



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Revisiting Accepted Science: The Indispensability of the History of Science

Author(s): George E. Smith

Source: *The Monist*, OCTOBER 2010, Vol. 93, No. 4, Philosophical History of Science (OCTOBER 2010), pp. 545-579

Published by: Oxford University Press

Stable URL: <http://www.jstor.com/stable/27904167>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Oxford University Press is collaborating with JSTOR to digitize, preserve and extend access to *The Monist*

REVISITING ACCEPTED SCIENCE: THE INDISPENSABILITY OF THE HISTORY OF SCIENCE

1. *Theses*

My theses are synopsized in my title, so let me begin by expanding on it, starting with the subtitle. The main point of the paper is to argue for the indispensability of the history of science to the philosophy of science, yet by the end I hope to make clear how the converse holds as well. Because both history and philosophy of science involve diverse pursuits, my claim of their mutual indispensability applies only insofar as each of them concerns itself with the nature and scope of the “knowledge” achieved in modern science. (The shudder quotes serve not to foreclose from the outset on those who question whether the word is strictly applicable to it.) Although adopting Kant’s phrasing exaggerates the situation, in some respects, I am going to argue, philosophy of science without history of science is empty, and history of science without philosophy of science is blind. My hope for some time has been a rapprochement between the two. The crucial step, nevertheless, is for philosophers of science to recognize a need for a certain sort of history of science. The principal task of the paper will be to spell out just what sort that is.

Turning to the main title, the key term is ‘revisiting’. I cannot explain it, however, until I first clarify ‘accepted’, which I use a little differently from most others. The underlying thought is common with the usual use of the term: a claim is accepted when questions to which it is an answer are taken, at least provisionally, to have been settled. Here, nevertheless, I intend the term to apply more narrowly: a claim has become accepted within a scientific community when the community begins to presuppose it as a *constitutive* element in their ongoing research. If that research is challenged on the grounds that there is not yet sufficient warrant for presupposing the claim, then it has not yet been accepted. Equally, if difficulties emerge in

“Revisiting Accepted Science: The Indispensability of the History of Science” by George E. Smith,
The Monist, vol. 93, no. 4, pp. 545–579. Copyright © 2010, THE MONIST, Peru, Illinois 61354.

the research, then the presupposed claim has not yet been accepted unless it is *prima facie* immune from being viewed as responsible for them. In analogy with the presumption of innocence until proved guilty, a scientific community has accepted a claim only when they grant it a presumption of truth until ongoing research gives clear evidence to the contrary.

Being accepted, as I use the term, has nothing as such to do with what individual scientists happen to believe. To offer a trifling example, once the galvanometer became the standard means for measuring electric current in the 1830s and 1840s, Ampère's law had become accepted even if those using the instrument were entirely unaware of its presupposing that Ampère's law holds to very high precision. Similarly, from the 1660s to the 1960s, the period and length of pendulums served to determine the strength of surface gravity. I wonder how many who used the resulting values gave any thought to the fact that those values were predicated on Galilean gravity—that is, uniform gravity acting along parallel lines—in direct contradiction with Newtonian centrally directed inverse-square gravity, the strength of which is what they thought was being measured. Being accepted, as I use the term here, has only to do with what the community is prepared to take for granted, perhaps tacitly, in ongoing research, and not with how individuals in the community thought about the matter.

The truly narrow aspect of my use of 'accepted' derives from the requirement that it be a *constitutive* element in ongoing research, and not merely *heuristic*. Heuristic elements, no matter how much they are taken for granted and how heavily they are relied on to guide research, can subsequently be discarded without jeopardizing the results obtained from the research; to discard a constitutive element, however, *prima facie* undercuts the results that presuppose it.¹

An example I have discussed at length elsewhere² involves the turn-of-the-century measurements J.J. Thomson and others made of the mass-to-charge ratio of the negatively charged constituents of cathode rays, of thermionic and photoelectric emissions, and of the discharge from uranium—that is, of what subsequently became known as the electron. Those measurements constitutively presupposed that the negatively charged constituents in question satisfy certain laws that theretofore had been shown to hold for charged particles in motion. Everyone making the measurements at the time surely thought of those constituents as particles, thereby precluding their exhibiting such wave-like behavior as diffraction.

That this was a discardable heuristic emerged only in the 1920s when electrons were experimentally confirmed to undergo diffraction, yet none of the measurements of their mass-to-charge ratio made over the thirty prior years were affected.³

A more nuanced example is the nineteenth century etherial substrate of Fresnel's transverse waves in optics. Mary Somerville in 1840 remarked, "The existence of an ethereal fluid is now proved,"⁴ and went on to speculate on its long-term effects on the motions of orbiting bodies. The "proof" at the time consisted merely of an inability to conceptualize the transverse waves without a medium to support them—this quite independently of the subsequent efforts to treat electromagnetism in terms of nonuniformities in the ether. Yet virtually all of the results in wave optics during the nineteenth century remained intact after the ether was abandoned at the end of the century, for the ether itself was entering only heuristically into those results, and not constitutively. By contrast, it entered constitutively into the Michelson-Morley experiment, which is why the null results of that experiment gave reason to reconsider it.

The ether example highlights my reason for confining "accepted" to claims that enter constitutively into ongoing research: That research puts them to a test in ways that it does not put heuristic elements. More of that in a moment, however. For now let me just concede that the distinction between constitutive and heuristic elements is not at all straightforward in practice. Whether an element is entering ongoing research constitutively or heuristically has to be an historical question not only because it is a question about *ongoing* research, and hence the answer can change over time, but even more so because the answer can vary from one piece of research to another. It is also a philosophical question because it involves a retrospective "rational reconstruction" of earlier evidential reasoning, as illustrated by the wave-like features of the electron and the Michelson-Morley experiment, even though in those cases, and most others in which something has gone wrong in research, the reconstruction was carried out by scientists. Because the rest of this paper will be about how to combine historical and logical analyses of evidence within extended research traditions, and the constitutive-heuristic distinction requires just such analyses to resolve it on a case by case basis, I am not going to elaborate further on it now.⁵

Anyway, the most important aspect of accepted results for my purposes does not depend on that distinction. If the last fifty years of studies in the

history of science have taught us nothing else, they have taught us that a great many fundamental claims of science first became accepted, in my sense, on the basis of remarkably little evidence. Examples abound: the law of inertia and Boyle's law in the seventeenth century; more recently Einstein's general theory of relativity, which initially had only the small anomaly in the precession of the perihelion of Mercury and a few problematic measurements of the bending of light in support of it; and Bohr's model of the atom, the only empirical test of which at the time it became accepted in my sense was a theoretical value of Rydberg's constant that differed from the measured value by six percent. Steven Shapin and Simon Schaffer made the limitations of evidence behind accepted results of experiments the central theme of their *Leviathan and the Air-Pump*. Several other historians of science, following their lead, have appealed to the limited evidence scientific communities require to accept not just experimental results, but all sorts of other fundamental claims, to argue that modern science does not merit the lofty epistemic standing that scientists have long claimed for it.

Some philosophers have reacted to that view by trying to show, in the face of mounting historical evidence to the contrary, that science really does have strict canons of acceptance that set it apart from all other forms of inquiry and give it a unique claim to achieving knowledge. I am not going to take that tack here. I agree with historians and sociologists of science that the evidence available at the time many of the most fundamental principles of science, and many results of experiments as well, became accepted in my sense fell far short of showing, with any finality, that those principles and results were true.

I have worded the scope of my agreement with historians and sociologists of science carefully. I am not granting them their customary further claim that acceptance of results in science is dominated by political, sociological, and other extrascientific considerations that are orthogonal to the goal of acquiring knowledge. To the contrary, my view is that, more often than not, the claims they say are accepted in the face of inadequate evidence are in fact accepted on proper evidential warrant. The warrant in question is not, however, for the final truth of those claims, or even for a high probability of final truth. Rather, the evidence gives warrant for accepting the claims as constitutive working hypotheses enabling further research. That is, (1) the evidence gives *prima facie*

support for the claim; (2) predicating further research on the claim has promise for markedly increasing the ability to marshal evidence within that research; and (3) the research in question promises to include ways of quickly exposing shortcomings in the working hypothesis that would otherwise threaten to lead down a long garden path—that is, an extended body of research that ultimately has to be thrown out because it was predicated on a fundamentally mistaken working hypothesis. Accepting a claim that meets these three requirements has everything to do with the goal of acquiring knowledge, though admittedly whether what emerges is actually knowledge depends overwhelmingly on the future research itself, and not on the grounds for accepting the working hypothesis on which it is predicated.

Which brings me to the key word of my title, ‘revisit’. Everyone knows that accepted science sometimes gets revisited, for Einstein’s theory of gravity replaced Newton’s, the ideal gas law was found not to hold precisely of any real gas, and the classical principles presupposed in the Rayleigh-Jeans radiation law gave way to radical new principles with Planck’s law. My point is broader than that. If some accepted principle is constitutively presupposed in ongoing research, then in some respect or other it is generally being tested in that research. Granted, the aim of the research is rarely to test that principle, for it has become accepted. Rather, the test is indirect, *en passant*, as it were. No one usually even notices that the accepted principle is being tested unless some problem emerges, an anomaly that turns out to be sufficiently recalcitrant that the assumptions underlying the research come to be reconsidered.

