Gaining Access to the Early Universe

Chris Smeenk

March 30, 2018

1 Introduction

Particle physics and cosmology each have a Standard Model consistent with an astonishing array of observations and experimental results. Both Models are, to some extent, victims of their own success: there are few clear empirical anomalies that could serve as signposts guiding physicists' next steps. The reach of experiments and observations barely extends past the domains they cover, providing few glimpses of the undiscovered country of novel phenomena. There are theoretical accounts of this territory. Yet the almost complete lack of empirical access to this domain makes it difficult to determine whether these are reliable. Research programs such as string theory and eternal inflation have been successful in one sense – winning widespread acceptance in the relevant communities – without the record of correct novel predictions that is often taken as a prerequisite for empirical success. Trust in these proposals is seen as either justified by a reasonable extension of scientific methodology, in light of changed circumstances, or a sign that physicists have ventured into a worrisome new phase of "post-empirical" — or even "post-scientific" — inquiry.

Debates regarding the status of such theories reflect a fundamental disagreement regarding what constitutes success, and how to establish it. A commitment to empiricism is often taken to imply that how well a theory fits the data is the only relevant factor in assessing its truth, or to what degree current evidence confirms it.[1] Competing accounts of confirmation take a broader view, and allow for what is sometimes called "indirect" confirmation, based on factors other than compatibility with the data.

One motivation for these broader views is that scientists arguably do take a variety of factors over and above successful predictions into account in assessing theories; they demand that theories do much more than merely fit the data. For example, the theory should be compatible with other relevant theories. Direct evidence for one theory may extend outward to support several related theories.[2] It is more contentious whether features like elegance should be taken into consideration. An advocate of "elegance" owes us at least an account of how to assess theoretical elegance, along with an argument that questions of elegance are not best left to tailors, as Einstein quipped. The challenge is to provide such a defense, without tacitly assuming that the world has an order or structure that we will find beautiful. One central question in debates about what I will call "supra-empirical" physics regards how much weight scientists should give to a theory's

[1] There is a contrast between the way (most) philosophers and scientists use "confirm." Here I will follow the philosophers: evidence in favor of a theory "confirms" it, even if it only leads to an incremental boost in confidence. On this usage, confirmation admits degrees, whereas scientists often use "confirm" as a success term — applied only to cases of extremely strong evidence. (Readers who adopt the latter usage should substitute "incrementally confirm" for "confirm" throughout.)

[2] Laudan and Leplin (1991) argue that theories can be confirmed by direct evidence for a seemingly unrelated theory, if both are themselves consequences of a more general theory.

provenance.[3] Suppose that a particular research strategy has led to successful theories when it has been employed in the past. To what extent should evidence for the reliability of the strategy, in the form of its past successes, boost confidence in the strategy's latest "output"? There is no reason to expect a strategy that has been reliable in the past to falter just as it reaches past accessible domains. As a descriptive claim it is clear that scientists often take factors like these into account in theory assessment; the more delicate question is whether they have abandoned empiricism in doing so.

A second motivation for a broader account of confirmation starts from the observation that mere compatibility with the data is too weak to justify the confidence we have in successful theories. A narrow construal of evidential support leaves us without sufficient reason to eliminate rival theories. Underdetermination is the claim that for a theory T_0 supported by a given body of evidence E, there are rival theories $\{T_1, T_2, ...\}$, incompatible with T_0, that are also compatible with E.[4] Without an effective way to limit underdetermination, we would have no grounds to trust T_0's claims about a given domain, or its extension to as yet unobserved domains, over the competing claims of its rivals. Scientists would only be justified in accepting low-level generalizations of the data, about which all the competitors agree. But we clearly do have sufficient evidence to justify, for example, reliance on Maxwell's equations within the domain of weak, slowly-varying electromagnetic fields. At a minimum, we need an account of theory assessment that clarifies the contrast between the appropriate level of credibility in mature, reliable theories, and in theories that, however promising, remain speculative.

Ideally these two motivations would dovetail, leading to an account of theory assessment that both describes how scientists have in fact evaluated theories (at least when they are at their best), and makes a case that these considerations are a reliable guide to true theories, compatible with the empiricist requirement that substantive knowledge of the world is grounded in experience.

Dawid (2013) proposes such an account, based on the idea that three arguments working in conjunction can be used to effectively limit the scope of underdetermination. The first of these, the "meta-inductive argument," concerns the provenance of a theory: our credence in a theory should be enhanced if it is the outcome of a successful research strategy. The second regards the failure to find alternative theories for a given domain ("no alternatives"), and the third, a theory's ability to explain things it was not designed to account for ("unexpected explanatory coherence"). Dawid gives two different defenses of this account: a formal analysis within the framework of Bayesian confirmation theory, and a historical analysis tracing the role of these arguments through several case studies. More strikingly, he applies this account to the supra-empirical phase of physics, and makes the case that we *could* still have strong support for theories, focusing in particular on string theory.[5]

[3] I am using "supra-empirical" as an alternative to "post-empirical," because the latter implies that empirical evaluation is no longer relevant. This does not need to be the case for an account of indirect confirmation, and it is not for Dawid's view. This debate regards physical theories that cantilever out into new domains from some empirical foundation. "Supra-" is intended to emphasize that further direct tests of these aspects of the theory are impossible. Yet evidence is still relevant; for example, further evidence that shores up the foundation can increase confidence in speculative extensions. My thanks to Richard Dawid for comments on this point.

[4] There are a number of variations on this theme. For example, we have a temporary or transient form of underdetermination if E is taken to be evidence available *at a given time* (or within a restricted domain); and permanent underdetermination if E includes *all available evidence*. Our focus will be on transient underdetermination. I am assuming that theories make claims about the world that extend beyond what is reflected in the evidence, and that the account given by rival theories differs. An instrumentalist would deny that this contrast makes sense, instead regarding some parts of the theories as "mere instruments," not to be interpreted literally.

