Experimental Knowledge in the Face of Theoretical Error

Kent W. Staley

Saint Louis University

Contents

Abstract

This paper attempts to make progress toward a greater understanding of the problem of theoretical error in experimental contexts. Drawing upon two eras of research in gravity physics (Newton's argument for Universal Gravitation and modern tests of gravity theories interpreted through parametric frameworks), I will consider how physicists have confronted the problem of theoretical assumptions that are potentially inadequate for the inferences they seek to draw. Using a conceptual apparatus drawing upon epistemic notions of modality, I will defend a guardedly optimistic position: Experimenters can avail themselves of methods that allow them to secure (but not guarantee) their experimental conclusions against the possibilities of erroneous theoretical assumptions.

1 Indtroduction

Much of the advancement of scientific knowledge is due to or consists of advances in experimental knowledge. Such experimental progress, moreover, owes a great deal to advances in our understanding of how to systematize, quantify, and eliminate error. It is therefore no surprise that the turn by philosophers of science toward error as a subject of interest in its own right is strongly associated with their turn toward experiment, beginning with the "New Experimentalists" in the 1980s. Much of the work that characterized the New Experimentalism in philosophy of science emphasized the idea that, in Hacking's words, "experimentation has a life of its own," a slogan that could be interpreted differently depending on whether one wished to emphasize the continuity of experimental methods across radical changes in theory (Galison 1987), the possible independence of experimental evidence from theoretical assumptions (Franklin 1986; Hacking 1983; Mayo 1996), or the growth of knowledge at the experimental rather than the theoretical level as the driver of scientific advancement (Franklin 1986; Mayo 1996; Galison 1997).

It was perhaps inevitable that there would be some backlash against this emphasis on the independence of experiment from theory. Even authors who contributed to the turn toward experiment began to point out the important role played by theoretical assumptions in arriving at experimental results (A. Chalmers 2009; Heidelberger 2003; Radder 2003; Staley 2008). Recognizing this role, however, makes it incumbent upon philosophers of science to gain clarity about the effects it might have upon the possibility of cumulative knowledge at the experimental level (i.e., both knowledge gained by inferences made from experimental data and knowledge about how to reliably produce, control, and analyze experimental data).

The present paper attempts to make progress toward a greater understanding of the problem of theoretical error in the context of drawing conclusions from experimental data. More precisely, my discussion will focus on the problem of theoretical assumptions that threaten to lead investigators to either misinterpret their data or to incorrectly evaluate the support the data provide for their conclusions. Such erroneous theoretical assumptions demand philosophical attention because unlike, say, possible error due to variability in a sample, we presently have no satisfactory quantitative statistical framework for taking such errors into account. I will consider a number of different ways in which errors in theoretical assumptions can manifest themselves at the empirical level. Theoretical assumptions pose dangers when drawing inferences from experimental data for several reasons. Here I will discuss examples highlighting problems that fall into two broad categories: (1) *problems of scope*: because of their great generality theoretical assumptions are difficult to test exhaustively and sometimes must be applied in domains far from those in which they have previously been tested; (2) *problems of stability*: the rare-but-recurring historical phenomenon of unforeseen and dramatic *change* in general explanatory theories arguably constitutes some reason to worry about similar upheavals overturning theoretical assumptions that presently seem sound; moreover, conceptual innovation sometimes enables erroneous *implicit* assumptions to be brought to light.

In the face of these dangers, I will defend a guardedly optimistic position: Experimenters can avail themselves of methods that allow them to secure their experimental conclusions against the possibilties of erroneous theoretical assumptions, provided they confine themselves to certain kinds of conclusions, viz., those that I shall characterize as resting upon *secure* premises. Through two episodes in the history of gravity physics, I show how physicists have developed strategies for securing their inferences regarding gravity against possibilities of error. I argue that these strategies should be construed as directed at *epistemically* *possible error scenarios*, i.e., ways that the world might be in an epistemically modal sense. Such strategies are central to the *secure evidence framework*.

The plan of the paper is as follows: In §2 I lay out the secure evidence framework for discussing the possibilities of error in drawing conclusions from experimental data. Within that framework, three types of strategies can be distinguished for securing evidence: strengthening, weakening, and robustness. The next two sections survey episodes from the history of gravity physics that exemplify these strategies and show how physicists have used them to confront the problem of relying on possibly erroneous theoretical assumptions. §3 discusses the strategies used by Newton to reason securely about gravity from the data available to him. In §4 I discuss the successors to Newton who are using a parametric framework to explore systematically the possibilities of error in General Relativity. I clarify the theoretical assumptions behind that framework, and show how physicists are able to deploy still further advances in the ability to relate data to fundamental physics to secure those theoretical assumptions. I also show that robustness plays a role in understanding how the conclusions they draw are compatible with the possibilities of error in those theoretical assumptions that remain. I summarize my conclusions in §5.

2 Securing evidence

The issues that are of concern here relate to at least two central problems for experimental science. (1) Given a body of data, what propositions may justifiably be inferred from those data? Alternatively, for what hypotheses do those data serve as good evidence? (2) Given a number of hypotheses of interest, which constitute potential answers to some question, how might one go about generating data that could serve as good evidence for or against one or more of those hypotheses?

For either question, the experimenter must consider what she may safely assume to be true. Even for purely statistical inferences from data, one must rely on assumptions that function as additional premises. Such assumptions might consist of invoking a statistical model of the data-generating process (as in the case of, for example, an inference using Neyman-Pearson statistics), or of a prior probability distribution across the hypotheses of interest (as in a Bayesian inference). Such statistical assumptions might themselves be founded upon material assumptions, as varied as the experimental techniques themselves, about the conditions under which data are gathered: are the reagents sufficiently pure? is the shielding against cosmic rays sufficient? are the responses of interview subjects influenced by the order in which the questions are asked? are there dietary differences between the control and treatment groups that contribute to any apparent difference in outcomes?