But what about when no problem or anomaly emerges? To some extent or other the presupposed principle has passed a test, and, just as interest accrues to money in a bank, further evidence has accrued supporting it. The further evidence, like the interest I am now getting from my bank, may amount to next to nothing. Millions of galvanometers are out there measuring electric current, but they are scarcely making much difference to the body of evidence supporting Ampère’s law. I am going to argue, however, that ongoing research can subject accepted science presupposed in it to far more stringent *en passant* tests than is obvious at first glance. The only way to tell is to examine the details of that research.

So far I have said little beyond what Pierre Duhem pointed out a century ago: principles of accepted science continue to be tested to the

extent that they enter into ongoing research, and hence the evidence for them typically grows with time. As he noted, the evidence for Newton's laws of motion in 1787, when Laplace thought he had removed the last orbital discrepancy with his explanation of the secular acceleration of the Moon, was far greater than when the *Principia* first appeared in 1687, and it was far less than the evidence for them a century later in 1887, at the time of the Michelson-Morley experiment. I want to say more than that:

Whatever claim any of the sciences have to being epistemically different from other areas of inquiry, it lies at least as much (usually far more) in the way in which they have constantly revisited previously accepted claims, testing them anew—often much more stringently than those claims had been tested when they first came to be accepted—as it lies in any canons of initial acceptance; and this even when the testing has been only *en passant*.

But then philosophers of science who want to assess the nature of the knowledge achieved in any area of research are going to have to examine the history of that research, determining how evidence for various claims has or has not grown over time during the course of it.

That is the sense in which I say that history of science—more narrowly, the history of evidence—is indispensable to philosophy of science. But a straightforward history of research in an area is generally by itself not going to tell us how stringently various claims have continued to be tested, for the research is not focused on testing those claims, but instead on discovering new things about the world. Careful critical analysis of the details of the research is needed to reveal the ways in which accepted claims are being tested in the process and the stringency of those tests. Simply put, history of science is indispensable to the philosophy of science because the question, “How has evidence for a claim grown over the course of continuing research?” is an historical question; and philosophy of science is indispensable to history of science because critical analysis of the sort philosophers typically engage in is needed to assess the evidence, especially when continued testing is *en passant*. That is why we need a rapprochement between history and philosophy of science.

I am going to defend these theses in two steps in the rest of the paper. First I am going to appeal to my studies of the history of Newtonian orbital mechanics to demonstrate how rewarding a history of evidence can turn out to be. Then I am going to offer a few examples involving other

areas of science that have raised important philosophical questions which, I claim, can be answered only through similar historical studies.

2. *Revisiting Newton's Theory of Gravity*⁶

No one would deny that Newton's theory of gravity continued to be tested as calculations based on it were made of orbital motions in our solar system and compared with the observed motions. Although such calculations began in Book 3 of the *Principia*, they became central to orbital astronomy only after the Euler-Mayer lunar theory of the early 1750s and Laplace's breakthroughs on the Jupiter-Saturn interaction and the secular acceleration of the Moon in the mid-1780s. They have remained central ever since. Two moments in that history are usually singled out: the discrepancy between the calculated and observed motions of Uranus that led to the discovery of Neptune in the mid-1840s—which Norwood Russell Hanson called the “zenith” of Newtonian gravity; and the 43 arc-second per century residual discrepancy between the calculated and observed precession of the perihelion of Mercury—Hanson's “nadir” of Newtonian gravity.⁷ Other than these two, the view, at least among philosophers of science, seems to be that the only advance from the first two centuries of comparing calculated with observed orbits was increasingly closer agreement as the calculations were refined and the observations became more precise. In other words, the exactness of the theory was being tested, in the process markedly tightening the bands of accuracy with which it could legitimately be said to hold, but nothing more.

A decade ago I began looking at the history of post-Newtonian orbital research, asking if that was the only respect in which his theory was tested over its first two hundred years. I came away with seven surprises that, as a matter of autobiographical fact, are what convinced me of the value of the detailed study of the history of evidence in individual areas of research.

My first surprise was that the “test-question” was not, “Do the calculated motions agree with the observed motions?,” but instead “Can robust physical sources compatible with Newtonian theory be found for each clear, systematic discrepancy between the calculated and the observed motions?” By “robust” I mean a source that has other observable and hence confirmable consequences besides accounting for the discrepancy. I should have realized that this was the test-question all along

just from comparing the Neptune and Mercury-perihelion cases. What finally drove it home was the realization that there was at least one widely recognized discrepancy between calculated and observed orbits from 1687 until 1993, save for a quarter century beginning in 1787 during which Laplace's breakthroughs were being assimilated and the comparisons that were being made were more anticipatory than meaningful. The small anomaly in the precession of the perihelion of Mercury proved to be the "nadir" of Newtonian gravitation only after the failure of a half-century effort to find a source for it compatible with Newtonian theory. All the other discrepancies ended up revealing some detail of our planetary system, the least subtle of which was Neptune, that theretofore had not been taken into account in the calculations.

The reason why the test-question cannot simply concern agreement with observation is worth noting. The calculations based on Newtonian theory not only require a good deal of empirical information, like the masses of the planets. They also include a proviso, to use Carl Hempel's term, that no other forces need to be taken into account.⁸ Agreement with observation would have consequently been evidence at least as much for this proviso as for the accepted theory. The different test-question puts those engaged in the recondite work of calculating orbits in a different light. They were not just retesting Newtonian theory again and again, in a manner that Thomas Kuhn tended to dismiss when writing about "normal science." They were pursuing theretofore unnoted details of our planetary system that make a difference in the orbital motions and the subtle, but nonetheless detectable, differences those details make.

My second surprise concerned what was being tested. Take for example the anomaly in the motion of Uranus that led to the discovery of Neptune. It would have been masked if the significantly larger gravitational effects of Saturn on Uranus had not been included in the calculation first. So, the discovery of Neptune provided evidence not only for Newton's theory, but also for the specific aspects of Saturn that entered into calculating its effects on Uranus, for these were no less presupposed in the anomaly that emerged than Newton's theory was. The point generalizes. Each time a discrepancy emerges and a robust physical source for it is found, that source is incorporated into the new calculations, and the process is repeated, typically with still smaller discrepancies emerging that were often theretofore masked in the calculations. So, what

was being tested each time when a new discrepancy emerged and a physical source for it was being sought was not only Newtonian theory, but also all the previously identified details that make a difference and the differences they were said to make without which the further systematic discrepancy would not have emerged.

Notice that as the loop shown in Figure 1 is repeated with more and more details incorporated into the calculations, and smaller and smaller discrepancies emerging, tighter and tighter constraints are being placed on finding robust physical sources for new discrepancies. They had become tight enough in the case of Mercury's perihelion that no source for the residual anomaly in it could be found that was compatible with Newtonian theory.⁹ Notice further what it is about the theory of gravity that was being tested in this process. It was not merely how well the theory represents observations—the sort of question one asks of curve-fits. It was whether the theory can serve to establish which physical features make detectable differences in orbital motions and what differences they make. As the discrepancies became smaller, that became an increasingly stringent test of Newtonian theory.

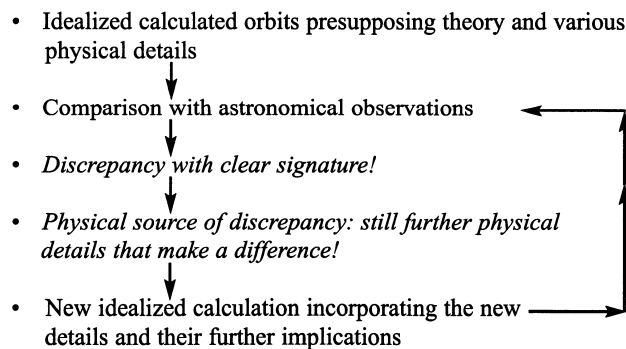


Figure 1. "Closing the Loop": the Logic of Testing in Orbital Mechanics

This shows that increasingly strong evidence was accruing to Newtonian theory over the first two hundred years of orbital research based on it. My third surprise was just how strong that evidence was. To illustrate this I am going to use an example that was of paramount

importance in the history of evidence in orbital research, but has been entirely missed by philosophers of science. As most any textbook on orbital astronomy tells us, the first fully successful account of lunar motion was the Hill-Brown theory of 1919. That theory contains some 1400 perturbational terms, the vast majority of them derived from Newtonian gravity. What textbooks rarely tell us is that it also included a fudge, what Brown called “the Great Empirical Term,” shown in Figure 2 as the solid line. That there was some recalcitrant discrepancy in the motion of the Moon Newcomb had picked up in 1870, but it took the 1400 terms of Hill-Brown theory for it to emerge with the clear signature displayed in the figure. The question it posed was whether some still further force was acting on the Moon or whether instead the rotation of the Earth was fluctuating slightly over a roughly 200 year period.