[5] He is more cautious regarding whether we *do* in fact have strong support; he elucidates a pattern of argumentation that (arguably) could support string theory, while acknowledging that some of the assumptions such an argument must make may turn out to be false.

Clearly Dawid is correct that a satisfactory account of theory assessment should explain how phyisicists have effectively limited underdetermination. But when it comes to extending historically successful strategies to our unfortunate current state, the devil is in the details of how these arguments are characterized. Below I propose an alternative characterization of the ways in which physicists have limited underdetermination, with an emphasis on the context of application. There is a family resemblance between the arguments that I see as playing a role in historical cases, and Dawid's analysis. However, there is a striking contrast between the historical cases and extensions of them to contemporary fundamental physics. In the historical cases, several issues regarding the content of these arguments (such as: what are the viable alternatives? what counts as explanatory coherence?) can be settled relatively straightforwardly; whereas in supra-empirical physics, the lack of concrete applications leaves the same issues open to vigorous debate.

The paper begins by clarifying what it means to trust theory. We typically take mature theories to deserve a strong sense of trust, as we rely on them to guide practical actions in a variety of ways. Justifying this level of trust requires limiting the scope of underdetermination. In §3, I argue in favor of taking the content of physical theories to be reflected in the possibilities they introduce for measurements of theoretical quantities. Given this account of content, the threat of underdetermination is limited to two different types of cases. First, a given theory may be compatible with different extensions to new domains, and second, a rival research tradition may reject the theory's claim to have achieved stable, convergent measurements of theoretical quantities outright. I consider strategies that have been used historically in response to both types of underdetermination in §4, and argue that these do not extend to supra-empirical physics. Finally, §5 shows how these problems play out in eternal inflation.

2 Why Trust Theory?

There is an ambiguity in asking whether scientists should "trust" theories. Trust comes in various forms, and one contrast is particularly relevant here. When a theory is introduced, the evidence that can be marshalled in its support is usually provisional. Scientists often "trust theory," nonetheless, in the sense that they develop an understanding of some domain of phenomena based on the theory. The theory may allow scientists to gain access to new phenomena, as well as providing guidance in the search for particularly telling types of evidence. This sense of "trust" or acceptance is heuristic and pragmatic: how much can be gained by assuming that the theory holds? This question is distinct from an assessment of a theory's credibility: how likely is it that the theory is true? This is not to say that credibility is irrelevant to the pragmatic choice; it would be unusual for a scientist to accept a theory to which they assign very low credibility. Scientists who assign the same credibility may nonetheless make different choices about whether to accept a theory as a basis for further work – reflecting their risk tolerance, training, potential for fruitful work in the area, and so on.

Trust as provisional acceptance differs from the trust granted to theories that we take as reliably representing nature. Our trust in Newton's gravitational theory is reflected in the confident assertion that, despite its flaws, errors that arise from using it to describe the motion of bodies can be made arbitrarily small for sufficiently weak fields and low relative velocities.[6] The evidence available to Newton, and even to Laplace, was not sufficient to justify this level of trust. But Newtonian ideas became the warp and woof of celestial mechanics. The law of gravity has undergone

[6]Even though Newton's theory only provides an approximate description of the gravitational interaction, in light of general relativity, the errors introduced by using it are negligible in this domain (at least, with this crude characterization filled out appropriately); see also the end of §3 below.

ongoing tests through the development of a dynamical account of the solar system, incorporating a steadily increasing body of physical details and matching ever more precise and comprehensive observations. Newton's characterization of gravity is but one example of a scientific claim that we plausibly regard as permanent, in that it will be included, at least as an approximation valid in some domain, in any future scientific description of the world. It is hard, albeit not impossible, to imagine a way in which all of the evidence acquired in its favor could simply unravel. To reject it completely, we would need to explain how a line of research based on a false assumption had nonetheless appeared to make steady progress. Our trust in using Newtonian gravity for a variety of practical purposes rests on our confidence in its permanence.

Just as in personal relationships, developing such deep and abiding trust is a long-term achievement. In the case of scientific theories, such trust should be based on a detailed quantitative comparsion between the actual world, and the possible worlds described by a theory. Typically the comparison proceeds based on accepting a theory provisionally.[7] The theory guides the choice of experiments or observations that will be particularly revealing, and also plays the role of a tool used to extend scientists' reach, making more detailed assessment possible. As Laplace remarked, for example, Newton's law of gravity provided a tool for the study of celestial dynamics that was as essential as the telescope.

One concrete way theories play this role is in underwriting theory-mediated measurements.[8] Physical theories typically achieve clarity and conceptual economy by introducing various quantities that are not directly observable. They then owe us an account of how these theoretical quantities are revealed in properties that are accessible to us – that is, an account of how these quantities can be measured. Schematically, what is required is an explanation of what kind of system can be used as a measurement apparatus, with functional dependencies linking some readily accessible property of the apparatus (the "pointer-variable") to the theoretical quantity of interest. The functional dependencies linking the pointer-variable and target quantity follow from the dynamical description of the combined system (measuring device plus target). Further questions about the utility and reliability of the measuring apparatus can be answered based on this dynamical description. (In what domains is a particular type of device a reliable way of measuring the target quantity? Etc.) As a simple illustration, according to Newtonian gravity, local surface gravity (the target quantity) can be reliably measured by a pendulum (whose length is the pointer-variable). The domain in which pendulums of different types (small-arc circular, cycloidal, etc.) provide reliable measurements is determined by the theory.

There is clearly a great deal to be gained by accepting a theory. Theory-mediated measurements make it possible for a given body of data to constrain and inform a richer description of the phenomena. Yet there is also a risk to accepting a theory. The new quantities introduced may merely fit a given body of data, without accurately representing the phenomena. How should we avoid the fruitless line of research that would result from accepting such a theory? One response would be to demand a higher evidential threshhold for acceptance. This strikes me as misguided as a general policy, given the points above, in addition to being descriptively inaccurate (see also Kuhn, 1961). Scientists do often accept theories on the basis of remarkably weak evidence, with the expectation that subsequent work will reveal any mistakes. In most cases the only way to reveal mistakes or limitations is through a long-term assessment of the theory's potential, which is guided by the theory itself. The crucial question is whether the theory is in fact subject to further critical evaluation, or if it instead shapes further inquiry in a way that shields it from ongoing scrutiny. This risk is particularly salient when accepting a theory that grants exclusive access to

[7] This is one of the main themes of Kuhn's work; see, in particular, Kuhn (1961).