We must note at the outset the distinction between the exact truth of the explicit statement of an assumption and the *adequacy* of such a precisely stated assumption. That, for example, survey responses are independent of question order is a statement that can, in a given setting, be rendered with considerable precision. In practice, experimenters are often able to rely on such statements without them being, in the sense of that precise statement, exactly true. If survey response has a sufficiently weak dependence on question order, then perhaps one can simply ignore the dependence and proceed *as if* it did not exist at all.[1] Even so, truth remains relevant. What is being assumed to be true is not the precise statement 'survey response is independent of question order', but a less precise statement such as 'survey response may safely be taken, for present purposes, to be independent of question order'.

We can thus articulate a little more carefully the problem with which we are concerned. The problem of "theoretical error" is the problem that, in drawing inferences from data, investigators might draw upon theoretical assumptions that are *inadequate* for the inference at hand, in the sense that those assumptions are likely to lead investigators to misinterpret the data or to mistake the extent to which the data support their inferences.

Before moving forward, it will be helpful to distinguish between two phases or modes of reasoning with regard to experimental data. In what we might call the *use mode*, scientists use assumptions, including theoretical assumptions, to arrive at substantive conclusions from experimental data. In the *critical mode*, scientists turn their attention to those assumptions themselves, subjecting them to testing and criticism of various sorts. In the case of model-based statistical inference (e.g., Neyman-Pearson inference), this distinction corresponds to that between the primary inference, which is based on an assumed model, and "model criticism," which is concerned with the statistical adequacy of that model (Spanos 1999, 2008;

Staley 2012). At least implicitly, the use of a set of assumptions in arriving at substantive conclusions commits one to the claim that one knows enough for inferences based on those assumptions to be reliable. The critical mode is crucial for warranting such a commitment.

I wish to characterize the epistemic function of the critical mode of reasoning in experimental contexts in terms of the *securing of evidence*, a term that I will next undertake to explain.

The investigator who turns a critical eye toward the assumptions employed in her inferences from data must be concerned with the state of her own knowledge and the constraints it sets upon the ways in which her inferences might go wrong. We can think of her as contemplating the ways the world might be that would, were they actual, result in the falsehood of her claims about what the data support.

Here, the relevant modality of the word 'might' must be an epistemic modality. Merely logical possibilities of error can be safely ignored if one knows that they are not actual. The same can be said for counterfactual possibilities, while what philosophers like to call "metaphysical possibility" is simply too remote from the concerns of empirical science to be relevant to this question. There is an emerging literature and an ongoing disagreement about the semantics of epistemic modal propositions (D. Chalmers 2011; DeRose 1991; Hintikka 1962; Kratzer 1977; MacFarlane 2011; Salerno 2009). Fortunately, that debate can safely be ignored here. The relevant feature of epistemic modality for our purposes is that what is epistemically possible in epistemic situation K is limited by the propositions that are known in K. More precisely, *gaining* knowledge renders scenarios that were possible, impossible[2]. Conversely, knowing less means confronting a greater range of possibilities.[3]

In what follows I will apply the concept of epistemic possibility to *error scenarios*, and here I explain that term. If C is an evidence claim (i.e., a statement expressing a proposition of the form 'data $\mathbf{x}$ from test T are evidence for hypothesis H), and S is an epistemic agent contemplating asserting C whose epistemic situation (her situation regarding what she knows, believes, is able to infer from what she knows and believes, and her access to information) is K, then a way the world might be, relative to K, such that C is false, we will call an *error scenario* for C relative to K.

It is common in scientific discourse to find investigators, when presenting and justifying their inferences from data, addressing directly the question of what error scenarios might and might not be possible, given their epistemic situation (typically this is the epistemic situation of a collective agent, such as a research team). I propose that we can conceptualize the justificatory function of such practices with the help of the notion of security.

Suppose that, relative to a certain epistemic situation K, there is a set of scenarios that are epistemically possible, and call that set Ω_0. If proposition P is true in every scenario in the range Ω_0, then P is *fully secure* relative to K. If P is true across some more limited portion Ω_1 of Ω_0 (i.e., $\Omega_1 \subseteq \Omega_0$), then P is *secure throughout* Ω_1.

To put this notion more intuitively, then, a proposition is secure for an epistemic agent just insofar as that proposition remains true, whatever might be the case for that agent. Thus defined, security applies to any proposition, but the application of interest here is to evidence claims and inferences. Specifically, an *inference* from fact e to hypothesis h is secure relative to K insofar as the proposition 'e is good evidence for h' is secure relative to K.

It is important to stress that the methodological benefit of the security concept derives not from full security but rather from the ways in which various practices serve to *increase relative security.* I do not suppose that inquirers are ever called upon to determine the degree of security of any of their inferences. The methodologically significant concept turns out to be not security *per se*, but the *securing* of inferences, i.e., those practices that increase the relative security of an evidence claim.

Such practices can be classified as falling into two broad types of strategy: *weakening* and *strengthening.* In weakening, the conclusion of an evidential inference is logically weakened in such a way as to remain true across a broader range of epistemically possible scenarios than the original conclusion. Strengthening strategies operate by adding to knowledge, reducing the overall space of epistemically possible scenarios so as to eliminate some in which the conclusion of the evidential inference would be false.

A third security-related justificatory strategy, which we might label *robustness*, aims at the analysis and clarification of the security of an evidence claim, by showing, for some class of error scenarios, that an evidence claim put forth remains true in those scenarios. This strategy is often used in combination with either weakening or strengthening strategies. A significant literature has grown up around the issue of robustness, and the term has been given a variety of interpretations (see, e.g., Stegenga 2009; Woodward 2006, and Coko & Schickore 2011 for a survey). My use of it here is restricted: the robustness strategy with regard to security is a strategy whereby one clarifies the justification of an evidence claim by showing that the claim remains true across some range of error scenarios. I have previously argued that robustness of evidence can be epistemically valuable by contributing to the security of an evidence claim (Staley 2004).

In what follows I will discuss experimental undertakings in the face of uncertainty about theoretical uncertainty that illustrate the three strategies just mentioned.

