
Precision measurement, scientific personalities and error budgets: the *sine quibus non* for big G determinations

James E. Faller

Phil. Trans. R. Soc. A 2014 **372**, 20140023, published 8 September 2014

References

This article cites 19 articles, 4 of which can be accessed free
<http://rsta.royalsocietypublishing.org/content/372/2026/20140023.full.html#ref-list-1>

Subject collections

Articles on similar topics can be found in the following collections

[mechanics](#) (37 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

Review



Cite this article: Faller JE. 2014 Precision measurement, scientific personalities and error budgets: the *sine quibus non* for big G determinations. *Phil. Trans. R. Soc. A* **372**: 20140023.

<http://dx.doi.org/10.1098/rsta.2014.0023>

One contribution of 13 to a Theo Murphy Meeting Issue ‘The Newtonian constant of gravitation, a constant too difficult to measure?’

Keywords:

Newtonian constant, big G , precision measurement, scientific personalities, errors, systematic errors

Author for correspondence:

James E. Faller

e-mail: fallerj@jilau1.colorado.edu

Precision measurement, scientific personalities and error budgets: the *sine quibus non* for big G determinations

James E. Faller^{1,2}

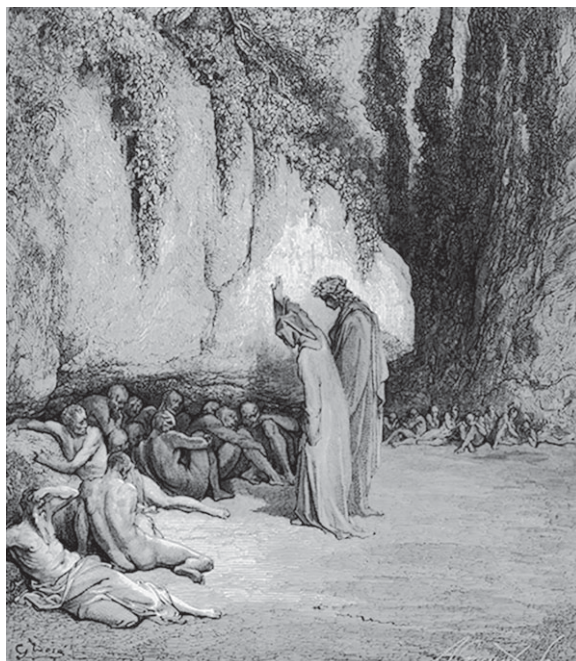
¹JILA, University of Colorado and National Institute of Standards and Technology, Boulder, CO 80309, USA

²Institute for Gravitational Research, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK

Determinations of the Newtonian constant of gravitation (big G) fit into the oftentimes-unappreciated area of physics called precision measurement—an area which includes precision measurements, null experiments and determinations of the fundamental constants. The determination of big G —a measurement which on the surface appears deceptively simple—continues to be one of Nature’s greatest challenges to the skills and cunning of experimental physicists. In spite of the fact that, on the scale of the Universe, big G ’s effects are so large as to single-handedly hold everything together, on the scale of an individual research laboratory, big G ’s effects are so small that they go unnoticed...hidden in a background of much larger forces and noise sources. It is this ‘smallness’ that makes determining the precise value of this (seemingly unrelated to the rest of physics) fundamental constant so difficult.

1. Introduction

An underlying theme in this paper will be the interconnectivity and commonality of all precision measurement experiments. Big G determinations are connected to the rest of physics, if by nothing else, through their associated error budgets that ‘look’ like the precision measurement error budgets from the rest of physics. Because of this connection, other precision measurements can benefit from the 300-year history of determinations of the Newtonian constant of gravitation. Furthermore, as the Mount Everest of



And you're certain that none of you want to learn about science?