[8] Here I am following the account of measurements in Newton's physics developed in Harper and Smith (1995); Smith (2002); see also Chang (2004).

some set of novel quantities, in the sense that all measurements of a target quantity (or set of such quantities) rely on a single theory. If the interpretation of available data depends on the very theory that the data is being used to evaluate, there is an obvious risk of circularity.

One further clarification regarding what it is that scientists accept or trust will be essential in what follows. On my view, the object of trust or acceptance is not restricted to theories, because what is used as a tool to guide further work may be more specific or more general than what philosophers usually call "theories." The functional dependencies supporting inferences from pointer-variable readings to target quantities may follow from a theory, but this need not be the case. In many fields, empirically discovered regularities, not yet integrated into a general theory, are used to gain access to new quantities.[9] In other cases, one example of which is discussed below, quite general background assumptions (shared by many theories) may suffice for some inferences from the data.

Furthermore, scientists routinely make judgments about how evidence bears on claims that are more specific, or more general, than theories. For example, astronomers in the 18^{th} century carefully analyzed what parts of Newton's characterization of gravity were actually supported by successful descriptions of orbital motion. This success bears directly on the dependence of gravitational force with distance, but the same cannot be said for the further claim that gravitational attraction is proprotional to the mass of the attracting body.[10] On the other hand, scientists often have compelling arguments for principles that stand above specific theories, such as symmetry principles in contemporary physics.

Philosophical accounts that treat questions of evidence as a relation between a collection of data and a monolithic "theory T" hide this complexity from view (see also Wilson, 2006). A more fine-grained analysis of what principles are actually relevant for specific lines of evidential reasoning makes it possible to see whether a set of results rely on independent principles. The evidence provided by convergence of independent lines of reasoning would be overlooked if the results are regarded as all simply following from "theory T". In addition, in some cases principles more general than a theory are sufficient to support inferences from the data, as we will see below in considering eliminative programs.

These considerations lead to a sharper formulation of the question of trust: how have scientists successfully made the transition from a risky, pragmatic choice to accept a theory, to taking it as a permanent contribution to our knowledge of the world and a reliable guide for action? And what role does the differential evaluation of components of a theory, or more general principles it satisfies, play in that transition? A theory is worthy of trust in the stronger sense to the extent that it is not merely compatible with the data, but is the only viable characterization of the law-like relationships that hold in the relevant domain. In the next section, I will turn to the challenge that apparently arises to establishing trust due to the underdetermination of theory by evidence.

---

[9]Theories may support the idea that such a relationship should exist, while leaving the details to be fixed by observation; or, in other cases, empirically stable regularities are discovered that cannot be accounted for theoretically. Understanding how these regularities fit into a larger theory makes it possible to answer a number of further questions, such as those regarding reliability and domain of applicability. But a regularity may support useful inferences from the data without being successfully integrated. Reliability may be established empirically through repeated successful applications, even if there are open theoretical questions as to why the regularity holds. I owe recognition of this point to conversations with George Smith.

[10]As Euler and others pointed out, purely gravitational interactions among celestial bodies cannot distinguish between the contributions of the gravitational constant G and the mass of the attracting body M to the overall attractive force. The mass can be assigned freely if the value of G is allowed to vary, unless the third law is taken to apply to this interaction. (Euler, among others, regarded this extension of the third law to action between non-contiguous bodies as not sufficiently supported by direct evidence.) The strongest evidence that the force of gravity is proportional to M comes from experimental measurements establishing that G is a universal constant. See Smith (2007) for a detailed assessment of this issue.

3 Underdetermination

One of Dawid (2013)'s main themes is that trust, in the stronger sense, can only be earned through an effective response to underdetermination. While I agree on this point, the details of how scientists have effectively limited underdetermination in the past matter a great deal in trying to evaluate whether they can succeed, using similar approaches, in a context where theories cannot be developed and assessed through concrete application to accessible phenomena.

The first task is to clarify the nature of underdetermination and the threat it poses. There is a striking contrast between the treatments of underdetermination by philosophers and physicists: philosophers, following the pronouncements of Quine and others, hold that there are *many* rival theories that fare equally well with respect to some body of evidence. By contrast, physicists tend to emphasize how challenging it is to find even *one* theory compatible with the evidence (as emphasized by Norton, 1993). The contrast reflects differing views about the empirical content of theories, and what success with respect to some body of data implies.

Philosophers who regard rival theories as plentiful usually adopt a narrow conception of empirical content: it is limited to a set of "observational claims," or to an "empirical substructure" (part of a theoretical model). By contrast, I endorse Wilson (1980)'s proposal that a theory's content should be characterized in terms of the possibilities it introduces for theory-mediated measurements. Rather than thinking of the evidence as exhausted by a set of observational claims, on this view the evidence consists of a characterization of systems that qualify as measuring devices for specific quantities, along with the results of measurements performed using these devices (schematically, pointer-variable readings).

There are several reasons to prefer an understanding of empirical content along these lines. Perhaps the most compelling is that the dynamical description of measurements makes it possible to assess their stability and reliability, which are clearly both essential to scientific practice.[11] Measuring devices can often be thought of as amplifying some feature of a target system, so that it is registered in an accessible pointer-variable. When does the device reliably amplify the target quantity, and to what level of precision can it be trusted? Answers to these and related questions depend on the theoretical description of the functional dependence holding between the target quantity and the pointer-variable. Suppose that a community of scientists has established a reliable means of measuring a target quantity, at some level of precision, that is *stable* in the sense that repeated measurements in varying circumstances yield consistent results. This success lends support to the theoretical account of measurement, including its counterfactual implications (e.g., to what extent we could still use the measuring device confidently in a new setting). This extends beyond what would be included on a narrow construal of content.