3 Newton's arguments for the law of universal gravity: strengthening and weakening strategies

In Book III of *Principia* Newton lays out a step-by-step argument for his universal law of gravitation. Beginning with "phenomena" that describe the ways in which the planets and their respective satellites satisfy Kepler's laws, Newton argues for a series of propositions, which cumulatively add up to the claim that "gravity exists in all bodies universally and is proportional to the quantity of matter in each" (Proposition 8), and that this gravitational force is a force of attraction that is inversely proportional to the square of the distances between the centers of the bodies in question (Newton 1999, 802–811).

Here I do not propose to recapitulate Newton's primary positive argument for the universal law of gravity, which can be understood as having been advanced in the 'use mode' referred to above (Harper 2002, 2011). Rather, I wish to consider Newton's means, in the 'critical mode,' of securing the conclusion of that argument against possible error scenarios. Some of his responses to possible error scenarios draw upon aspects of his primary positive argument, but others require him to go beyond that argument.

First, could the universal law of gravitation be subject to a large error in the direction of the attractive gravitational force? This error Newton rules out on the basis of the fact that the rate at which areas are swept out by orbital radii can serve (Kepler's "harmonic law") as a *measurement* of the direction of the attractive force (Harper 2002, 175–177). Second, could the law be subject to a large error in its specification of the distance-dependence of attraction? Again, Newton can use a theory-mediated measurement as the basis for his rejection of this possibility: According to Corollary 1 of Proposition 45 in Book I, the power of the distance dependence is measured by the precession of the moon's orbit around the earth (Newton 1999, 539–545; Harper 2002, 180–181).

These points are crucial to Newton's primary argument for the law of gravity. By appealing to such methods, Newton is able to argue that these phenomena do constitute good evidence for the law of gravity. What I would like to consider a little more closely is how Newton, having made that argument, secures his inference against error scenarios that threaten the underlying theoretical premises of the argument.

The first such challenge was raised by Cotes. Cotes challenged Newton's application of the Third Law of motion (that "the actions of two bodies upon each other are always equal and always opposite in direction" (Newton 1999, 417)) to gravitation. Granting that one may justifiably apply the Third Law to contact forces, can one apply it to forces of attraction relating two bodies separated in space? If not, then this would undermine Newton's inference that makes gravitation universal: perhaps not all bodies exert gravitational forces upon one another. Cotes expressed this concern with the help of an imaginary scenario:

> Suppose two Globes *A* & *B* placed at a distance from each other upon a Table, & that whilst *A* remains at rest *B* is moved towards it by an invisible Hand. A bystander who observes this motion but not the cause of it, will say that *B* does certainly tend to the centre of *A*, & thereupon he may call the force of the invisible Hand the Centripetal force of *B*, or the Attraction of *A* since ye effect appears the same as if it did truly proceed from a proper & real Attraction of *A*. But then I think he cannot by virtue of the Axiom [Attractio omnis mutua est] conclude contrary to his Sense & Observation, that the Globe *A* does also move towards the Globe *B* & will meet it at the common centre of

Gravity of both Bodies. (Letter to Newton, March 18, 1713, quoted in Harper 2011)

Cotes' proposes an error scenario that threatens Newton's inference by calling into question a theoretical assumption about the applicability of the third law of motion to forces of attraction between bodies that are not in contact. The error scenario in question holds that what appear to be motions produced by mutually acting centripetal forces are really produced by mechanical pushes on the orbiting body. Since the appearance of a force exerted upon one body by another is, under this scenario, an illusion, the third law does not apply.

Were Newton to simply assume hypothetically that the third law applies in this case, as alleged by Stein (1990), his inference would fail to be secure against this error scenario. As shown by William Harper, however, Newton does have resources for answering Cotes' challenge. By appealing to these resources, Newton is able to strengthen his epistemic situation, ruling out the kind of error scenario invoked by Cotes (Harper 2011). I will not attempt here to rehearse Harper's account of Newton's response to Cotes' challenge in its entirety, nor to explain the arguments in detail. Nonetheless, what follows will, I hope, suffice to establish that Newton is able to use a strengthening strategy to secure his inference against this particular error scenario.

First, Newton is able to appeal to other kinds of attractive forces between spatially separated bodies that evidently *do* exert equal and opposite forces upon one another: namely, magnetic attractions between a lodestone and a sample of iron, as Newton found in his own experiments (Newton 1999, 427–28). Even if these motions are the result of some kind of ether, Harper argues, as the mechanical philosophy would require, the mutual endeavor of the two bodies to approach one another would still be subject to the third law. Cotes' invisible hand does not qualify as such a mutual endeavor.

Newton's second argument appeals to the gravitational equilibrium of the parts of the earth itself. Were not the attractive forces exerted by, say, one hemisphere upon the other, not equal and oppositely directed, then there would result a net acceleration of the planet as a whole (Newton 1999, 428).

Finally, Harper explains how applying the third law of motion to the Sun and Jupiter leads to convergent results when combining independently evaluated acceleration fields. In a nutshell, the argument goes as follows: Newton can use estimates of the distances between the Sun and the planets to estimate the centripetal acceleration field directed toward the Sun, yielding an estimate of Jupiter's acceleration towards the Sun. Likewise Newton can use James Pound's data for Jupiter's four moons to estimate the acceleration field centered on Jupiter. Using the third law to extrapolate this field to the distance separating the Sun from Jupiter yields an estimate of the Sun's acceleration towards Jupiter. Using the ratios of these two quantities, one can then define a common center of rotation about which the Sun and Jupiter will orbit with a common period. This yields measures of the weight of each of the two bodies towards the other, weights that remain oppositely directed to each other as the two bodies orbit their common center. As Harper writes,

> They, therefore, fulfill one major criterion distinguishing what Newton counts as attraction from Cotes' invisible hand pushing one body toward another. To have these oppositely directed weights count as a single endeavor of these bodies to approach one another requires, in addition, that they be equal so that they satisfy Law 3. (Harper 2011)

Newton thus shows how the assumption that the third law does apply to gravitational accelerations leads to convergent estimates of acceleration fields based on distinct bodies of data, adding, according to his own methodology, to the evidence for the premises of his main argument.