Figure 1. Public interest in science. (Engraving by Gustav Doré, in *The vision of Purgatory and Paradise* by Dante Alighieri, London and New York: Cassell, Petter and Galpin 1868.)

precision measurement, big G 's determination promises to be a measurement on which metrologists will want to hone their laboratory skills for generations to come.

In the summary, I mention that precision measurement also includes null experiments and the determination of fundamental constants. Some might ask, 'What is a null experiment?' A null experiment is simply an experiment in which the question, asked of Nature, is framed in such a manner that the expected result will be zero. In this case the 'gain' (sensitivity) of the experiment can then be turned up as high as possible to increase the precision of the null result. On the other hand, in the best (and also sometimes the worst) of all possible worlds, the experimentally determined answer may prove not to be zero thereby reflecting 'new physics' or (more likely) one or more unrecognized error sources which have caused the answer to differ from zero. (A reader who is interested in learning more about null-type experiments might begin by looking at §1.2, pages 5–9, 'The Inverse Square Law or the Mass of the Photon', in the second edition of *Classical Electrodynamics* by J. D. Jackson (John Wiley & Sons, Inc. 1975). It might be of some interest to those in the big G community that the first of the four (electrostatics) experiments cited in that section was performed by Henry Cavendish!) Even for non-null-type experiments, measurements of a particular quantity or constant are oftentimes treated by scientists as if they were 'null-type' experiments, because when the expected answer is known, any result differing by more than one sigma from the expected result will likely trigger a search for additional error sources to 'correct' the obtained answer to bring it into agreement with the anticipated result.

One problem facing science and scientists today is image. This is not a new problem, and I would suggest that much of the problem is self-inflicted. Because of the Public's oftentimes disinterest in science (figure 1) scientists need to work on making themselves and the importance of their work more accessible to layperson understanding. In today's world of specialization, one experiment, for which the basic idea of the measurement is accessible and interesting to all scientists, as well as to lay persons, is the measurement of G . The reason: because they can 'understand' it.

2. Credo

At this point before going on—and for full disclosure—let me tell you (confess to you) what I believe.

I believe that measurement capability is the enabler of almost all scientific progress. (By extending the reach of our hands and quickening the response of our eyes, new measurement methods and instrumental capabilities have driven and implemented much of scientific progress.)

I believe that Universities as well as some National Institutions are not unlike game preserves. (They provide a protective environment where certain types of individuals survive far better than they would in the real world outside of these ‘preserves’. And, given this, the inhabitants of these preserves owe something in return to those who are paying for the ‘fences’.)

I believe, as Gauss once wrote, that ‘Ideas are like violets in the springtime; they spring up in many places at the same time’ (Gauss writing to the elder Bolyai about Bolyai’s son’s discovery of non-Euclidean geometry in which he urged the younger Bolyai to publish!) ... and, I would add, these same ideas spring up in different centuries.

I believe that more often than not it is a chance observation which results in scientific progress; but it is only those chance observations seen with a prepared mind which have a chance of giving rise to scientific progress. (Fleming on receiving the Nobel Prize in 1945 said that, ‘In my first publication I might have claimed that I came to the conclusion as a result of serious study of the literature and deep thought, that valuable antibacterial substances were made by moulds and that I set out to investigate the problem. That would have been untrue and I preferred to tell the truth that penicillin started as a chance observation. My only merit is that I did not neglect the observation and that I pursued the subject as a bacteriologist’.)

3. A few words about science

One question which the public as well some managers of science find difficult to understand is ‘Why does scientific progress have to take so long?’ A friend of mine (Tuck Stebbins) once quipped, ‘They want acorns without oak trees’. (An oak tree, as does science, requires many years to mature.)