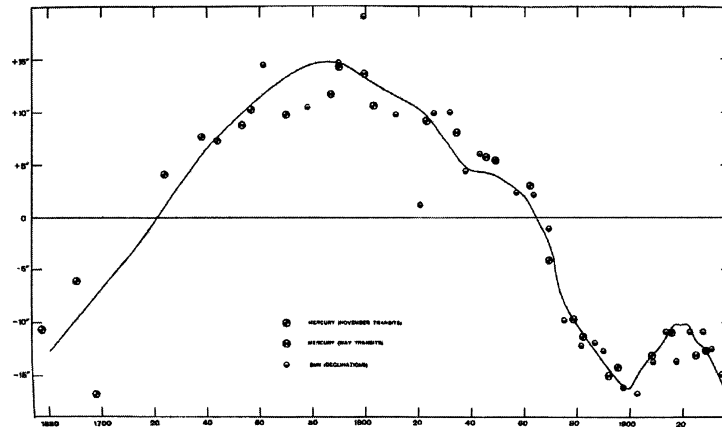


Figure 2. Residual Discrepancies in the Longitudes of The Moon, Mercury, and the Sun in 1939

Harold Spencer Jones resolved this question in 1939 when he established that a parallel discrepancy was showing up in the motions of Mercury, Venus, and the Earth once their mean-motions were renormalized to allow comparison with the Moon's.¹⁰ Same-effect-same-cause reasoning implied a fluctuation in the rotation of the Earth. That has turned out to be

robust, as demonstrated by lunar-laser-ranging following space-shots to the Moon and the subsequent discovery of lesser fluctuations in the Earth's rotation that the fluctuation shown in the figure had masked.¹¹ This detail that makes a difference led in 1950 to the end of the two thousand year tradition of taking sidereal time as our ultimate standard of time.

This example brings out more than just how stringent continued testing of Newtonian theory became over the course of the history of orbital research. It also underscores that not all the discoveries in that research concern gravity. The robust fluctuation in the Earth's rotation corroborated 1400 some terms concerning details in our planetary system that make a detectable difference in the Moon's motion. Most of those terms involve gravitational forces, but not all of them—most notably, terms involving viscous forces in our oceans that are slowly transferring angular momentum from the Earth to the Moon.

There is a still more important respect in which the strength of the evidence coming out of this example surprised me. One might naturally think the evidence issuing from any comparison of calculation with observation is greatest when they match one another within observational accuracy. This example shows that that is not automatically true. Least-squares statistical methods are used in orbital mechanics to set the values of such parameters as Keplerian orbital elements and masses. As a consequence, when orbital calculations agree with observation, some ambiguity always remains about the extent to which the agreement is arising from mere curve-fitting versus the extent to which it confirms gravitation theory. Just such ambiguity would have been present had Hill-Brown theory achieved agreement with observation from the outset and Brown had not needed his Great Empirical term. That discrepancy, however, together with Spencer Jones's success in identifying what has turned out to be a robust physical source for it, removed the ambiguity in that case. For, the evidence that the discrepancy itself is physical became evidence that all the elements entering into the calculation from which it emerged are physical too. *The evidence for physicality was thus stronger with the discrepancy than it would have been without it!* That was my fourth surprise.

My fifth and sixth surprises came from the residual anomaly in the precession of Mercury's perihelion and what the transition to Einstein's theory of gravity has told us about claims to knowledge coming out of two

centuries of research based on Newtonian gravity. As is well known, Einstein required Newtonian gravitation to hold in an asymptotic limit as he developed his new theory of gravity—specifically in a static, weak-field limit. That he did so was just as well because the 43 arc-seconds per century anomaly in the perihelion of Mercury that was initially the sole evidence for his theory presupposes Newtonian gravity. The 43 arc-seconds is not something anyone can observe. It is what one gets after subtracting the 531 arc-seconds per century precession produced by Newtonian gravitational effects of the other planets on Mercury from the observed 574 arc-seconds per century—that is, the 574 arc-seconds that remain once one introduces a roughly 5600 arc-seconds correction of the raw observations to compensate for the wobble of the Earth.¹² So, the 43 arc-seconds could count as evidence for Einstein's new theory only if that theory left intact, at least to the requisite level of precision, all the Newtonian contributions to the 531 arc-seconds that yielded the 43 arc-seconds.

Einstein's immediate goal in requiring that Newtonian theory hold in the static, weak-field limit was most likely to assure that all the evidence that had accumulated in support of that theory carry over immediately as evidence for his new theory. The stringency with which Newtonian theory had been tested gave him good reason to require continuity of evidence in the transition to the new theory.¹³ None of this, however, seemed especially surprising. What did surprise me was what Einstein's asymptotic limit did for the hundreds of details that Newtonian theory had singled out as making detectable differences in orbital motions in our planetary system. The conclusion just reached about the details contributing to the 531 arc-seconds per century generalizes. As a matter of historical fact, all of the details singled out as making detectable differences during the two centuries of prior research carried over intact into post-Einstein orbital mechanics. *Save for some qualifications concerning levels of precision, the same details are still making the same differences as before.* The only substantive change is that new details that make a difference have subsequently been added to them, details that do not invariably make the differences within Newtonian theory that Einstein's theory says they make.

A natural conclusion to draw from this is that the difference-making details identified during the two hundred years of orbital research predicated on Newtonian theory had, and still have, more of a claim to knowledge—to being permanent—than the theory itself ever had. But that

cannot be correct. In an experiment one can confirm that the change in the value of some variable makes a difference simply by changing that value and seeing what happens. But orbital mechanics in those two centuries included no experiments. All conclusions about what changes in variables make what differences came out of Newtonian theory. We are still getting such conclusions out of Newtonian theory, the most prominent being the existence of dark matter. So, Newtonian theory must still have some sort of claim to being knowledge. Of course, it cannot have any claim to being the final word. But all along and still now it has had a claim to holding, projectably, *to sufficiently high accuracy over the domain of our solar system to establish, with seeming finality, a huge number of detailed features that make detectable differences in orbital motions in that system.* The causal aspect of this claim makes it very different from saying merely that Newtonian theory yielded a representation of those motions to a certain level of accuracy—something a curve-fit could have done. That was my sixth surprise, the specific strong respect in which Newtonian theory continues to have claim to permanence.¹⁴

One final surprise. I have all along been speaking of the stringency with which Newtonian theory was tested during the course of orbital research. Now I have to add a surprising qualification. A fundamental part of Newtonian theory is that the strength of the gravitational attraction toward any body varies as the mass of that body. The first two centuries of Newtonian orbital research never tested that claim in any way at all. Throughout its history the masses of celestial bodies have been inferred from the strengths of the centripetal acceleration fields surrounding them. But then the only way in which those masses entered into any of the orbital calculations was as the product GM , the strengths of the fields. In other words, those masses were all along mere placeholders for the field strengths from which they were inferred and which they were then used to calculate. The thesis that gravitational attraction varies as the mass of attracting bodies thus got a free ride over the history of orbital research because it was *not* entering constitutively into that research.¹⁵

Lest anyone jump to a conclusion that those engaged in Newtonian gravitational research were unaware that claims about the masses of attracting celestial bodies were not being tested during that research, I best add something about how the claim that the gravitational field around any body is proportional to its mass was tested.¹⁶ The second volume of Laplace's

Mécanique Céleste (1799) was largely devoted to confirming that claim not from orbits, but from what we now call physical geodesy, in particular the strict relationship between the nonspherical flattening of the Earth and the variation of surface gravity with latitude implied by Clairaut's theorem for an Earth in hydrostatic equilibrium under Newtonian universal gravity.¹⁷ Uncertainties in the geodetic measurements of the flattening and worries about the effects of surface irregularities on surface gravity limited Laplace to concluding that the data were generally supportive of Newton's universal gravity, "even though it is not so strictly verified in this case, as in the motion of the planets."¹⁸ Discrepancies between the measured values of the flattening and the variation of surface gravity with latitude remained throughout the nineteenth century, largely because of nonuniformities in the latter from one meridian to another, thereby continuing to limit the conclusiveness of the evidence from physical geodesy for the claim that the gravitational field around a body is proportional to its mass.

If Cavendish had varied the masses and distances between them in his famous experiment of 1798 in which he was, in effect, determining the value of the gravitational constant G , he would have provided direct evidence for the claim along lines that Newton himself had originally proposed.¹⁹ But Cavendish did not vary the masses; he instead simply assumed Newton's law in order to infer the mean density of the Earth from his experiments, employing the same masses in all his trials. Over the course of the second half of the nineteenth century, however, Cavendish-like experiments were carried out by a number of individuals in order to determine the value of G , and the attracting masses (as well as their shapes) did vary substantially from most of those experiments to the others. Though this was not their recognized objective, for they were taking Newton's law of gravity for granted just as Cavendish had, those individuals, through the convergence of their measured values of G across their several experiments, produced the strongest evidence as of the beginning of the twentieth century that gravity is proportional to the mass of the attracting body. Strikingly, a physicist, Jonathan Zenneck, reached exactly that conclusion in a review article of 1901.²⁰

There is a crucial lesson to learn from the fact that claims about the masses of celestial attracting bodies got a free ride during the two centuries of Newtonian research on orbital motions. The only way to tell just what aspects of a theory were being tested during ongoing research that presupposed

it, and how stringent those tests were, is to examine the history of that research in detail, analyzing the logic of just what was and what was not being tested at each stage. This is just what Laplace did in turning from orbital mechanics to physical geodesy with Volume 2 of his *Mécanique Céleste* and Zenneck did a century later in preparing his review article. The importance of examining the history of evidence in gravity research was already clear, I hope, from the discussions of the first six of my surprises. Regardless, my seventh surprise really drives the point home that, when I say the history of science is indispensable, I am talking about fine details of that history.

3. Some Further Examples Requiring Study of the History of Evidence

A decade of effort on the three hundred-year history of evidence in orbital mechanics is what convinced me of the indispensability of studying the history of evidence in highly specific areas of research in order to answer questions about knowledge achieved in those areas. I do not see why a piece of my autobiography, however, should be all that compelling to anyone else, especially to anyone who has to be taking my word for what they are going to find in that history. So, let me turn to three examples of philosophic issues raised in work done by others for which I do not presently know what a detailed investigation of the history of evidence is going to reveal. Even without knowing what the outcome of such an investigation will be, I am convinced that the philosophic issues raised in each case can be resolved only by means of it. For, in each case prior results have been revisited during ongoing research, and the resolution of the issues turns entirely on how those results came to be tested further, *en passant* or otherwise, as they were revisited.