Consider, for example, the use of the length of a pendulum (ℓ) of a given period to measure the variation of surface gravity with latitude ($g(\theta)$). On a narrow construal of content, a rival theory would need to recover a set of data points (roughly, ordered pairs $\{\ell, \theta\}$); on the proposed alternative, by contrast, a rival theory would need to recover not just these results, but also account for the reliability and limits of using pendulums for these measurements. Any rival theory recovering content in this stronger sense would agree on the fragment of Newtonian dynamics needed to establish the functional dependencies holding between ℓ and g.

Theories of broad scope typically provide multiple ways to measure theoretical quantities, making it possible to triangulate on them using independent methods. Continuing with the example, Newton argued that the same force causes the motion of pendulums and the moon, because they both give consistent measurements, in this general sense, of the Earth's gravitational field. A

[11] This is not the only reason to adopt an alternative conception of content; for further related discussions, see Bogen and Woodward (1988); Roberts (2008).

further expectation of successful theories is that the multiple measurements they make possible will converge on a consistent assignment of theoretical quantities.[12] Insofar as each distinctive measurement method employs a different fragment of the theory, they provide independent constraints. The convergence of a set of independent measurements on a common value for the target quantity is a particularly powerful reply to a skeptic who holds that the theory succeeds by simply "fitting" parameters to match the data.

A crucial further aspect of assessment regards how the theory fares in response to the pressure of systematically improving standards of measurement precision. As noted above, a theory is often accepted in order to gain access to a new domain of phenomena. The initial applications of a theory typically start with simple models of the measurement interaction as well as the system being studied. As standards of precision improve, discrepancies between observations and these simple models are used to guide the development of more detailed models. A theory is successful to the extent that it can be consistently applied to develop successively more detailed and precise models of the relevant phenomena, without abandoning core principles. The development of models incorporating further physical details to account for discrepancies also provides more opportunities for corroboration – for example, astronomical observations of Nepture, which was introduced in order to resolve discrepancies in Uranus's orbit.

Returning to the question of underdetermination, a theory that is successful in the sense of supporting stable, convergent measurements of the theoretical quantities it introduces, and is consistently applied as measurement precision improves, leaves little room for rival theories covering the same domain. On the proposal I have been considering, the content of the theory includes all the elements needed to account for theory-mediated measurements of the relevant set of target quantities. If we construct a rival theory that captures the same content, as Wilson (1980) argues, we will recover very nearly the same theory — a reformulation rather than a true rival.[13]

There are two important qualifications to this uniqueness claim.[14] First, there may be "locally indinstinguishable" rival theories, that share the same content in this sense, to some specified level of precision within a specified domain, yet differ in other domains. For example, by construction general relativity reduces to Newtonian gravity in an appropriate limit. Within the domains to which Newtonian theory applies, general relativity is a locally indistinguishable rival.[15] Even though Newton's theory only provides an approximate description of the gravitational interaction, in light of general relativity, the errors introduced by using it are negligible in this domain. The existence of rivals in this sense does not undercut trust in the theory. If general relativity is true then the Newtonian results hold as good approximations within the relevant domain. To undermine the reliability of Newtonian gravity in the relevant sense, we would need to discover a rival theory that substantively differs from it *even within this domain.*

The limits of the domain within which we can trust a given theory are, however, often only clear retrospectively. The existence of locally indistinguishable rivals reflects the ineliminable risk associated with inductive generalizations. Newtonian gravity adequately describes the law-like

[12]Harper (2012) argues that Newton introduced an "ideal of empirical success" along these lines. See Chapters 4 and 6 for a particularly thorough discussion of the comparison between pendulum results and the lunar orbit (called the "moon test"), which defends attributing this account to Newton.

[13]Wilson (1980) makes the case that classical mechanics is essentially the only theory compatible with the phenomena, in this rich sense, within its domain of applicability.

[14]There is a third qualification regarding the scope of this approach: my primary focus is on physical theories, which are distinctive in usually requiring an explicit account of measurement of novel quantities on their own terms. It is not clear how much of this approach carries over to areas of science that treat measurement in different ways, e.g. if access to quantities is mediated by theories from another field.

[15]It is also an "unconceived alternative" in the sense discussed by Stanford (2006), because it was not explicitly formulated as an alternative when Newton's theory was originally accepted.

relations revealed in solar system motions, but it would be a mistake to take these laws describing weak-field effects as characterizing the gravitational interaction more generally.[16] Of course, it is extremely difficult to assess to what extent the evidence available at a given stage of inquiry is parochial or limited in this sense. These limits are often uncovered by extrapolating the theory boldly and determining where it begins to lose fidelity.

Second, there may be what I will call "rival research traditions" that differ sharply from an existing theory. Although this idea is admittedly somewhat vague, for my purposes the defining characteristic of such a rival is that it rejects a theory's claims to have achieved stable, convergent measurements of theoretical quantities. In other words, the initial theory failed to establish that there are real phenomena that must be preserved. From the standpoint of a rival research tradition, the apparent successes have to be explained away as misleading — the result of systematic errors, coincidences, or something else along these lines. Such explanations undercut the rationale for preserving the content of the theory in the sense defended above. All that needs to be preserved is the raw data, although it may not have much intrinsic interest. The rival research tradition has to further provide some account of how to understand the earlier data, without recovering the theory-mediated measurements and associated commitments. Such rival research traditions typically also pursue different aims for inquiry in a given area, prioritizing an altogether different set of problems and suggesting alternative approaches to resolving them.

4 Eliminating Rivals

To what extent have physicists been able to eliminate rival theories, in these two different senses? An answer to this question effectively rebuts a skeptic, who objects that the apparent success of a theory may merely reflect its flexibility, and is an essential part of establishing trust in the stronger sense. I will briefly discuss two concrete cases from the history of physics exemplifying successful strategies.[17] This is not intended to exhaust the approaches physicists have used in eliminating rival theories, but I expect that two cases will be sufficient to raise questions regarding whether similar approaches can be extended to supra-empirical physics.

