the Earth ' s rotation : 720 BC to AD 2015 Subject Areas :. *Proceeding of the Royal Society A 472*(2196).

Stephenson, F. R., L. V. Morrison, and C. Y. Hohenkerk (2016b). Measurement of the Earth ' s rotation : 720 BC to AD 2015 Subject Areas : The Supplement. *Proceeding of the Royal Society A*.

Stojanović, M. (2013). Exploratory experimentation and taxonomy of experimentation. *Filozofija i drustvo 24*(4), 199–217.

Tal, E. (2013). Old and New Problems in Philosophy of Measurement. *Philosophy Compass 8*(12), 1159–1173.

Taylor, J. H. and J. M. Weisberg (1982). A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16. *Astrophysical Journal 253*, 908–920.

Teller, P. (2001). Whither Constructive Empiricism? *Philosophical Studies 106*(1-2), 123–150.

Turok, N. (2003). The Ekpyrotic Universe and Its Cyclic Extension. In G. W. Gibbons, E. P. S. Shellard, and S. J. Rankin (Eds.), *The Future of Theoretical Physics and Cosmology: Celebrating Stephen Hawking's 60th Birthday*, pp. 781–800. Cambridge: Cambridge University Press.

van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon Press.

van Fraassen, B. C. (1984). Theory Comparison and Relevant Evidence. In J. Earman (Ed.), *Minnesota Studies in the Philosophy of Science: Testing Scientific Theories*, pp. 27–42. University of Minnesota Press.

van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Clarendon Press.

van Fraassen, B. C. (2012). Modeling and Measurement: The Criterion of Empirical Grounding. *Philosophy of Science 79*(5), 773–784.

van Helden, A. (1976). The Importance of the Transit of Mercury of 1631. *Journal for the History of Astronomy 7*.

Weinberg, D. H., M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo (2013). Observational probes of cosmic acceleration. *Physics Reports 530*(2), 87–255.

Will, C. M. (2014). The confrontation between general relativity and experiment. *Living Reviews in Relativity 4*.

Woodward, J. (2004). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Woodward, J. F. (2011). Data and phenomena: a restatement and defense. *Synthese 182*, 165–179.