My *first* example derives from Shapin's and Schaffer's *Leviathan and the Air-Pump*, which I am told has sold more copies over the last 50 years than any other book in the history of science except Kuhn's *The Structure of Scientific Revolutions*. The historiographical purpose of their book, they say late in it, was "to show the inadequacy of the method which regards experimentally produced matters of fact as self-evident and self-explanatory."²¹ The three matters of fact they employ to that end are the Torricellian vacuum, the effectiveness of the air-pump in producing a vacuum, and Boyle's spring of the air. All three are presented as outcomes of experiments in one of the Harvard Case Studies,²² where the experiments

are presented not only as straightforwardly establishing their scientific results in the face of “non-scientific” resistance to them, but also as illustrating to students of humanities the nature of experimental knowledge. In contrast to most science textbooks, that Harvard case study was careful to include the disputes at the time over the vacuum, the difficulties with the newly invented air-pump, and the vagaries in the data from which Boyle extracted his law. Shapin and Schaffer did not dismiss it as sloppy history.

Their complaint was that the case study presented the experimental results as if they settled matters when the step from them to the Royal Society’s endorsing the vacuum, the air-pump, and Boyle’s so-called law was anything but straightforward. However much I might wish in other contexts to dispute their analysis of what was involved in settling those matters during the seventeenth century, that is not my concern here. I am perfectly willing to grant that the experimental results in question fell short of establishing the vacuum at the closed end of Torricellian tubes, the effectiveness of the air-pump, or the relationship between the pressure and density of air expressed by the Boyle-Mariotte law. My complaint with Shapin and Schaffer is, why stop the discussion with 1680? Three and a half centuries of research have subsequently been carried out on those three topics. Are we supposed to take it as “self-evident and self-explanatory” that any lacunae in their initial acceptance have never been revisited?

Consider the air-pump, which has continued to be developed over those three centuries, not merely as a laboratory instrument, but for a wide range of practical applications as well. Until flat-screen displays became the vogue, we all sat for many hours a day in front of cathode-ray tubes that had been evacuated to a measured pressure of around a millionth of an atmosphere. Many experiments are presently conducted in so-called ultra-vacuums of a billionth of an atmosphere, though it is worth noting that science now tells us that the number of molecules remaining in a single cubic centimeter at room temperature in an ultra-vacuum far exceeds a billion. The word ‘vacuum’ has come to have rather complex reference compared to what it had in the 1660s.

The Boyle-Mariotte law was revisited as early as Newton’s *Principia* of 1687, but far more intensively during the research of Gay-Lussac and others early in the nineteenth century, leading to its expansion into what, following Regnault’s research of the 1840s, came to be known as the ideal gas law. Figure 3 shows a table of Regnault’s results that he stated “prove

that the law assumed correct by physicists, viz. that air expands the same fraction of its volume at 0° , whatever its density, is not accurate.”²³ A philosophically intriguing aspect of Regnault’s work is that the instrument he used to measure temperature was a constant-volume air-thermometer, which presupposes the ideal gas law—this in experiments in which he was measuring systematic deviations from that law. During the twentieth century the virial expansion emerged as the rigorous replacement for the ideal gas law. When I, working as an engineer, need really accurate values for the compressibility of air, I turn to tables calculated, to four significant figures, from the experimentally determined first three terms of the virial expansion.²⁴

Pressure at 0°	Pressure at 100°	Density of air at 0° , that of air at 0° under a pressure of 760 mm. being taken as unity.	$1 + 100a$
109.72	149.31	0.1444	1.36482
174.39	237.17	0.2294	1.36513
266.06	395.07	0.3501	1.36542
374.67	510.35	0.4930	1.36587
375.23	510.97	0.4937	1.36572
760.00		1.0000	1.36650
1678.40	2286.09	2.2084	1.36760
1692.53	2306.23	2.2270	1.36800
2144.18	2924.04	2.8213	1.36894
3655.56	4992.09	4.8100	1.37091

The third column of the table comprises the densities of the gas at the temperature of melting ice; it will be seen that these vary from 0.1444 up to 4.8100, that is, from 1 to 33.3, and for so great a change in density, the coefficient of expansion of the gas changes only from 0.3648 to 0.3709.

The above determinations, therefore, prove that the law assumed correct by physicists, viz., that air expands (on heating) the same fraction of its volume at 0° , whatever its density, is not accurate; air expands, between the same temperature-limits, by amounts which are greater in proportion as the density of the gas is higher, or, in other words, as the molecules are closer together.

Figure 3. Regnault’s “Proof” that the Ideal Gas Law is not Exact

These examples scarcely begin to cover the richness of the three hundred fifty years of research in pneumatics since Torricelli proposed his vacuum and Boyle conducted his experiments on the spring of the air. If we are going to reach conclusions about the nature and scope of the knowledge achieved in pneumatic research, we cannot restrict our attention to its first twenty five years. We need to adopt a long vista to see how claims that had become accepted by 1680 have been revisited during the subsequent 330 years.

Understand, however, that just pointing to those 330 years of further research does not answer the complaints of Shapin and Schaffer. To do that we are going to have to subject the various stages of continuing research over those centuries to the same critical scrutiny to which they subjected the first decades. I know a fair amount about the conceptual history of the theory of gases, but I have little idea of what surprises anyone is going to find if they look with that kind of scrutiny at the three and a half centuries of experimental research in pneumatics. Perhaps some fundamental tenets of pneumatics have gotten the same sort of free-ride in those experiments that Newton's claim that attraction varies as the mass of the attracting body got over two hundred years of orbital research. Philosophers of science need to appreciate that surprise is the ultimate redeeming feature of the labor that goes into historical research!

Early in their book Shapin and Schaffer remark:

The Harvard history has itself acquired a canonical status: through its justified place in the teaching of history of science it has provided a concrete exemplar of how to do research in the discipline, what sorts of historical questions are pertinent to ask, what kinds of historical materials are relevant to the inquiry, what sorts are not germane, and what the general form of historical narrative and explanation ought to be. *Yet it is now time to move on* from the methods, assumptions, and the historical programme embedded in the Harvard case history and other studies like it. . . . We became increasingly convinced that the questions we wished to have answered had not been systematically posed by previous writers.²⁵

I do not want to deny the indispensability of local studies of "making knowledge" with their characteristically short vistas for addressing questions of the sort Shapin and Schaffer raised, and questions about initial acceptance generally. When, however, the questions concern the nature and scope of knowledge achieved in specific areas of modern

science, with pneumatics as one among many examples, short vistas cannot suffice. This gives reason to say anew that it is now time to move on beyond such studies.

My *second* example comes out of Hasok Chang's more recent book on thermometry, *Inventing Temperature*.²⁶ Chang, in keeping with Shapin and Schaffer, emphasizes the experimental—and social—complications of the century of effort that went into devising the two fixed points and unit interval between them that formed our temperature scale. He, however, did not stop with 1780.²⁷ His book continues through most of the next century, examining how the scale initially specified in terms of the expansion of mercury was revisited, giving way first to gas thermometers and then to the Kelvin-Joule theoretical characterization of absolute temperature in terms of its relation to energy and entropy, and as a by-product of this a temperature scale tied, as it continues to be, to the ideal gas law.

Chang's book thus ends up striking the same theme I am pressing here, that whatever is epistemically distinctive about science lies more in the process through which accepted results are revisited and refined than in their initial acceptance. If Chang had stopped with 1780, in the manner in which Shapin and Schaffer stopped with 1680, readers would have been left with the sense that temperature as we know it is an artificially devised quantity pieced together through a complex social process. That would have left them wondering whether temperature as we know it could have so little claim to being a real physical quantity. By going on, Chang indicated how it came to have a century later much more of such a claim.

Let me explain what I mean by that. The table displayed in Figure 4, taken from Chang's book,²⁸ compares the Kelvin-Joule ideal-gas absolute temperature scale with the constant-volume air-thermometer scale. I am less interested in the pattern of correction between the two than in the fact that the two fixed-points—refined conceptions of the freezing and boiling points of water²⁹—were carried over into the ideal-gas-based absolute temperature scale. One might think this a mere matter of convention were it not for the fact that the absolute scale includes a third salient point besides those two, absolute zero. As such, it provides a test of the physicality, in contrast to artificiality, of the two fixed-points. If the absolute zero extrapolated from the two fixed-points by Thomson and Joule had turned out not to be physically characterizable, there would

have been reason to reconsider and perhaps reconstitute the fixed-points as well as the unit between them that specifies the quantity of temperature employed in physics.³⁰ As we noted, that is precisely what happened in the middle of the twentieth century when the sidereal day was abandoned as the fundamental unit of time after Spencer Jones provided clear evidence that it, as such a unit, is parochial in a systematically misleading way. The history of thermometry thus offers an answer to a philosophically interesting question: what claim does temperature, defined as we know it in terms of its two classic, socially constructed fixed-points, have to being a physically real quantity?