First, physicists in many areas have pursued eliminative programs that proceed by constructing a space of possible theories, and then winnowing this down to a small subset – or even a single theory – compatible with the evidence.[18] In the first step, the parametrized possibility space includes known alternative theories as well as merely possible competitors that have not yet been explicitly formulated. This allows comparison between existing theories and an entire class of rival theories. Furthermore, in the second step, the implications of different types of data follow from assumptions held in common among all these competing theories. It is then possible to use data to constrain the space of allowed theories, ideally eliminating almost all the possibilities, without needing to perform calculations for each theory. Success in this program supports a local version of what Dawid calls the "no alternatives argument," by unambiguously identifying the "best" theory among a clearly articulated space of competitors.

To take one prominent example, the "Parametrized Post-Newtonian" (PPN) formalism is a systematic framework designed to allow observations to choose among general relativity and various

[16] See Smith (2014), especially §3, for a discussion of the relationship between Newtonian theory and GR, which I draw on here.

[17] In both cases, I am isolating one line of argument that has been used to make the case for the theory discussed; obviously, there is much more to be said about evidence in favor of GR or QED than I have the space to discuss here.

[18] Dorling (1973), Earman (1992, Chapter 7) and Norton (1993) all emphasize the importance of eliminative inferences in physics; it also plays a central role in Kitcher (1993)'s account of scientific progress.

competing accounts of gravity. It is relatively straightforward to compare competing theories in the post-Newtonian limit, in which the differences among spacetime metrics for a broad class of competing theories can be characterized in terms of ten coefficients. The PPN formalism represents a "possible gravitational theory" with a point in a 10-dimensional parameter space, although distinct theories may correspond to the same point.

Two main claims justify taking this parameter space as delimiting the relevant alternatives. First, there is broad agreement regarding the domain of "gravitational" phenomena and how facts about this domain should be handled. Candidate theories are all expected to account for facts about this domain based primarily on the dynamics of the gravitational interaction. The second more substantive argument limits consideration to so-called "metric theories" of gravity, in which gravity is treated as an effect of spacetime curvature (Will, 2014). The eliminative program thus begins with a carefully circumscribed set of possible theories. Over the last half century, experimental tests have constrained all 10 PPN parameters to be very close to the values for GR (summarized in Will, 2014). At an abstract level, at least, these tests are relatively straightforward: based on an account of some phenomena valid for all metric theories, including explicit dependence on the PPN parameters, observations are used to determine the best values of the parameters using standard statistical techniques. This is possible because the structure shared by metric theories is sufficient to link observations to the PPN parameters.

This argument for GR is local in the sense that it applies to this specific regime. This focus enables the construction of parametrized possibility space, and identification of common principles needed to link observations to parameters. Yet due to its local character, this argument leads to a limited conclusion: GR, or any locally indistinguishable rival, outperforms alternative metric theories of gravity in accounting for solar system dynamics. The map from specific PPN parameters to dynamical theories of gravity is one to many, however, so these results will not distinguish among metric theories that share the same limiting behavior in this regime. Because of this degeneracy, a thorough eliminative program requires local tests in a variety of regimes. GR and competing dynamical theories of gravity specify links among these various regimes, and the tests can be complementary if they have different degeneracies. Recent gravitational wave observations have provided an opportunity to pursue tests of GR in the strong-field regime, providing a powerful complement to the solar-system tests.

This case illustrates that tests in several distinct regimes can effectively eliminate locally indistinguishable rivals. Tests of gravitational effects in the strong-field regime and in cosmology promise to differentiate among theories that match GR's success in describing the solar system. This strategy is based on explicitly defining the space of possible theories in terms of common principles similar to those that hold in GR, along with assumptions regarding what domain of phenomena a gravitational theory should explain. These assumptions define a research tradition in gravitational physics. Our assessment of GR as a permanent contribution to science, a characterization of the gravitational interaction that will at least be recovered as a limiting case of some future theory, is based on accepting this research tradition. The argument in favor of accepting this research tradition itself has a different character, based on assessment of how much progress has been made in developing increasingly detailed models of phenomena in light of a growing body of observational and experimental results.

Second, the convergence of theory-mediated measurements sometimes takes a particularly sharp form: multiple measurements of a single fundamental constant. Perrin (1923)'s famous case in favor of the existence of atoms was based on 13 independent measurements of Avogadro's number. Physicists have often emphasized the "overdetermination of constants" (borrowing Norton (2000)'s term) as a reply to skeptical worries that a theory fits the data merely because of its flexibility.

The strength of this argument depends on two claims. As the number and diversity of ways of determining the value of a fundamental constant increase, the odds of attributing the agreement to systematic error associated with each measurement should decrease. The chance that the various experimental measurements would agree, even if the theory were fundamentally false, is also expected to decrease.

Consider, for example, the evidence in favor of quantum electrodynamics (QED) based on agreeing measurements of the fine-structure constant α. Remarkable levels of precision have been reached in low-energy tests of QED: state of the art measurements of the anomalous magnetic moment of the electron (a_{e^-}) achieve a precision to better than one part in 10^{12} (Gabrielse et al., 2006). For these measurements, Gabrielse and his collaborators have effectively created something that is as close to a "pure QED system" as possible: in effect, a single electron cyclotron. The system can be described without needing to worry about the complicated structure of protons, as would be required to perform spectroscopic measurements of comparable precision. QED determines a theoretical value of a_{e^-}, through a perturbative expansion in terms of the fine structure constant α. Schwinger (1948)'s success in calculating corrections to Dirac's value for the magnetic moment of the electron initially inspired confidence in QED. Seventy years later the quantitative comparison of QED and the world that has been achieved is simply astonishing: theoretical calculations have now been carried out to 10^{th} order in α, including a total of 12,672 Feynman diagrams (Aoyama et al., 2015). These results lead to a consistency check (rather than a direct prediction of a_{e^-}): given an independent determination of α, the computed value of a_{e^-} can be compared with the precision measurements.[19]