Next I would like to show how Newton uses a weakening strategy to respond to a different kind of challenge. In a weakening strategy, the investigator weakens the conclusion of an inference, or opts for the weaker of two possible conclusions to an inference, so that an error scenario that would threaten to undermine the inference to a stronger conclusion is no longer threatening. The scenario may remain possible, but even if it is actual the inference remains probative.

Consider the following error-scenario: Maybe it is changes in motion of invisible particles in a vortex, such as that postulated by Huygens, that pushes the planets into orbital motion, rather than mutual gravitation of the Sun and planets.

The first thing to be said here is that, as Harper explains, Huygens' theory itself predicted that the planets would follow exact Keplerian orbits. As such the theory was only successful insofar as it predicted orbits consistent with the data available at the time. But what Harper calls Newton's "richer" ideal of empirical success requires more than this. Newton's application of the third law of motion, for example, yielded converging measurements of relative intertial masses in the solar system, something that Huygens' theory was unable to accomplish. Later work demonstrating perturbations to Keplerian orbits for solar system bodies, beginning in the 1740's but becoming truly compelling only in 1785 with Laplace's "Théorie de Jupiter et de Saturne," subsequently eliminated Huygen's theory decisively (though it had ceased to attract adherents well before this) (Harper 2011, ch. 9.IV.2).

But Newton's weakening strategy is a response to a less well-defined challenge. This strategy is essentially captured in his famous "hypotheses non fingo." The challenge takes the form of the following potential error scenario (or more precisely, a class of potential error scenarios specified only by a feature that they all share): Perhaps *somehow* etherial particles act in such a way as to *produce* the mutual attractions that Newton describes using the law of gravity.

Here Newton's weakening strategy is his decision to opt for a conclusion from the data available to him that is weak enough to be compatible with this possibility. I refer to this as a *potential* error scenario because it would be an error scenario were Newton to draw the stronger conclusion from his data that bodies endeavor to approach one another with forces that are attributable to gravitational attraction *and nothing else*. That he does not intend for the universal law of gravity to be given such a strong interpretation is clearly indicated by his oft-quoted statement that "I have not as yet been able to deduce from phenomena the reason for these properties of gravity, and I do not feign hypotheses" (Newton 1999, 943). In other words, his conclusion should be understood as "Gravity really exists and acts according to the laws that we have set forth and is sufficient to explain all the motions of the heavenly bodies and of our sea," rather than as, "Gravity consists of the mutual attractions of bodies acting upon one another at a distance in a way that does not involve locally acting causes (such as etherial particles)."

To be sure, Newton's refusal to draw this stronger conclusion has an independent motivation insofar as he believes it would commit him to regarding gravitational attractions as instances of action at a distance, which, as expressed in a letter to the clergyman Richard Bentley, he regarded as an "absurdity" (Newton 2004, 102; Janiak 2010). Even had he not held this commitment, however, it would remain true that such a conclusion would have been stronger than warranted, precisely because he would not have been able to rule out the potential error scenario in question.

3.1 summary

Newton faced a significant problem of scope with regard to his application of the third law of motion to gravitational forces, because these forces acted between bodies that were not in contact and because it was difficult to test directly the hypothesis that the third law applied to gravitational attractions between celestial bodies. Newton was able to secure his assumption by appealing to the gravitational equilibrium of the parts of the earth and to the agreeing estimates of centripetal acceleration fields that resulted from thus applying the third law of motion. He could also appeal to the third law's applicability to similar terrestrial phenomena involving magnetism. Finally, he drew a conclusion that was sufficiently weak to remain compatible with possible mechanical explanations of the cause of gravity.

4 Parametric frameworks for testing fundamental physics: strengthening and robustness

The fate of Newton's conception of gravity as it was supplanted by Einstein's General Relativity exemplifies the problem of stability. That gravity should be thought of in terms of the curvature of spacetime rather than forces of mutual attraction is clearly an error scenario against which Newton's inferences were not secured, even though, as Harper argues, Newton's own methodology provided the engine of his own theory's demise (Harper 2007). In this section I will discuss the ways in which recent gravity researchers have secured their inferences regarding modern theories of gravity against error using a more systematic approach that constitutes a striking methodological advance. I will also argue that the implementation of the methods I describe involves ruling out possible errors in the assumptions of that method, and that the appropriate conception of the possibilities thus invoked is epistemic.

The Parametrized Post-Newtonian (PPN) formalism was developed to enable the comparison of metric theories of gravity with each other and with the outcomes of experiment, at least insofar as those theories are considered in the slow-motion, weak-field limit. Metric theories of gravity can be characterized by three postulates:

1. spacetime is endowed with a metric $\mathbf{g}$,

2. the world lines of test bodies are geodesics of that metric, and

3. in local freely falling frames (Lorentz frames) the nongravitational laws of physics are those of special relativity. (Will 1993, 22)

The PPN approach facilitates comparison of such theories using a common framework for writing out the metric $\mathbf{g}$ as an expansion, such that different theories are manifested by their differing values for the constants used in the expansion. As Clifford Will writes, "The only way that one metric theory differs from another is in the numerical values of the coefficients that appear in front of the metric potentials. The [PPN] formalism inserts parameters in place of these coefficients, parameters whose values depend on the theory under study" (Will 2006, 29).

Using her error-statistical framework, Deborah Mayo has emphasized the positive role played by the PPN framework in facilitating, not only the comparison of existing theories, but also the construction of new alternatives as a means of probing the various ways in which General Relativity (GR) could be in error. In addition, she argues that the resulting proliferation of alternatives to GR was not a manifestation of a theory in "crisis," but rather of an exciting new ability to probe

Parameter	What it measures relative to GR	Value in GR
γ	How much space-curvature produced by unit rest mass?	1
β	How much "nonlinearity" in the superposition law for gravity?	1
ξ	Preferred-location effects?	0
α_1	Preferred-frame effects?	0
α_2		0
α_3		0
α_3	Violation of conservation of total momentum?	0
ζ_1		0
ζ_2		0
ζ_3		0
ζ_4		0

Table 1: The PPN parameters, adapted from Will (2006).

gravitational phenomena and prevent the premature acceptance of GR. A key to the strength of this approach is the way in which the PPN formalism allows for the combination of the results of piecemeal hypothesis tests, not only to show that some possibilities have been eliminated, but to indicate in a positive sense the extent to which and ways in which gravitation is a phenomenon that GR (or theories similar to GR) gets right (Mayo 2002, 2009).