Absolute temperature, minus 273.7	Air-thermometer temperature ^b
0	0
20	20 + 0.098
40	40 + 0.0403
60	60 + 0.0366
80	80 + 0.0223
100	100
120	120 – 0.0284
140	140 – 0.0615
160	160 – 0.0983
180	180 – 0.1382
200	200 – 0.1796
220	220 – 0.2232
240	240 – 0.2663
260	260 – 0.3141
280	280 – 0.3610
300	300 – 0.4085

*Figure 4. Joule-Thomson Comparison of
Absolute and Air-Thermometer Temperatures*

That leads me to a further question that Chang's book does not answer. The history of thermometry did not end with Kelvin, Joule, and Rankine. The discovery of the Seebeck electromotive force, tying an electric potential to temperature, ultimately led to the development of thermocouples as instruments for measuring temperature; and the discovery of the relationship between temperature and phenomena of radiation similarly led to the development of pyrometers. Temperature

occurs linearly in conjunction with the universal gas constant R in the ideal gas law:

$$p = \left(\frac{n}{V}\right) RT$$

In the Stefan-Boltzmann law for radiation energy, by contrast, temperature occurs to the fourth power, in conjunction with Boltzmann's constant k :

$$\overline{E_{Rad}} = \frac{8\pi^5 V}{15 c^3 h^3} k^4 T^4$$

And in Planck's black-body radiation law, the preferred basis for pyrometry, temperature together with Boltzmann's constant occurs in an exponential:

$$E_\nu = \frac{8\pi h \nu^3}{c^3} \times \frac{1}{\exp(h\nu/kT) - 1}$$

I can now pose my question: to what extent has radiation research provided still more stringent tests of our temperature scale and its fixed-points—or, alternatively, of the now preferred³¹ fixed-point for the absolute thermodynamic scale given by the value at which ice, liquid water, and water vapor are in equilibrium? Some scales we invent are clearly artificial. My favorite example is the Richter scale, which derives from a correlation between a specific measurement on a specific type of seismometer and damage done by earthquakes to buildings in California. Useful though it may be for engineering purposes, it has been of little use in earthquake research, where instead the quantity called the *moment magnitude* is now used.³² Orbital research revisited our fundamental scale for time, ultimately replacing it. But the two fixed-points of the temperature scale remain in place—or at least the one of the two singled out at the beginning of this paragraph. Is this because they have withstood increasingly stringent tests?

As Newton indicated³³ and Duhem subsequently emphasized, questions about the physicality versus the conventionality and artificiality of any scientific quantity turn on the range of well-behaved lawlike relationships that quantity bears to other quantities. Few, if any, scientific quantities are more ubiquitous than temperature, entering as it does into relationships governing phenomena far removed from those that served to

determine its original fixed-points and fundamental interval between them. If there are shortcomings in temperature as we have constituted it, they are more likely to show up in relations in which it occurs in the fourth power and exponentially than when it occurs linearly. So, has the last century of radiation research given us added reason to regard the classic two fixed-points of temperature to have picked out a real physical quantity? This depends on whether that research has really tested, *en passant* or otherwise, those points. I don't know the answer to the question, for while I am familiar with what the *Handbook of Physics* says about the matter, I have not gone back through the history of radiation research, critically examining the data and what they showed. And that is my point: it can be answered only through a detailed study of the history of evidence in radiation research.

My *third* example involves measurement of such fundamental quantities as the universal gas constant R and Boltzmann's constant k , which is just R divided by Avogadro's number. One way in which the physicality of the thermodynamic temperature scale might have been revisited was through precise measurements mediated by radiation theory of k and comparison of the value with other values obtained for k or with values obtained for R and Avogadro's number.

Such fundamental constants as the speed of light in a vacuum and the aberration-of-light constant have been central to astronomy for a long time. Full appreciation, nevertheless, of the evidence that can be obtained from precise measurement of fundamental constants seems to have emerged only in the period from 1897 through 1913, starting with J.J. Thomson's measurement of the mass-to-charge ratio m/e of what we now call the electron and continuing through the precise measurements of its charge e , Planck's constant h , Avogadro's number, Faraday's constant from electrolysis, and Rydberg's constant, that is, the wavelength of the principal line in the spectrum of hydrogen.³⁴ Jean Perrin won the Nobel Prize in 1926 for his marshalling those measurements into a "proof," so to speak, of the discontinuous nature of matter—what philosophers call the "reality" of molecules and atoms.

Several philosophers have recognized the importance of the measurement of fundamental constants to philosophic disputes concerning realism, especially in microphysics;³⁵ but neither philosophers nor historians, so far as I know, have given attention to the subsequent history of the measurement of fundamental constants. Perhaps as a consequence of Perrin's Nobel Prize,

a lesser known physicist, Raymond Birge, took it upon himself to publish preferred values for the six constants cited above and some twenty- five others in a 73 page article in 1929. He followed that a dozen years later with a second article listing preferred values, after which colleagues of his, Richard Cohen and Jesse DuMond, issued a series of further articles over the next quarter century, as well as a 1957 book they dedicated to Birge, *Fundamental Constants of Physics*, surveying the entire topic and its history.³⁶

Throughout that period the task of identifying and publishing the preferred values remained the domain of a handful of otherwise less prominent physicists in California. The number of fundamental constants for which they tabulated values grew from fewer than 40 in Birge's initial publication of 1929 to more than 50 in 1965, and of course the values listed gained some in precision.

All this began to change in 1969 when the international physics community decided that the preferred values of the fundamental constants are important enough to warrant a committee, CODATA (the Committee on Data for Science and Technology), which has been responsible for the "official" values since then. Table 1 lists the sequence of articles listing recommended values, starting with Birge's 1929 article. One thing to notice is that the committee is now publishing new sets of preferred values every four or so years, in memoranda of a hundred or so pages. A more important thing to notice is that the number of "fundamental" constants has grown, from the fewer than 40 with which Birge began to now more than 130. (Because of the theoretical relationships among them, the precise number of *separate* constants in each of the publications is open to interpretation; the full table in the 2006 CODATA report has 222 entries, but what I deem to be 137 separately measurable constants.) The overwhelming majority of the new constants, typical of which is the fine-structure constant, pertain in one way or another to microphysics.

TITLE, AUTHOR	JOURNAL	NUMBER OF CONSTANTS
"Probable Values of the General Physical Constants," Raymond T. Birge	<i>The Physical Review Supplement</i> , 1929, pp. 1–73	<40
"The General Physical Constants, as of August 1941, with details on the velocity of light only," Raymond T. Birge	<i>Reports on Progress in Physics</i> , 1941, pp. 90–134	<60

(continued on page 568)

TITLE, AUTHOR	JOURNAL	NUMBER OF CONSTANTS
"Least-Squares Adjustment of the Atomic Constants, 1952," Jesse W.M. DuMond and E. Richard Cohen	<i>Reviews of Modern Physics</i> , 5, 1953, pp. 691–708	<70
"Analysis of Variance of the 1952 Data on the Atomic Constants and a New Adjustment, 1955," E. Richard Cohen, Jesse M.W. DuMond, Thomas Layton, and John S. Rollett	<i>Reviews of Modern Physics</i> , 27, 1955, pp. 363–90	see preceding
"Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965," E. Richard Cohen and Jesse M.W. DuMond	<i>Reviews of Modern Physics</i> , 37, 1965, pp. 537–94	30 primary constants
"The Least-Squares Adjustment of the Fundamental Constants," E. Richard Cohen and Barry N. Taylor, under the auspices of CODATA	<i>Journal of Physical and Chemical Reference Data</i> , 2, 1973 pp. 663–734	>50
"The 1986 Adjustment of the Fundamental Physical Constants," E. Richard Cohen and Barry N. Taylor, representing CODATA	<i>Reviews of Modern Physics</i> , 59, 1987, pp. 1121–48	>100
"CODATA Recommended Values of the Fundamental Constants, 1998," Peter J. Mohr and Barry N. Taylor	<i>Reviews of Modern Physics</i> , 72, 2000, pp. 351–495	>120
"CODATA Recommended Values of the Fundamental Constants, 2002," Peter J. Mohr and Barry N. Taylor	<i>Reviews of Modern Physics</i> , 77, 2005, pp. 1–107	>120
"CODATA Recommended Values of the Fundamental Constants, 2006," Peter J. Mohr, Barry N. Taylor, and David B. Newell	<i>Reviews of Modern Physics</i> , 80, 2008, pp. 633–730	>130

Table 1. History of Publication of Preferred Values of Fundamental Constants

As with the six constants I singled out from the 1897 to 1913 period—all of which are included in the "abbreviated list" from the 2006 CODATA report shown in Figure 5—the majority of the ninety added constants are related to others, within physical theory, often in more than one way. Measurements of almost all of them, needless to say, are theory-mediated, but thanks to their multiple theoretical interrelations, there are generally multiple ways of measuring them, involving different theoretical presuppositions. Much of the committee's task is to decide which way of measuring each of them is currently yielding the most accurate value, and that requires comparisons not only of values determined in different ways, but of the different ways themselves.

The "abbreviated list" indicates how markedly the number of significant figures has increased from the time of Perrin; it is continuing to increase.