QED underwrites numerous experimental measurements of α. Alongside high-precision measurements of a single electron in a Penning trap, the value of the fine-structure constant α can be determined based on atomic recoil experiments, spectroscopic measurements, the quantum Hall effect, the AC Josephson effect, and various scattering amplitudes (see, e.g., Kinoshita, 1996). The agreement among these determinations is required to hold, insofar as QED correctly describes electromagnetism in these domains. The fact that they coincide within the bounds of experimental error provides powerful evidence in favor of QED: it provides a coherent account of a wide range of phenomena, at an astonishing level of precision. As the number of theory-mediated measurements increases, along with their precision, it is harder to preserve the connections between these diverse experimental situations in a theory that truly differs from QED in the relevant domains.

The strength of the conclusion of overdetermination arguments depends on whether this agreement would be expected on other grounds. What is the probability that these various determinations of α agree within experimental error, if QED were false? The argument thus relies implicitly on an assessment of the space of competing theories.

Dawid (2013) advocates a "new paradigm of theory assessment" in which arguments to limit underdetermination play a central role. There is a family resemblance between the arguments Dawid identifies and the strategies just described. The "no alternatives argument" (NAA) holds that a search for an alternatives to a given theory T can provide evidence for T if it fails to turn up a better alternative. Successfully implementing an eliminative program provides a precise version of this idea, by determining what theory among a space of competitors fares best with respect to the evidence. A theory's unified description of diverse phenomena, which were not taken into account when the theory was initially formulated, is an example of "unexpected explanatory co-

[19]The resulting comparison does not straightforwardly vindicate QED: there is a discrepancy between the QED calculation and observations. Agreement can be recovered by including hadronic loop contributions and contributions from the weak interaction. In other words, the low-energy precision tests have reached such remarkable precision that it is in fact the Standard Model being tested, rather than QED alone.

herence" (UEC). Several of the phenomena that now provide precision constraints on α, based on QED, played no role in the historical development of the theory, and hence provide examples of unexpected coherence. Finally, I noted above that theories are successful insofar as they are applied consistently even as the scope, detail, and precision of available observations increases. This idea presupposes that there is a distinction between maintaining the same strategy – maintaining core principles, while changing or adding details – through applciations. Dawid's third, "meta-inductive" argument (MIA) assumes a similar notion of constancy of research strategy: the provenance of a theory, as the output of a research strategy with a good track record, also provides evidence in its favor. Dawid takes a "good track record" to include, in particular, successful novel predictions.[20] The appealing idea is that these arguments have played an essential role in inquiry in the past, but – unlike more direct forms of empirical confirmation – they can continue to guide inquiry even as we depart the realm of the accessible.

This position, appealing though it may be for other reasons, is not compatible with the characterization of responses to underdetermination given above. I have emphasized the local character of two arguments that have been important in eliminating rival theories. Any eliminative program has to start somewhere, with an explicit choice regarding what type of theories qualify as legitimate competitors. Similarly, the overdetermination of constants argument exemplified by QED requires some assessment of the space of rival theories. The assessment of relevant competing theories is based on what I have called, loosely, a research tradition. In the case of the PPN framework, the competing theories are all similar to GR: metric theories satisfying the Einstein equivalence principle, taking similar phenomena as their explanatory target. Following through on the eliminative program makes it possible to eliminate nearby rival theories in favor of GR. The assumptions defining this framework are motivated by the success of a long research tradition in gravitational physics.

This acknowledgment of the limitations of these strategies does not undercut trust in theories, however, because there is more to the story. The assessment of rival research traditions has a different character than these local responses to underdetermination. This kind of argument addresses a different concern. Once a theory is accepted as the basis for inquiry, much of the subsequent reasoning is heavily theory-dependent. How can the resulting line of inquiry be compared with a rival approach that shares little common ground?

Without minimizing the difficulties in doing so, on my view competing research traditions can usually be evaluated effectively based on concrete applications. In many cases, a theory-dependent line of reasoning leads to striking claims that can be evaluated by entirely independent methods. The validity and value of these claims is obvious, even without contentious decisions regarding how to characterize the phenomena, or how to explain them. An example discussed by Smith (2014) illustrates the general point beautifully. The Hill-Brown lunar theory, published in 1919, is an enormously complicated Newtonian description incorporating roughly 1,400 different physical sources of perturbations. Brown discovered a discrepancy between this theoretical description and observed motion that he called the "Great Empirical Term." Jones (1939) argued that this discrepancy should be attributed to a fluctuation in the Earth's rotation with a 225-year period, which has subsequently been confirmed through a variety of independent methods. The agreement of these different precision measurements of the Earth's rotation is a striking result regardless of how the domain of gravitational phenomena is delineated.

Results of this sort can be decisive between competing research traditions, and provide compelling evidence that a research tradition has accurately identified law-like dependencies and fun-

[20]The language of "research strategies" is my own, whereas Dawid formulates his position in terms of a broader category, "research fields." There is an interesting question here regarding the appropriate unit of analysis, but I do not have the space to pursue it here – and it is not important for what follows.

damental quantities within a given domain. The application of a theory to an accessible domain of phenomena is essential to making this case. The challenge for supra-empirical physics is to find similarly compelling arguments when the fruits of successful applications are not available.