To take just one example of this use of the PPN framework to search for possible departures from the predictions of GR and thus set limits on the extent to which GR could be mistaken, consider the use of Very Long Baseline Interferometry (VLBI) to measure the PPN parameter λ, which measures the curvature of space. The relationship between λ and the predicted angle of deflection θ for an electromagnetic ray from a distant source due to the Sun is given by

$$\theta \cong \frac{(1+\lambda)GM_{\odot}}{c^2 b}(1+\cos\phi), \tag{1}$$

where G is the gravitational constant, c is the speed of light in a vacuum, $M_{\odot}$ is the mass of the Sun, b is the distance of closest approach from the center of the sun to the ray's path, and ϕ is the angle between the source of the ray and the sun as viewed from Earth.

General Relativity, in which the only dynamical field is the metric $\mathbf{g}$, assigns a value $\lambda = 1$, but some alternative theories of gravities include additional parameters that allow λ to take other values. In scalar-tensor theories the matter fields couple not only to the metric but also to an additional gravitational scalar field. In Brans-Dicke theory, for example, the effects of this additional scalar field show up in the form of an additional parameter ω_{BD}, such that $\lambda = (1 + \omega_{BD})/(2 + \omega_{BD})$. The greater the effects of the scalar field, the smaller the value of ω_{BD}, and the greater the departure of λ from unity.

Interferometry, which has been the basis of crucial experimental insights going back to the Michelson-Morley experiment, looks for differences in times of arrival for light signals originating from a single source that have traveled different paths. VLBI employs telescopes at different locations around the earth equipped with local atomic clocks as timing devices to allow for the combination of data from widely separated detectors. The idea behind using VLBI to measure λ, first suggested by Irwin Shapiro (I. I. Shapiro 1967), is to apply such interferometric techniques to light from celestial radio wave sources.

As one example of such an experiment, consider the results reported by S. S. Shapiro, Davis, Lebach, and Gregory (2004). They base an estimate of λ on twenty years' worth of VLBI data collected from 1979 to 1999, with signals from 541 sources producing data at 87 VLBI sites. This yielded 1.7×10^6 measurements. Bypassing the somewhat intricate error analysis attempting to take into account possible errors in assumptions regarding matters such as atmospheric propagation delay and atmostpheric refractivity gradients, I will simply note that they report an estimate of $\lambda = 0.9998_3 \pm 0.0004_5$, which is "within one standard deviation of the value predicted by GR" (S. S. Shapiro et al. 2004). (A value of $\lambda = 0.9994$ corresponds to $\omega_{BD} = 1.665 \times 10^3$.) As Will notes, for large values of ω, scalar-tensor theories make predictions for the current epoch that agree with those of GR "for all gravitational situations – post-Newtonian limit, neutron stars, black holes, gravitational radiation, cosmology" to within the order of ω^{-1} (Will 1993, 126). What Shapiro et al. thus allow us to conclude is that for all these phenomena, any effects of a scalar gravitational field, should there be one, will be such that we can rely on GR to give us the correct prediction to at least one part in a thousand.

That is what we may conclude, *in the use mode*, supposing that we may safely accept the assumptions of the PPN framework itself, from which the parameter measured in the experiment performed by Shapiro et al. derives its meaning. Recall that the PPN framework encompasses only *metric* theories of gravity. Such theories, which treat gravity as a manifestation of curved spacetime, satisfy the Einstein Equivalence Principle (EEP). EEP is in turn equivalent to the conjunction of three apparently distinct principles — Local Position Invariance (LPI), Local Lorentz Invariance (LLI) and the Weak Equivalence Principle (WEP).[4] Thus, experimental conclusions placing limits on the values of the PPN parameters rest upon a substantial theoretical assumption: The Einstein Equivalence Principle. In a shift from the use mode to the critical mode, Will writes, "The structure of the PPN formalism is an assumption about the nature of gravity that, while seemingly compelling, could be incorrect" (Will 1993, 207).

Will's comment involves two assertions that both demand attention in order to how experimental knowledge can be secure in the face of possible theoretical error. First, he claims that EEP is "seemingly compelling." Secondly, he admits that it “could be incorrect.”

Taking these claims in order, it first must be clarified that Will’s use of the term “seemingly” is not meant to suggests that the support for the EEP is merely illusory. Indeed, he elsewhere writes that there is sufficient evidence that “gravitation . . . must be described by a ‘metric theory’ of gravity” (ibid., 10) and that this evidence “supports the conclusion that the only theories of gravity that have a hope of being viable are metric theories, or possibly theories that are metric apart from very weak or short-range non-metric couplings (as in string theory)” (Will 2006, 26). (I comment below on the latter qualification.) It is evident that Will maintains that we *may*, at least for the purposes of using the PPN parameters, safely assume the EEP. The basis for this claim is that the EEP itself has been subjected to some very high precision tests that can be utilized in the context of other parametric frameworks that facilitate the testing of WEP, LLI, and LPI.

For example, in 1973, Lightman and Lee developed the $TH\epsilon\mu$ formalism, which functions analogously to the PPN framework for tests of GR (Lightman & Lee 1973).[5] The class of theories that can be described within the $TH\epsilon\mu$ formalism includes all metric theories. It also includes many, but not all, non-metric theories.[6] The ability to put non-metric theories into a common framework such that limitations can be put on EEP violations in a systematic way provides a powerful extension of the program of testing within PPN.