In the case of the converging measures of the universal gravitation constant, G , the number of significant figures gives us error bands on the precision to which the gravitational field strength varies with the mass of the attracting body.³⁷ What the increased precision is giving us in the case of the other 130 some constants is the key question. One curiosity worth mentioning is that far more frequently than one might expect, new officially recommended values have fallen outside the published error-bands or uncertainty of their immediate predecessors. Not surprisingly, that has made the committee now chary about publishing error bands.³⁸

Quantity	Symbol	Numerical value	Unit	Relative std. uncert. u_r
speed of light in vacuum	c, c_0	299 792 458	m s^{-1}	(exact)
magnetic constant	μ_0	$4\pi \times 10^{-7}$	N A^{-2}	(exact)
electric constant $1/\mu_0 c^2$	ϵ_0	$= 12.5666 370 614 \dots \times 10^{-7}$	N A^{-2}	(exact)
Newtonian constant of gravitation	G	$8.854 187 817 \dots \times 10^{-12}$	F m^{-1}	(exact)
Planck constant	h	$6.674 28(67) \times 10^{-11}$	$\text{m}^3 \text{kg}^{-1} \text{s}^{-2}$	1.0×10^{-4}
$h/2\pi$	$\hbar$	$6.626 068 96(33) \times 10^{-34}$	J s	5.0×10^{-8}
elementary charge	e	$1.054 571 628(53) \times 10^{-34}$	J s	5.0×10^{-8}
magnetic flux quantum $h/2e$	Φ_0	$1.602 176 487(40) \times 10^{-19}$	C	2.5×10^{-8}
conductance quantum $2e^2/h$	G_0	$2.067 833 667(52) \times 10^{-15}$	Wb	2.5×10^{-8}
electron mass	m_e	$7.748 091 7004(53) \times 10^{-5}$	S	6.8×10^{-10}
proton mass	m_p	$9.109 382 15(45) \times 10^{-31}$	kg	5.0×10^{-8}
proton-electron mass ratio	m_p/m_e	$1.672 621 637(83) \times 10^{-27}$	kg	5.0×10^{-8}
fine-structure constant $e^2/4\pi\epsilon_0\hbar c$	α	$1.836 152 672 47(80)$		4.3×10^{-10}
inverse fine-structure constant	α^{-1}	$7.297 352 5376(50) \times 10^{-3}$		6.8×10^{-10}
Rydberg constant $\alpha^2 m_e c/2h$	R_∞	$137.035 999 679(94)$		6.8×10^{-10}
Avogadro constant	N_A, L	$10 973 731.568 527(73)$	m^{-1}	6.6×10^{-12}
Faraday constant $N_A e$	F	$6.022 141 79(30) \times 10^{23}$	mol^{-1}	5.0×10^{-8}
molar gas constant	R	$96 485.3399(24)$	C mol^{-1}	2.5×10^{-8}
Boltzmann constant R/N_A	k	$8.314 472(15)$	$\text{J mol}^{-1} \text{K}^{-1}$	1.7×10^{-6}
Stefan-Boltzmann constant $(\pi^2/60)k^4/\hbar^3 c^2$	σ	$1.380 6504(24) \times 10^{-23}$	J K^{-1}	1.7×10^{-6}
Non-SI units accepted for use with the SI				
electron volt: $(e/C) \text{ J}$	eV	$1.602 176 487(40) \times 10^{-19}$	J	2.5×10^{-8}
(unified) atomic mass unit $1 \text{ u} = m_u = \frac{1}{12} m(^{12}\text{C})$ $= 10^{-3} \text{kg mol}^{-1}/N_A$	u	$1.660 538 782(83) \times 10^{-27}$	kg	5.0×10^{-8}

Figure 5. “Abbreviated List” of the Recommended Values of Fundamental Constants, 2006

An international committee that decides every few years which eight- or ten-significant-figure measurement is the most accurate is right up the alley of Shapin and Schaffer, not to mention their colleague in Edinburgh, Donald MacKenzie. How does this international institution that persists over time with changing members, most of whom remain virtually unknown to those outside the scientific community, reach their decisions?

Philosophers of science have still better reason to look at the continuing history of the values of the fundamental constants. The evidence for much of what physics claims to know about microphysical structure, from the level of molecules to that of electron-photon interactions, lies within those measurements. How stable is each of the different ways of measuring each constant, and what does that stability tell us about the specific theoretical principles entering that way of measuring it? To what extent do different ways of measuring the same constant, with their different theoretical assumptions, converge on the same value, and what conclusions about those assumptions can we draw from the degree of convergence?

The stringency with which the theoretical assumptions are being tested in the measurement of any one constant depends first on the precision of the measurement and second on the number of cross-checks provided by the relationships between the constant in question and other independently measured constants. Issues about which parts of microphysics have been more stringently tested than others can be answered only by looking carefully at the history of the measurement of each inter-related group of constants. Physicists are not going to worry about a question like this until problems surface in ongoing research. Philosophers need not wait for such problems to surface.

Hopefully, my three examples suffice to show that the importance of the history of evidence to philosophy of science is not limited to gravity research. I could have chosen any number of other examples. I chose the three I did because they are especially pertinent to philosophic issues about the knowledge gained in microphysics. Had I more space, the one other example I would have included is the initial half-century inability to reconcile the theoretical and measured ratios of the specific heats of gases. As Maxwell pointed out in 1860, this conflict between theory and observation gave clear evidence that one of the most fundamental premises of statistical mechanics, the equipartition of energy, is not true. The first proposal that began to get anywhere on the problem was at the initial Solvay Conference of 1911,⁴⁰ and even then it took fifteen more

years until the last discrepancy was eliminated and some quite stunning microstructural conclusions could be drawn from the measured values.⁴¹ What do fifty or more years in the face of clear evidence that one of the most fundamental assumptions of that research is false tell us about modern science?

4. Toward a Rapprochement Between History and Philosophy of Science

I proposed at the outset that, whatever claim any science has to lofty epistemic standing, it is generally going to lie in the extent to which fundamental principles of that science have been tested not only again and again over the historical course of research in it, but more importantly ever more stringently. My summary of the way in which such further testing has happened over the course of the history of gravity research has supported my proposal for that science. The further examples I have offered pose questions that suggest similar support for my proposal may be coming out of them. What we need are studies in the history of evidence in individual, narrow areas of research, demarcated by the extent to which the principal research-concerns in them have remained the same. The fact that any continued testing of fundamental principles has likely been largely *en passant* means that philosophers and historians have some hard work ahead of them, digging critically into little known publications and even reconstructing past experiments. I can attest from my efforts on the history of evidence in gravity research that such hard work can be highly rewarding. I am sure that Hasok Chang would say the same about his efforts on thermometry.

I am asking a good deal of philosophers of science in proposing that they undertake such studies. For one thing, I am asking them to put to one side, at least for the moment, many of the issues that have dominated the field over the last century—confirmation theory, issues of realism versus, say, constructive empiricism, and abstract concerns about theory-change and the semantics of theories. If I am right, we are still woefully unprepared to address such issues, for we do not really know what the evidence consists of in most areas of research. That is what I mean when I say that so much of philosophy of science is empty, for it is about a stick-figure, and at times even a fantasy, version of science, not the real thing. I am asking philosophers of science to start reading primary source material not simply of the greats, like Newton and Einstein, but of such less heralded figures as Simon Newcomb and Harold Spencer Jones, in my case, and

Jean-André De Luc and Victor Regnault, in Chang's. The hard work is not finding and reading such sources, but learning to read them in their historical context, as others at the time and immediately afterwards read them.

Philosophers do bring one special training to this task, their skill in exposing and critically evaluating hidden assumptions in evidential reasoning and thereby making explicit what the logic of that reasoning was. Scientists of course have that skill too, but they rarely bring it to bear on the past unless something has gone wrong in current research *or* they are charged with writing a review article. What I am asking of philosophers of science in reviewing the history of evidence in any field is to put themselves in a position to write a sequence of review articles from the perspective of key moments in that history, critically assessing what the evidence had shown to that moment, its limitations, and the relationship between these and the chief questions with which research at that time was grappling. The retrospective stance of history will generally provide some insight into shortcomings in the evidence that those at the time would have had trouble formulating. But that is as much a problem as a blessing insofar as care is needed not to become too omniscient in viewing matters from the perspective of the time. When and how insights about shortcomings in the evidence that are clear to us first emerged is part of the history of evidence.

Surprisingly, I am asking much less of historians of science. Admittedly, much of history of science has concerned how conceptual frameworks have developed over time, usually through novel shifts made by the greats; and more recently it has focused on how the social context of scientific research in specific locales has affected it. There are, nonetheless, some well-known studies in the history of evidence, such as the one that got me started on the history of gravity research some twenty-five years ago, Curtis Wilson's "From Kepler's Laws, So-called, to Universal Gravitation: Empirical Factors."⁴² Shapin and Schaffer, for that matter, were themselves engaged in a study of the history of evidence, though granted more with the goal of showing how scientists at the time and philosophers and historians of science more recently have misconceived that evidence. My chief request of historians of science is simply not to rule out as illegitimate, even before the work begins, studies involving such long, narrow vistas as the history of comparisons made by specialists of calculated versus observed motions of the planets, say from Ptolemy to today, exploring exactly why increasingly greater agreement has been achieved.

Fashion has a way of turning into dogma. *Leviathan and the Air-Pump* produced a fashion in the history of science of concentrating on the local making of knowledge, local both geographically and temporally. That fashion has led to doubts about the legitimacy of questions stretching across multiple generations of specialists in many different countries in the history of science. Shapin and Schaffer accused the Harvard historians of being blind to what was involved in establishing matters of fact in seventeenth century experiments in pneumatics. When I say that much recent history of science is blind, what I mean is that it has, like the drunk looking for his car keys where the street is lighted, been looking myopically in the wrong place if the question is whether the sciences have any extraordinary claim to knowledge. One of the wonderful things about history is that one can legitimately ask so many different questions. The chief thing I am requesting of historians of science is not to dogmatically reject certain questions that have recently been out of fashion.