In a talk which I have given on the physics of basketball, I show a photograph (figure 2) from an early (*ca* 1893) basketball game which shows that initially the (peach) basket attached to the top of the pole had no hole in its bottom. This greatly slowed up the game for, after each score, the ball had to be ‘extracted’ from the basket by either shinnying up the pole or standing on another player’s shoulders (not to ‘see further’ but simply to get higher in order to access the ball). Accordingly, early scores of basketball games were quite low... until someone had the brilliant idea of cutting out the bottom of the basket—an idea which transformed the character (i.e. flow) of the game. (I was recently told that an intermediate strategy had been devised whereby a small hole was put in the bottom of the basket through which a long stick could be inserted to poke out the ball.) In any case, it took over 10 years for this ‘obvious change’, i.e. cutting out the bottom of the basket, to be made. Ten years!!! Though obvious in retrospect as the thing to do, it was not obvious until someone, who was living in a world of basketball games being played without a hole in the bottom of the basket, actually suggested doing this. (Copernicus, who lived in an Earth-centred Universe, faced this same kind of thought-steering mindset which his thinking had to overcome.)

It has been said that ‘Science came down from heaven to earth on the inclined plane of Galileo’ [1, p. 273]; and indeed, much of the joy in doing science is related to building things (doing experiments) which help us understand Nature’s processes.

Scientists are not always as effective as they should be in communicating their research to the public. On the other hand, if we are to communicate science—be it in the classroom or in the popular press—then I believe this communication needs to be done by real (practising and understanding) scientists. There is an old adage that goes, ‘Research is to teaching as sin is to



Figure 2. Basketball game *ca* 1892 with solid-bottomed peach basket. (Photo from Naismith Memorial Basketball Hall of Fame.)

confession...if you haven't been involved with the first, then you need not involve yourself in the second'. There are, of course, exceptions to this adage. I was a friend of Prof. Eric Rogers who taught for years at Princeton; although he was not doing any research at that time, he was nevertheless one of the greatest teachers, and also one of the best physicists, that I have ever known—without a doubt one of the exceptions.

In 1940, James Bryant-Conant, a very good chemist who was at that time President of Harvard University, said that 'The stumbling way in which even the ablest of scientists in every generation have had to fight through thickets of erroneous observations, misleading generalizations, inadequate formulations and unconscious prejudice, is rarely appreciated by those who obtain their scientific knowledge from textbooks'. Though no longer a popular view, I continue to believe that the researcher part of the researcher-teacher paradigm is important in revealing and communicating the true picture of science and discovery.

4. Precision measurement

Let us now turn to the first subject mentioned in the title of this paper, precision measurement. Precision measurement has as its purpose raising the quality of new and old measurements and measurement procedures through an increased understanding of the physical laws involved and of the material processes which underlie and dictate the quality of measurement. In precision measurement new concepts, new techniques and new insights are used to deal with fundamentally related high-precision and applied measurement problems which are encountered in a broad range of science and technology. As measurement requirements become more and more demanding it requires sophistication both to advance present-day technology and to apply it to these problems. It is this continuous challenge to our skills and to our creativity which makes working in this field challenging, interesting and fun. Precision measurement is not, however, a field which has received much applause from the scientific community or the general public.

In their 1983 *Science* article, 'Precision Measurements and Fundamental Constants' [2], Pipkin & Ritter begin their article with the (discouraging) sentence: 'An important but *rarely recognized* area of the physical sciences consists of metrology, the science of measurement, and

Throughout the Year we are Celebrating ALL Areas of Physics

Astrophysics
Biological
Computational
Fluids
Optical
Nuclear
Plasma
History



Atomic
Chemical
Condensed Matter
Polymer
Materials
Particle
Education

Photos courtesy of (from left to right): Contemporary Physics Education Project, University of Pittsburgh, Anglo-Australian Observatory, Epping Australia, ISIS, Rutherford Appleton Laboratory, Charles Townes, Brookhaven National Laboratory, AIP Neils Bohr Library, and P. Tuffy, University of Edinburgh.

Figure 3. There was room. (American Physical Society centennial year viewgraph.)

the determination of the fundamental constants required for relating measurements to theory'. One can only wonder what has happened to the 1930s phrase, 'The romance of the next decimal place' [3].