This formalism has proven to be adaptable to the pursuit of tests of null hypotheses for each of the components of EEP. By taking various combinations of the four $TH\epsilon\mu$ parameters, one can define three "non-metric parameters," Γ_0, Λ_0, and Υ_0, such that if EEP is satisfied then $\Gamma_0 = \Lambda_0 = \Upsilon_0 = 0$ everywhere. Tests of the components of EEP can then be investigated in terms of null tests for these parameters. A non-zero value for Υ_0 is a sign, for example, of a failure of LLI. Will describes how the results of the Hughes-Drever experiment ("the most precise null experiment ever performed" (Will 1993, 31)), can be analyzed so as to yield an upper bound of $\Upsilon_0 < 10^{-13}$ and concludes that "to within at least a part in 10^{13}, Local Lorentz Invariance is valid" (ibid., 62). Eötvös experiments have tested WEP and yielded limits on "non-metric parameters" of $|\Gamma_0| < 2 \times 10^{-10}$ and $|\Lambda_0| < 3 \times 10^{-6}$.

More recent tests of Lorentz invariance have employed another formalism, the Standard Model Extension (SME) (Colladay & Kostelecký 1998; Kostelecký 2004). Because violations of Lorentz invariance are expected in a number of proposals for physics beyond the Standard Model (SM), experimental tests that might reveal such violations without having to assume any particular non-Standard Model dynamics offer the possibility of insights into such daunting theoretical problems as quantum gravity. The SME uses an effective field theory approach to generalize the SM and GR so as to allow for the quantitative description of violations of both LLI and CPT invariance, adding new terms to the SM that allow for CPT and Lorentz violation in different sectors. Experimental tests can then be used to set limits on the coefficients for these terms. For example a clock-comparison experiment like the Hughes-Drever experiment looks for variation in the frequency of a clock (typically an atomic transition frequency) as its orientation changes. A descendant of the Hughes-Drever experiment, using a ^{3}He/^{129}Xe maser system, has yielded a constraint on a combination of the Lorentz-violating coefficients in the neutron sector of the SME (coefficients characterizing a possible Lorentz-violating coupling of neutrons to a possible background tensor field traceable to spontaneous symmetry breaking in a fundamental theory that need not be specified) at the level of 10^{-32} (Bear, Stoner, Walsworth, Kostelecký, & Lane 2000, 2002).

It is this test and myriad other high-precision tests of the components of EEP (Mattingly 2005; Will 2006) that stand behind Will's assertion of the "seemingly compelling" evidence for EEP. We must now turn to his comment that, nonetheless EEP "could be incorrect." An alternative formulation of this statement is: "It is possible that the EEP is false." But in what sense of "possible"?

We can quickly dispense with an easy question: Is it logically possible that the EEP is false? Surely, it is. (More carefully, we haven't any reason to think that the denial of EEP in some hidden way harbors a logical contradiction.) But that is not relevant to whether we may safely assume EEP to be true. We can also quickly dispense with a difficult question: Is it physically possible that EEP is false? Assuming the standard view that what is physically possible is whatever is not ruled out by the laws of physics, the answer is that we do not know. It is precisely to *find out* whether violations of EEP are permitted by the laws of physics that we wish to test the EEP.

I would suggest that the important sense in which we need to consider the possibility of the failure of EEP is an epistemic one: Given what we do know, does it remain possible that the EEP is false? Not, presumably, if we know that EEP is true. This would constitute an application of what is sometimes called "Moore's Principle," which states that if a speaker truthfully asserts that she knows that a certain proposition P is true, then she is not in a position to assert that P might be false (DeRose 1991). I do not propose here to defend the truth of Moore's Principle, but only wish to claim that, provided that the modality implicated is epistemic, an assertion by S that P might be false is seriously in tension with S's truthful assertion that she knows P.

However, there is a difference between knowing that EEP has been well-tested and knowing that it is true full-stop. There are, as it turns out, two distinct bases for the statement that "EEP might be false." One basis is that of the determined skeptic, and falls prey to all of the difficulties that the skeptic faces. The second is more scientifically relevant, but poses no threat to the possibility of gaining experimental knowledge about gravity by means of tests of hypotheses of the PPN parameters.

First, there is the skeptic's gambit: Given any finite body of data that appears to lend support to a theory, it is always possible to come up with *some* other theory that can be made compatible with those data. To this one appropriate response is to say "Yes, but so what?" From Newton's fourth rule of reasoning (Harper 2007, 2011; Newton 1999) to Peirce's injunction that one must not "block the road of inquiry" to Deborah Mayo's critique of "gellerized" alternative hypotheses (Mayo 1996), a long tradition in philosophy of science has equipped us to provide a methodological critique of the strategy of avoiding experimental conclusions by invoking the mere possibility of alternative hypotheses. Here, Mayo's criticism will serve us nicely: the strategy is unreliable (Mayo 1996, ch. 6). More specifically, a person who is determined to avoid conclusions drawn from data because of the possibility of alternative hypotheses that can be constructed after the fact to fit those data will always fail to accept hypotheses that are true.

But, inspired perhaps by Moore's Principle cited above, one might object that admitting an alternative to EEP, compatible with all of the data generated in tests of EEP, to be *possible* amounts to admitting that we do not know EEP is true, and thus we cannot rely on EEP in the interpretation of tests conducted within the PPN framework. Clifford Will himself has written that "Nothing can be categorical, of course, because given any finite experimental error, one could always in principle conceive of a non-metric theory that satisfies all tests of EEP" (personal correspondence). Here there is a danger of landing in a philosophical muddle. We have conflicting accounts of knowledge and of epistemic possibility. Are we doomed to similar conflict over the possibility of knowledge about gravity?

The first reason I think that we are not thus doomed is that we have good reason to dismiss most such alternatives, which would have to be so extremely gerrymandered as to be utterly implausible (Staley 2008). These alternatives fall into the category of "conspiracy theories" — effectively, they suppose that Nature has gone to great lengths to make things look as if EEP is satisfied everywhere that we do look, and hides the exceptions from us very cleverly.