5 Eternal Inflation

Inflationary cosmology is often discussed, alongside string theory, as an example of "post-empirical" physics. The simplest models of inflation account for a variety of observed features of the early universe as the consequence of the dynamical evolution of a scalar field, which drove a transient phase of exponential expansion. Here I will briefly consider how the line of thought developed above applies in this case: to what extent can inflation be subjected to ongoing scrutiny if it is accepted, and is there a way to effectively eliminate rival theories?[21]

Observations have led to a remarkably simple picture of the early universe, which is well-described by a flat FLRW model, with Gaussian, adiabatic, linear, nearly scale invariant density pertubations. Inflation has remained the most widely accepted explanation of why the early universe has these features, even as three decades of increasingly precise observations of the cosmic microwave background radiation (CMBR) have successfully ruled out competing proposals.[22] The details of inflation are not directly accessible to experiments, given the features of the inflaton field in the simplest models. In principle an "inflaton" field has implications for accelerator physics, but in practice we will clearly never build an accelerator to probe its properties. Cosmological observations are thus the primary source of evidence for the inflaton.

Inflation might qualify as a "post-experimental," yet still observational science, if not for a further step. Many cosmologists hold that inflation is "generically eternal," in the sense that inflation produces a an ensemble of "pocket universes," quasi-isolated regions in which inflation has come to an end, within a still-inflating background (see, e.g., Aguirre, 2007). The mechanism leading to this multiverse structure is typically assumed to produce variation in low-energy physics among the different pocket universes. This version of inflation, in particular, exemplifies the risk identified earlier: it provides a framework for further inquiry that shields basic theoretical assumptions from substantive tests.

For the sake of contrast, consider first a view of inflation that takes observations to constrain the properties of the inflaton field – in effect, determining the form of the Lagrangian for the inflaton field. Observations of the CMB and large scale structure would give theory-mediated measurements of the Lagrangian in two main ways. Primordial fluctuations have implications for features of the inflaton field at a time well before the end of inflation.[23] The energy density of other matter fields is rapidly diluted during inflation, so most inflationary models require a

[21] This line of argument is similar in some ways to recent prominent criticisms of eternal inflation due to Ijjas, Steinhardt and Loeb (see Ijjas et al., 2014). Many of these points have been raised earlier, by critics including George Ellis, but these publications have brought renewed attention. For example, Ijjas et al. (2017), a *Scientific American* article presenting their arguments at a more accessible level, led proponents of inflation to defend their theory in a letter to the editor signed by 33 prominent physicists and cosmologists.

[22] Observations of acoustic peaks lend support to the idea that the fluctuations are primordial, as they are in inflation, contrasting with predictions from competing models of structure formation based on active sources for fluctuations (such as topological defects). (See, e.g., Durrer et al. (2002) for a review of structure formation via topological defects, and its contrasting predictions for CMB anisotropies.)

[23] Inflation generates scalar and tensor perturbations whose physical properties depend on the features of the effective potential $V(\phi)$ at horizon exit, with $\frac{k}{R} \approx H$. Perturbations relevant to CMB observations typically crossed the horizon at ≈ 60 e-foldings before the end of inflation, whereas those that are re-entering the horizon now were produced a few e-foldings later.

phase of "re-heating," during which the inflaton field decays into other particle species such that the post-inflationary state is compatible with a "hot" big bang. Observations can also be used to determine features of the inflaton Lagrangian, such as interaction terms with other fields, based on an account of reheating.

The best case scenario for this kind of approach obtains if the inflaton is identified within a specific particle physics model. The parameters appearing in the Lagrangian would then be constrained by cosmological data related to the details of inflation, along with experimental data relevant to the particle physics model. The promise of obtaining constraints from cosmology alongside particle physics was surely one of inflation's most appealing features. Agreement of theory-mediated measurements of the properties of the inflaton field from such strikingly different domains would be an astonishing example of unification.

Fulfilling this promise has proven to be elusive. No canonical candidate for the inflaton field has been widely accepted. There are now a wide variety of different inflationary models, and the prospects for establishing a tighter link to particle physics are bleak. This certainly makes the prospects for observationally constraining the properties of the inflaton field more daunting. Even though cosmological observations can be used to constrain properties of the inflaton field, it is much harder to make the case that they are independent, and that they provide consistent constraints on the same underlying feature.[24] Furthermore, in part because of the lack of a canonical model, inflation has not led to the identification of robust physical features of the early universe that can be tested independently. There is no analog of the discovery of Neptune based on the comparison of inflationary models with observations – that is, a result that is compelling independent of the research tradition that led to it.

In sum, on this first approach it is challenging to see how to make a more detailed empirical case for inflation, primarily due to the inaccessibility of the relevant energy scales. The compatibility of the observed state of the universe with inflation is a significant success. But this success has not yet been followed up with the development of a more detailed account of how inflation transpired. The worry is that the progress inflation has made is analogous to that in Ptolemaic astronomy: progress in developing a more precise model with better parameter fits, without leading to a more accurate model.

The situation is much worse if inflation leads to EI. Accepting eternal inflation undermines the observational program of attempting to constrain and fix the features of the inflaton field.[25] Developing strong evidence, or eliminating rival theories, relies on the exactness of a theory, and EI is anything but exact.

The central question can be put quite simply: what do we gain by accepting that EI is true? Advocates of EI often claim that we can make statistical "predictions" about, for example, the value of fundamental parameters such as Λ. Several further questions need to be answered in the course of making these predictions. The first regard a physical characterization of the ensemble generated by inflation: How is the ensemble of pocket universes characterized? (For example, does a given fundamental constant or other aspect of low energy physics vary among the members of the ensemble?) How do we count different types of pocket universes in the ensemble – or, in other words, what is the measure over the ensemble? A distinct type of question regards the selection effect associated with the presence of observers like us. What part of the ensemble includes observers? What are the necessary conditions for observers like us? None of these questions have

[24] The second difficulty reflects the fact that the observations constrain different aspects of the inflaton Lagrangian: different parts of the effective potential $V(\phi)$, or interaction terms.

[25] Ijjas et al. (2014) argue that many of the defenders of inflation fail to acknowledge the full impact of shifting to this "post-modern" version of the theory (see also Smeenk, 2014, 2017).

clear answers. But even if these are all satisfactorily resolved, the further assumptions needed to derive "predictions" put substantial distance between the basic physical description of inflation and our assessment of it.