Why do I think that this is important? Early in his book, *Making Natural Knowledge*, is an extraordinarily astute remark by Jan Golinski: "The history of science has had a long struggle to free itself from science's own view of the past."⁴³ This, on my view, has been the signal achievement of the last fifty years of the history of science, a goal Tom Kuhn was aiming for when he published *The Structure of Scientific Revolutions*, initiating a still continuing conflict between historians and philosophers of science.

The fact that a struggle was needed for history of science to free itself indicates that the shortcomings of science's own view of its past are too elaborate to cover here. A key unfortunate aspect of that view especially relevant to this paper can be summarized in terms akin to those used by Shapin and Schaffer: *once one of the greats proposed a breakthrough together with a modicum of reasons to take it seriously, anyone properly trained in science should have seen almost immediately that it had to be true—almost as if the proposal, once put forward, were self-evident*. Subsequent efforts to test it therefore amounted to little more than confirming the obvious in order to safeguard against any slight possibility of a mistake. That attitude, expressed succinctly in Kuhn's famous remark, "Mopping-up operations are what engage most scientists throughout their careers,"⁴⁴ grossly undervalues the contribution the vast majority of scientists over the centuries have made to the development of knowledge in their field.

The mistake lying behind this attitude is one of conflating a distinction that Newton went to pains to make in his “Rules of Reasoning” in the *Principia*: it is one thing to have sufficient evidence to *take* some claim to be true for purposes of ongoing research, and quite another for it to be true.⁴⁵ The mistake is understandable, for once scientists take something to be true and cease questioning it in their ongoing practice, they lose sight of the distinction between that and its simply being true—that is, they do so until some unexpected result gives reason to reconsider it. One nevertheless has to wonder how so many scientists can put so little value on their own work. In highlighting that value I am not denying the genius of the insights of the greats, even though I do think that scientists underappreciate the extent to which prior developments in a science create the opportunity for such moments of genius. All I am denying is that those groundbreaking insights are what give modern science its claim to special epistemic standing. That claim, when it is valid, lies instead in how stringently the groundbreaking proposals came to be tested over subsequent years, decades, and even centuries. In glossing over the subsequent history, science’s view of its past has done at least as much to create the “science wars” as historians and sociologists of science like Kuhn, Shapin, and Schaffer. Indeed, we should applaud them for enabling us all to free ourselves from that view of the past.

Historians of science have freed us from it by showing time and again how claims that scientists and philosophers of science have made about how science is different from other areas of inquiry do not withstand critical scrutiny. *But it is now time to move on.* Kuhn’s master project was not to undermine all claim that the sciences have to special epistemic standing, but to lay out a proper understanding of the respects in which the sciences represent humanity’s greatest epistemic achievement. Kuhn, on my view, deflected that project down a dead-end path not merely by being so dismissive of normal science, but also by overemphasizing aspects in which the sciences do not differ from other areas of inquiry. It is time we return, free of the classic misconceptions, to questions about how the sciences really *are* different from other intellectual pursuits. It is because these questions will require a concerted effort from both historians and philosophers of science that I urge a rapprochement between the two. We might not merely answer questions we want answered about the nature and limits of knowledge produced in the sciences. We might provide an

understanding of science that will be of value to scientists and the community at large.

George E. Smith

Tufts University

NOTES

1. My distinction between constitutive and heuristic elements somewhat parallels the one Pierre Duhem draws between the discardable “explanatory” elements of physical theory and the “representational” elements. See *The Aim and Structure of Physical Theory*, tr. Philip P. Wiener (Princeton: Princeton University Press, 1991) Part I, especially Chapter 3, pp. 31–38.

2. George E. Smith, “J.J. Thomson and the Electron: 1897–1899,” in *Histories of the Electron: The Birth of Microphysics*, ed. Jed Z. Buchwald and Andrew Warwick (Cambridge: MIT Press, 2001), 21–76; and “Getting Started: Building Theories from Working Hypotheses,” the second lecture in the series, “Turning Data Into Evidence: Three Lectures on the Role of Theory in Science,” given at Stanford University in 2007 and available online at <http://www.stanford.edu/dept/cisst/events0506.html>.

3. Ironically, J.J. Thomson’s son, George Thomson, shared the Nobel Prize in Physics in 1937 for his experimental efforts establishing the diffraction of electrons.

4. Mary F. Somerville, *On the Connection of the Physical Sciences* (London: Elibron Classics, 2005; facsimile of the 1840 edition), 27.

5. For more, see my lecture “Getting Started” cited in note 2.

6. This section summarizes the conclusions from two long essays of mine, “Closing the Loop: Testing Newtonian Gravity, Then and Now,” forthcoming, and “Pending Tests to the Contrary: The Question of Mass in Newton’s Law of Gravity,” in preparation. A shorter lecture version of the former is available online at:

<http://www.stanford.edu/dept/cisst/events0506.html>.

I do not have space to repeat the arguments for my conclusions in this paper. Those who find what I say here excessively *ex cathedra* should turn to those essays.

7. Norwood Russell Hanson, “Leverrier: the Zenith and Nadir of Newtonian Mechanics,” *Isis*, vol. 53, 1962, 359–78.

8. Carl G. Hempel, “Provisos: A Problem Concerning the Inferential Function of Scientific Theories,” in *The Limitations of Deductivism*, ed. Adolf Grünbaum and Wesley G. Salmon (Berkeley: University of California Press, 1988), 19–36.

9. In particular, a proposal Simon Newcomb had made in 1895 (“The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy,” *Supplement to the American Ephemeris for 1897*, Washington, 1895) of changing the value of the exponent in Newton’s inverse-square law from -2 to -2.0000001612 had become untenable once Ernst Brown had shown in 1903 (E. W. Brown, “On the Verification of the Newtonian Law,” *Monthly Notices of the Royal Astronomical Society*, Vol. 64, 1903, 396–97) that such a change was much too large to reconcile with the motion of the perigee of the Moon.

10. H. Spencer Jones, “The Rotation of the Earth, and the Secular Acceleration of the Sun, Moon, and Planets,” *Monthly Notices of the Royal Astronomical Society*, vol. 99, 1939, 541–58. Figure 2 of the present paper is Figure 1, p. 551, of that paper. The dots on

the figure display the comparable discrepancies for Mercury and the Sun, once Spencer Jones renormalized their mean-motions to be on the same basis with the mean-motion of the Moon, a renormalization that is needed in order to compare commensurately the contributions made by a fluctuation in the rotation of the Earth to their respective geocentric longitude discrepancies.

11. See Kurt Lambeck, *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge: Cambridge University Press, 1980).

12. To appreciate the complexity of extracting the residual 43 arc-seconds per century in the motion of the perihelion of Mercury, see G.M. Clemence, "The Relativity Effect in Planetary Motions," *Reviews of Modern Physics*, vol. 19, 1947, 361–64.

13. See J.Z. Buchwald and G.E. Smith, "Incommensurability and Discontinuity of Evidence," *Perspectives on Science*, Vol. 9, no. 4, 2002, 463–98.

14. See "The Revolution That Didn't Happen," in Steven Weinberg, *Facing Up: Science and Its Cultural Adversaries* (Cambridge: Harvard University Press, 2001), 187–206, and "T.S. Kuhn's Non-Revolution: An Exchange," *ibid.*, 207–209. I propose that the remarks made in the paragraph that ends with this note, properly expanded as in my "Closing the Loop," spell out grounds for the claims Weinberg makes in the latter of these two.

15. The one factor seemingly involving the masses of celestial bodies entering into orbital calculations was the two-body correction to Kepler's $3/2$ power rule, used to infer the mean distance of planets from their periods. Newton himself presents the correction in terms of the masses, but in doing so he assumed that the two bodies are orbiting about their common center of gravity. The two-body correction can be derived without the assumption that the neutral point the two bodies are orbiting is their common center of gravity; so derived, it can be stated directly in terms of the respective strengths of the centripetal acceleration fields without any occasion to mention masses. Moreover, observations were not precise enough to confirm the need for the two-body correction, even for Jupiter, until well into the twentieth century. For details, see my "Pending Tests to the Contrary: The Question of Mass in Newton's Law of Gravity."

16. My "Pending Tests to the Contrary: The Question of Mass in Newton's Law of Gravity" discusses the history of evidence for this claim in detail. The remarks here merely summarize the conclusions of that essay.

17. Clairaut's equation for a body in hydrostatic equilibrium held together by Newton's inverse-square universal—that is, particle-to-particle—gravity states that there is a fixed relation, given the ratio of the centrifugal force to pure gravity at the equator m , between the variation of the measured acceleration of gravity with latitude, g , and the flattening of the Earth f :

$$\frac{g}{g_{\text{equator}}} = 1 + \left(\frac{5}{2}m - f\right) \sin^2 \phi$$

This equation, which is here given only to first-order in m , can be rewritten to replace g/g on the left with the ratio between the lengths of a one-second pendulum at latitude and at the equator.

18. [Pierre Simon] Marquis de Laplace, *Celestial Mechanics*, tr. Nathaniel Bowditch, Vol. II (Bronx: Chelsea Publishing, 1966), 931. The strict verification that Laplace refers to in the case of the motion of the planets is of the inverse-square centripetal acceleration field around each body.

19. Henry Cavendish, "Experiments to determine the Density of the Earth," *Philosophical Transactions of the Royal Society*, vol. 88, 1798, 469–528. Although it has gone largely unnoticed, in Book 1, Proposition 92 of his *Principia*, Newton proposed a sequence of *experimenta crucis* first with spheres and then with bodies of different shapes to confirm his inverse-square particle-to-particle gravity.