The second reason we are not doomed is closely tied to the second basis for the admission that, after all, EEP "could be incorrect." We in fact do not know, and, for the purposes of interpreting PPN experiments, do not need to know that EEP is *exactly* correct. Many physicists expect EEP in some way or other to fail to hold in the regime of quantum gravity. Indeed, the very meaning of what a metric theory is becomes unclear in these contexts. Clifford Will notes that string-inspired theories

> are 'metric' in a deep sense, because they are built upon a metric foundation; however they can have additional fields (dilatons, moduli) that couple to matter in a non-universal way, and can lead to violations of EEP. One line of reasoning would treat these fields as gravitational fields, in the same class as the metric, and thus would call these theories non-metric. But another viewpoint would treat these fields as additional (admittedly exotic) MATTER fields, no different, really, from the electromagnetic field, which obviously couples to matter in a non-universal way. (personal correspondence)

PPN allows us to restrict our attention to settings in which these kinds of concerns can be set aside.

This final point can be appreciated as an application of a *robustness* strategy. There are indeed error scenarios in which the EEP fails to hold, but because the

PPN is meant to capture the description of gravitational phenomena in the slow-motion, weak-field domain in which such violations will not be manifest, the conclusions drawn via PPN would remain the same across such error scenarios.

4.1 summary

The PPN framework succeeds in part by restricting its attention to only metric theories of gravity, which amounts to assuming the Einstein Equivalence Principle, which might not be true. Securing the PPN results that allow us to draw conclusions about gravity draws upon both strengthening and robustness strategies. The strengthening strategy uses tests of the EEP commitments to eliminate possible scenarios that would undermine the conclusions drawn in the PPN framework. There are, however, possible scenarios, and serious ones, in which the EEP fails to hold exactly, or in which it is unclear what the EEP requires. But the conclusions drawn about the values of the PPN parameters remain valid in all of these scenarios, which do not make a difference to the slow-motion, weak-field approximation on which the PPN is based. This latter point invokes robustness as a means of safeguarding experimental knowledge about gravity against possibly erroneous theoretical assumptions.

5 Conclusion

These developments in the experimental study of gravity exemplify the importance of theoretical assumptions in drawing conclusions from experimental data. They also show that efforts to secure those assumptions so that investigators have the epistemic resources to reason from safely-held assumptions are equally a part of the legacy of experimental physics, as they presumably are of other experimental sciences.

If the challenges of reasoning from experimental data in the face of uncertain theoretical assumptions have been more or less constant, the resources developed for confronting those challenges have hardly been stagnant. The proliferation of parametric frameworks for bringing experimental data into contact with fundamental physical theories constitutes a recent example of *progress* at the interstices between experimental technique and theoretical elaboration. That such progress in the methodology of bringing data to bear on theory has a life of *its* own, and hence a history to be explored, is a promising prospect for historians and philosophers of the experimental sciences.

I have also argued that these efforts to secure inferences about gravity against possibilities of error should be construed as directed at error scenarios that are possible in the epistemic sense, as opposed to other modalities that have preoccupied philosophers. A recent surge of philosophical interest in epistemic modality has occurred independently of considerations in the philosophy of science and without the participation of philosophers of science (Egan & Weatherson 2011). The present discussion serves, however, to demonstrate the relevance of epistemic modality to understanding how scientific investigators warrant inferences from data.

References

Bear, D., Stoner, R., Walsworth, R., Kostelecký, V., & Lane, C. (2000). Limit on Lorentz and CPT violation of the neutron using a two-species noble-gas maser. *Physical Review Letters*, *85*, 5038–5041. doi: 10.1103/PhysRevLett.85.5038

Bear, D., Stoner, R., Walsworth, R., Kostelecký, V., & Lane, C. (2002). Erratum: Limit on Lorentz and CPT violation of the neutron using a two-species noble-gas maser. *Physical Review Letters*, *89*, 209902. doi: 10.1103/PhysRevLett.89.209902

Chalmers, A. (2009). Can scientific theories be warranted? In D. G. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, objectivity, and rationality* (pp. 58–72). New York: Cambridge University Press.

Chalmers, D. (2011). The nature of epistemic space. In A. Egan & B. Weatherson (Eds.), *Epistemic modality.* Oxford: Oxford University Press.

Coko, K., & Schickore, J. (2011). *Robustness and its kith and kin: A meta-analysis.*

Colladay, D., & Kostelecký, V. A. (1998, Oct). Lorentz-violating extension of the standard model. *Physical Review D*, *58*, 116002. Retrieved from `http://link.aps.org/doi/10.1103/PhysRevD.58.116002` doi: 10.1103/PhysRevD.58.116002

DeRose, K. (1991). Epistemic possibilities. *The Philosophical Review*, *100*, 581–605.

Egan, A., & Weatherson, B. (Eds.). (2011). *Epistemic modality.* New York: Oxford University Press.

Franklin, A. (1986). *The neglect of experiment.* New York: Cambridge University Press.

Galison, P. (1987). *How experiments end.* Chicago: University of Chicago Press.

Galison, P. (1997). *Image and logic: A material culture of microphysics.* Chicago: University of Chicago Press.

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

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions.* New York: John Wiley and Sonss.

Harper, W. (2002). Newton's argument for universal gravitation. In I. B. Cohen & G. E. Smith (Eds.), *The cambridge companion to newton* (pp. 174–201). New York: Cambridge University Press.

Harper, W. (2007). Newton's methodology and Mercury's perihelion before and after Einstein. *Philosophy of Science*, *74*(5), 932–942. Retrieved from http://www.jstor.org/stable/10.1086/525634

Harper, W. (2011). *Isaac Newton's scientific method: Turning data into evidence about gravity and cosmology.* New York: Oxford University Press.

Heidelberger, M. (2003). Theory-ladenness and scientific instruments in experimentation. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 138–151). Pittsburgh, PA: University of Pittsburgh Press.

Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions.* Ithaca, NY: Cornell University Press.

Huber, P. (1981). *Robust statistics.* New York: John Wiley and Sons.

Janiak, A. (2010). *Newton as philosopher.* Cambridge: Cambridge University Press.

Kostelecký, V. A. (2004, May). Gravity, Lorentz violation, and the standard model. *Physical Review D*, *69*, 105009. Retrieved from `http://link.aps.org/doi/10.1103/PhysRevD.69.105009` doi: 10.1103/PhysRevD.69.105009

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, *1*, 337–55.