The appeal to anthropic selection effects is particularly problematic. Consider the "prediction" we obtain from EI for the value of some parameter α_i. One part of this calculation is based on evaluating the range of values of this parameter that are compatible with our existence. (Imagine varying α_i to see what range of values is compatible with the existence of certain physical systems that are necessary conditions for our presence, such as gravitationally bound systems at various scales.) The EI proponents see this calculation as clarifying what part of the ensemble we might inhabit, and modulate their probabilistic predictions in light of the selection effect. Why not take the calculation of the range of values α_i as in itself explaining why we observe the value that we do? This calculation would then "screen off" the alleged explanatory value of the multiverse. Perhaps the multiverse can be said to explain how it is possible that α_i has the value that it does. There is a much cheaper explanation available, which simply takes the parameter value as a contingent feature of the universe. But more importantly, this line of argument does not provide information about the kinds of law-like relations among parameters or features of the universe that would support ongoing scrutiny of the framework.

Acknowledgements: I have benefitted from questions and discussions with several people following talks based on this material, including, Peter Achinstein, Andy Albrecht, Radin Darshati, George Ellis, Helge Kragh, and Kyle Stanford. It is a pleasure to thank Richard Dawid for discussions as well as comments on an earlier draft of the paper. The title alludes to a paper by George Smith, whose influence will be apparent to those familiar with his work. Finally, I gratefully acknowledge that work on this paper was supported by a grant from the John Templeton Foundation. The views expressed here are those of the author and are not necessarily endorsed by the John Templeton Foundation.

References

Aguirre, Anthony (2007). "Eternal inflation, past and future." *arXiv preprint arXiv:0712.0571* .

Aoyama, Tatsumi, Hayakawa, Masashi, Kinoshita, Toichiro and Nio, Makiko (2015). "Tenth-order electron anomalous magnetic moment: contribution of diagrams without closed lepton loops." *Physical Review D* 91, 033006.

Bogen, James and Woodward, James (1988). "Saving the phenomena." *The Philosophical Review* 97, 303–352.

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

Dawid, Richard (2013). *String Theory and the Scientific Method*. 1st ed. Cambridge: Cambridge University Press. ISBN 0521883490.

Dorling, Jon (1973). "Demonstrative induction: Its significant role in the history of physics." *Philosophy of Science* 40, 360–372.

Durrer, Ruth, Kunz, Martin and Melchiorri, Alessandro (2002). "Cosmic structure formation with topological defects." *Physics Reports* 364, 1–81.

Earman, John (1992). *Bayes or Bust?*. Cambridge: MIT Press.

Gabrielse, G, Hanneke, D, Kinoshita, T, Nio, M and Odom, B (2006). "New determination of the fine structure constant from the electron g value and qed." *Physical Review Letters* 97, 030802.

Harper, W. L. and Smith, George E. (1995). "Newton's new way of inquiry." In *The Creation of Ideas in Physics*, Jarrett Leplin, ed., vol. 55 of *University of Western Ontario Series in Philosophy of Science*. Dordrecht: Springer Verlag, 113–166.

Harper, W.L. (2012). *Isaac Newton's Scientific Method: Turning Data Into Evidence about Gravity and Cosmology*. Oxford: Oxford University Press.

Ijjas, Anna, Steinhardt, Paul J and Loeb, Abraham (2014). "Inflationary schism." *Physics Letters B* 736, 142–146.

— (2017). "Pop goes the universe." *Scientific American* 316, 32–39.

Jones, H Spencer (1939). "The rotation of the earth, and the secular accelerations of the sun, moon and planets." *Monthly Notices of the Royal Astronomical Society* 99, 541–558.

Kinoshita, Toichiro (1996). "The fine structure constant." *Reports on Progress in Physics* 59, 1459.

Kitcher, Philip (1993). *The Advancement of Science*. Oxford: Oxford University Press.

Kuhn, Thomas S (1961). "The function of measurement in modern physical science." *Isis* 52, 161–193.

Laudan, Larry and Leplin, Jarrett (1991). "Empirical equivalence and underdetermination." *Journal of Philosophy* 88, 449–472.

Norton, John D. (1993). "The determination of theory by evidence: The case for quantum discontinuity, 1900-1915." *Synthese* 97, 1–31.

Norton, John D (2000). "How we know about electrons." In *After Popper, Kuhn and Feyerabend: Recent Issues in Theories of Scientific Method*, R. Nola and H. Sankey, eds. Dordrecht: Kluwer Academic Publishers, 67–97.

Perrin, Jean (1923). *Atoms*. New York: Van Nostrand. Translated by D L Hammick.

Roberts, John T. (2008). *The Law-Governed Universe*. Oxford: Oxford University Press.

Schwinger, Julian (1948). "On quantum-electrodynamics and the magnetic moment of the electron." *Physical Review* 73, 416.

Smeenk, Chris (2014). "Predictability crisis in early universe cosmology." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 46, 122–133.

— (2017). "Testing inflation." In *The Philosophy of Cosmology*, Khalil Chamcham, John D. Barrow, Simon Saunders and Joseph Silk, eds. Cambridge University Press, 206 - 227.

Smith, George E. (2002). "The methodology of the *Principia*." In *Cambridge Companion to Newton*, I. B. Cohen and George E. Smith, eds. Cambridge: Cambridge University Press, 138–173.

— (2007). "The Question of Mass in Newton's Law of Gravity." Manuscript of talk given in Leiden and Austin.

— (2014). "Closing the loop." In *Newton and Empiricism*, Zvi Biener and Eric Schliesser, eds. Oxford: Oxford University Press, 262–351.

Stanford, P Kyle (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives.* Oxford: Oxford University Press.

Will, Clifford M. (2014). "The confrontation between general relativity and experiment." *Living Reviews in Relativity* 17, 4.

Wilson, Mark (1980). "The observational uniqueness of some theories." *The Journal of Philosophy* 77, 208–233.

— (2006). *Wandering significance: An essay on conceptual behavior*. Oxford: Oxford University Press.