20. Jonathan Zenneck, "Gravitation," in the *Encyclopädie der mathematischen Wissenschaften*, Vol. 5: Physics, ed. A. Sommerfeld (Leipzig: B.G. Teubner, 1903–1921, 25–67; reprinted (in English) in *The Genesis of General Relativity*, ed. Jürgen Renn, Vol. 3: *Gravitation in the Twilight of Classical Physics*, ed. Jürgen Renn and Matthias Schemmel (Dordrecht: Springer, 2007), 77–112. In particular, Zenneck concludes (p. 88):

The G-determinations of Poynting and of Richarz and Kriger-Menzel are of special value in relation to the question of how far the proportionality of the attractive force to the mass is guaranteed for masses of the same material. Both experimenters used unobjectionable laboratory methods carried out with the greatest care. Both determinations employed the same material (lead) and the same method of measurement, but masses of very different magnitudes (154 or 100,000 kg). Even though in one case the mass was 650 times greater than the other, the results agree to approximately 0.2%.

21. Steven Shapin and Simon Schaffer, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton: Princeton University Press, 1985), 225.

22. James Bryant Conant (ed.), *Harvard Case Studies in Experimental Science*, Vol. 1, Case 1: "Robert Boyle's Experiments in Pneumatics," ed. James Bryant Conant (Cambridge: Harvard University Press, 1957), 1–63.

23. Victor Regnault, "Researches upon the Expansion of Gases," Second Memoir, in *The Expansion of Gases by Heat: Memoirs by Dalton, Gay-Lussac, Regnault, and Chappius*, tr. and ed. Wyatt W. Randall (New York: American Book Company, 1902), 135.

24. V.V. Sychev, A.A. Vasserman, A.D. Kozlov, G.A. Spiridonov, and V.A. Tsymarny, *Thermodynamic Properties of Air*, National Standard Reference Data Service of the USSR, English-Language ed. Theodore B. Selover, Jr. (Washington: Hemisphere Publishing Corporation, 1987).

25. Shapin and Schaffer, *op. cit.*, 4, italics added.

26. Hasok Chang, *Inventing Temperature: Measurement and Scientific Progress* (Oxford: Oxford University Press, 2004).

27. My choice of the year 1780 is slightly arbitrary. "The Report of the Committee Appointed by the Royal Society to Consider the Best Method of Adjusting the Fixed Points of Thermometers; and of the Precautions Necessary to Be Used in Making Experiments with Those Instruments" was published in 1777, and Jean-André de Luc's *An Essay on Pyrometry and Areometry, and on Physical Measures in General*, published in 1779, first appeared in *Philosophical Transactions of the Royal Society* in 1778.

28. Chang, *op. cit.*, 196. The table originally appeared in James Prescott Joule and William Thomson, "On the Thermal Effects of Fluids in Motion, Part 2," *Philosophical Transactions of the Royal Society*, vol. 144, 321–64, the second of four articles with that title, the others appearing in *Phil. Trans.* in 1853, 1860, and 1862.

29. See Chang, *op. cit.*, Chapter 1 or the papers cited in note 27.

30. Such a revision, though not one undercutting my main point with this example, was made in 1954 when "the Tenth General Conference on Weights and Measures defined the thermodynamic temperature scale by selecting the triple point of water as the fundamental fixed point and assigning to it the temperature of exactly 273.16°K. This new absolute thermodynamic scale, called the Kelvin scale, does not differ from the thermodynamic

'centigrade' scale by an amount which can be determined experimentally at this time." (R.E. Wilson and R.D. Arnold, "Thermometry and Pyrometry," in E.U. Condon and Hugh Odishaw, *Handbook of Physics* (New York, McGraw-Hill, 1958), 5–30.) Other refinements of the temperature scale are discussed in that chapter.

31. See the preceding note.

32. Keiiti Aki and Paul G. Richards, *Quantitative Seismology*, 2nd edition (Sausalito: University Science Books, 2002), 48 and 655.

33. I specifically have in mind Newton's remarks about the measurement of time in his Scholium on space and time; see *The Principia: Mathematical Principles of Natural Philosophy*, tr. I. Bernard Cohen and Anne Whitman (Berkeley: University of California Press, 1999), 410.

34. See my "J.J. Thomson and the Electron: 1897–1899," cited in note 2 above, especially pp. 60–64; that paper stops short of including the seminal article by H.G.J. Mosely, "The High-Frequency Spectra of the Elements," *Philosophical Magazine*, vol. 27, 1913, 1024–34 and vol. 27, 1914, 703–13, which ties atomic number directly to x-ray spectra and indirectly to the constants I discussed.

35. See, for example, Wesley Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton: Princeton University Press, 1984), 213–29; William Harper and George E. Smith, "Newton's New Way of Inquiry," in *The Creation of Ideas in Physics*, ed. Jarrett Leplin (Dordrecht: Kluwer, 1995), 113–66; William Harper, "Isaac Newton on Empirical Success and Scientific Method," in *The Cosmos of Science: Essays of Exploration* (Pittsburgh: University of Pittsburgh Press, 1997), 55–86; Peter Achinstein, *The Book of Evidence* (Oxford: Oxford University Press, 2001); my lecture, "Getting Started," cited in note 2 above; and, most recently, Bas C. van Fraassen, "The Perils of Perrin, in the Hands of Philosophers," forthcoming.

36. E. Richard Cohen, Kenneth M. Crowe, and Jesse W.M. DuMond, *Fundamental Constants of Physics* (New York, Interscience Publishers, 1957). In 1983 an entire issue of *Philosophical Transactions of the Royal Society* (Series A, vol. 310, no. 1512) was devoted to the fundamental constants and their values, and two years later B.W. Petley published a book, *The Fundamental Constants and the Frontier of Measurement* (Bristol: Adam Hilger, 1985) on them and why their measurement is important.

37. Specifically, the value listed in the 2006 CODATA report implies that the dependence of gravitational force on the mass of the attracting body holds to within 1 part in 10 thousand. It was only 2 parts in a thousand in 1910, while Eötvös by then had shown the dependence of the forces on the mass of the attracted body holds to within 5 parts in 100 million and Brown had shown that the exponent of distance is -2.0 to within 2 parts in 100 million. That explains why the bounds on the precision of G , a constant of proportionality if Newton's law holds exactly, have long since become bounds on the precision with which gravitational forces are proportional to the mass of the attracting body.

38. Kenneth G. Wilson, personal communication. The obvious problem with bounding the error in theory-mediated measurements of fundamental constants is unrecognized sources of systematic error in the measurements, something the spread in the measured data from any one measurement cannot expose. Wilson also reminded me of a central point Donald MacKenzie makes in his *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge: MIT Press, 1990): error-bands are themselves socially constructed. For just this reason we should not expect them invariably to be monotonically decreasing as they are revisited during ongoing research.

39. Maxwell first raised this objection, somewhat crudely, in 1860 in a report of a meeting of the British Association for the Advancement of Science; see Document III, 8

in *Maxwell on Molecules and Gases*, ed. Elizabeth Garber, Stephen G. Brush, and C.W.F. Everitt (Cambridge: MIT Press, 1986), 320f. The objection is elaborated more carefully in his “On the Dynamical Evidence of the Molecular Constitution of Bodies,” of 1875 (*ibid.*, Document II, 24, 217–37), where (p. 229) he calls the ratio of the specific heats “the greatest difficulty which the kinetic theory of gases has yet encountered.”

40. See Diana Kormos Barkan, *Walther Nernst and the Transition to Modern Physical Science* (Cambridge: Cambridge University Press, 1999), especially Chapter 11.

41. See R.H. Fowler, *Statistical Mechanics: The Theory of the Properties of Matter in Equilibrium* (Cambridge: Cambridge University Press, 1929), Chapter III, 50–71; and J.R. Partington and W.G. Schilling, *The Specific Heats of Gases* (New York: D. Van Nostrand, 1924), Chapter 6, especially pp. 229–41.

42. Curtis Wilson, “From Kepler’s Laws, So-called, To Universal Gravitation: Empirical Factors,” *Archive for History of Exact Sciences*, Vol. 6, 1970, 89–170.

43. Jan Golinski, *Making Natural Knowledge: Constructivism and the History of Science* (Cambridge: Cambridge University Press, 1998), 2.

44. Thomas Kuhn, *The Structure of Scientific Revolutions*, second ed., (Chicago: University of Chicago Press, 1970), 24.

45. Newton’s precise phrasing in his third Rule of Reasoning, introduced in the second edition of the *Principia*, is, “Those qualities of bodies that cannot be intended and remitted and that belong to all bodies on which experiments can be made *should be taken* to be qualities of all bodies universally;” and in the fourth Rule, added in the third edition, “In experimental philosophy, propositions gathered from phenomena by induction *should be taken* to be either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exception,” where in both cases the Latin verb translated with my emphasis added is a form of *habere*. Newton’s initial formulation of the third Rule, handwritten in his annotated copy of the first edition, and also in John Locke’s copy, was, “The laws and properties of all bodies on which experiments can be made *are* the laws and properties of bodies universally.” (See I. B. Cohen, *Introduction to Newton’s ‘Principia’* (Cambridge: Harvard University Press, 1971) 25.) So, Newton came to emphasize the distinction between “take to be” and “are” only subsequently, but then took the care much later to phrase the fourth rule correspondingly.