Lightman, A., & Lee, D. (1973). Restricted proof that the weak equivalence principle implies the Einstein equivalence principle. *Phys. Rev. D*, *8*, 364–376.

MacFarlane, J. (2011). Epistemic modals are assessment-sensitive. In A. Egan & B. Weatherson (Eds.), *Epistemic modality.* Oxford: Oxford University Press.

Mattingly, D. (2005). Modern tests of Lorentz invariance. *Living Rev. Relativity*, *8*. Retrieved from `http://www.livingreviews.org/lrr-2005-5`

Mayo, D. G. (1996). *Error and the growth of experimental knowledge.* Chicago: University of Chicago Press.

Mayo, D. G. (2002). Theory testing, statistical methodology, and the growth of experimental knowledge. In P. Gärdenfors, J. Wolinski, & K. Kijania-Placek (Eds.), *In the scope of logic, methodology, and philosophy of science* (pp. 171–190). Dordrecht: Kluwer.

Mayo, D. G. (2009). Learning from error, severe testing, and the growth of theoretical knowledge. In D. G. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, objectivity, and rationality* (pp. 28–57). New York: Cambridge University Press.

Newton, I. (1999). *The principia: Mathematical principles of natural philosophy* (I. B. Cohen & A. Whitman, Eds. & Trans.). Berkeley: University of California Press.

Newton, I. (2004). *Philosophical writings* (A. Janiak, Ed.). Cambridge: Cambridge University Press.

Radder, H. (2003). Technology and theory in experimental science. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 152–173). Pittsburgh, PA: University of Pittsburgh Press.

Salerno, J. (2009). *Epistemic modals: Relativism and ignorance.* (Talk given at Midwest Epistemology Workshop, Saint Louis University, September 2009)

Shapiro, I. I. (1967). New method for the detection of light deflection by solar gravity. *Science*, *157*(3790), 806–808. Retrieved from http://www.jstor.org/stable/1721384

Shapiro, S. S., Davis, J. L., Lebach, D. E., & Gregory, J. S. (2004, Mar). Measurement of the solar gravitational deflection of radio waves using geodetic very-long-baseline interferometry data, 1979–1999. *Physical Review Letters Rev. Lett.*, *92*, 121101. Retrieved from http://link.aps.org/doi/10.1103/PhysRevLett.92.121101 doi: 10.1103/PhysRevLett.92.121101

Spanos, A. (1999). *Probability theory and statistical inference.* Cambridge: Cambridge University Press.

Spanos, A. (2008). *Misspecification, robustness, and the reliability of inference: The simple t-test in the presence of markov dependence.* Retrieved from http://www.econ.vt.edu/faculty/2008vitas_research/spanos_working_papers/2Spanos-reliability.pdf

Staley, K. W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science*, *71*, 467–488.

Staley, K. W. (2008). Error-statistical elimination of alternative hypotheses. *Synthese*, *163*, 397–408.

Staley, K. W. (2012). Strategies for securing evidence through model criticism. *European Journal for Philosophy of Science*, *2*, 21-43. Retrieved from http://dx.doi.org/10.1007/s13194-011-0022-x (10.1007/s13194-011-0022-x)

Stegenga, J. (2009). Robustness, discordance, and relevance. *Philosophy of*

Science, *76*(5), 650–661. Retrieved from
http://www.jstor.org/stable/10.1086/605819

Stein, H. (1990). "From the phenomena of motions to the forces of nature": Hypothesis or deduction? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1990*, pp. 209-222. Retrieved from http://www.jstor.org/stable/193069

Will, C. M. (1993). *Theory and experiment in gravitational physics* (Second ed.). New York: Cambridge University Press.

Will, C. M. (2006). The confrontation between general relativity and experiment. *Living Reviews in Relativity*, *9*(3). Retrieved from http://www.livingreviews.org/lrr-2006-3

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, *13*, 219–40.

Notes

[1]Whether, in a given case, the approximate satisfaction of an assumption makes that assumption safe to use for purposes of inference is not easily decided. The assumption that approximately satisfied premises lead to approximately correct conclusions is a major source of error in statistical inference and is the motivation for the development of both robust statistics (Hampel, Ronchetti, Rousseeuw, & Stahel 1986; Huber 1981) and mis-specification testing (Spanos 1999, 2008) (see Staley 2012 for a discussion).

[2]Occurrences of the word "possible" and its cognates should henceforth be understood epistemically, unless otherwise specified.

[3]It should not be assumed that what is epistemically possible for an epistemic agent will always be known to be epistemically possible for that agent. Arguably, nothing that Isaac Newton knew, and no information to which he had access, ruled out the possibility that gravity is a manifestation of the curvature of space. There is a sense of possibility, distinct from that here invoked, in which curved space *became* possible much later. To my knowledge, the semantics of this latter notion have not been seriously tackled by philosophers. For purposes of the present discussion, I will assume that *discovering* a new possibility does not *create* a new

possibility.

[4]I note here how these principles might be stated: *WEP*: if an uncharged test body is placed at an initial event in spacetime and given an initial velocity there, then its subsequent trajectory will be independent of its internal structure and composition. *LLI*: the outcome of any local nongravitational test experiment is independent of the velocity of the experimental apparatus. *LPI*: the outcome of any local nongravitational test experiment is independent of its spacetime location (Will 1993, 22).

[5]See (Mattingly 2005) for a recent review of this and other parametric frameworks as applied to tests of Lorentz invariance.

[6]The restriction, more specifically, is to theories that describe the center-of-mass acceleration of an electromagnetic test body in a static, spherically symmetric gravitational field, such that the dynamics for particle motion is derivable from a Lagrangian. The parameters T and H appear in the Lagrangian; ϵ and μ appear in the "gravitationally modified Maxwell equations" (GMM). Lightman and Lee argue (in 1973) that "all theories we know of" have GMM equations of the type needed, and that all but one theory (which they treat separately) can be represented in terms of the appropriate Lagrangian (Lightman & Lee 1973).