In 1999, we see this lack of recognition again on the occasion of the centenary of the American Physical Society, when a 'viewgraph' (figure 3) was sent to the 200 APS-designated centenary speakers entitled 'Throughout the Year we are Celebrating ALL Areas of Physics'. The first thing I discovered on receiving this APS-provided viewgraph was that somehow in making up their 'comprehensive' list they had overlooked both gravitational physics and precision measurement—the two areas I find most exciting. My sense is that talking about physics without recognizing the role of measurement science is like trying to write a poem without metre or an organ piece without a pedal line. On inquiring, I was told by APS's Judy Franz that they 'simply didn't have enough space;' and this was in spite of the glaringly empty space following the last (important but rather arbitrary) entry, Education. I would note that the builders of most edifices do find space for the foundation. A similar concern appears in the April 1993 editorial of the British publication *Physics World*, which begins with the words 'The science of measurement has a low profile'. This indifference—neglect of the pedal line—is difficult to understand.

Let me at this point mention a question which I used to puzzle about. That question is, 'Should this field be precision measurement or precision measurements—with or without the *s*?' The answer I believe is precision measurement without the *s* because were it precision measurements you would be given the idea that there is a precision measurement here, another precision measurement over there, and another one somewhere else, and they are all *unrelated*. If this were in fact true, then it would bode poorly for any substantial progress in this field because there would be absolutely no transfer-learning which can be applied to a new measurement from what one has learned from having done other precision measurements. On the other hand, if it is precision measurement, where what one learns from one experiment can also be applied to other 'different yet similar' experiments, then there will be a transfer-of-learning benefit which guides one's thinking on how to proceed and what to worry about in any newly undertaken experiment. Since I believe that I have learned much which is broadly applicable from each of many such experiments that I have been involved in, I see precision measurement science



Figure 4. Embellished Pastorale in F (Johann Sebastian Bach).

(without the *s*) as denoting a group of related experiments which, though they may have quite different sounding names, have much in common. . . and therefore a benefit is to be had by those working in this—broad but single—field of research.

To put the importance of this ‘learning transfer’ another way, consider two different ways of picking apples. One approach would be to take a ladder and a basket, disappear into an orchard, find a reasonable tree, and then proceed to pick every apple on that tree. Some areas of physics seem to proceed quite successfully this way with those scientists who are involved carrying out experiments which are very similar or even functionally identical. Another approach is to pick from a given tree a particularly interesting apple and then move on to another tree before picking some other particularly interesting apple—one of considerable difference in colour, size and shape. This latter approach is similar to what happens when one carries out individual precision measurements. Each picking (measurement), however different, still shares some degree of commonality. They may be different trees but they are situated in the same ‘field’. And it is for this reason that I feel measurement science should be thought of as involving precision measurement, not measurements. Even though the experiments are different, the techniques (fortunately) are common.

So what are some of the characteristics of precision measurement science? First, like the pedal line in an organ piece, precision measurement underpins, sustains and supports all of the activity in the rest of physics (the melodies played using the upper manuals; figure 4). Secondly, the experimentalist’s tongue-in-cheek motto that ‘a month or two in the laboratory will save you an hour in the library’ simply does not apply to this field. The subtleties of the precision measurement trade rarely, if ever, find themselves included in journal articles with the necessary detail. As a result, the subtleties almost always need to be discovered (alas oftentimes rediscovered) in the laboratory.



Figure 5. It is oftentimes very difficult to recognize the source of a problem. (© Ed Frascino/The New Yorker Collection/The Cartoon Bank.)

5. Scientific personalities

One thing which is both good and bad is that scientists are human; they want to get the ‘right’ (read, previously obtained) answer. Furthermore, the error budgets which scientists create in order to determine the accuracy of individual measurements are fundamentally flawed because they do not include what was not thought of. And, it is often very difficult to recognize the source of a problem (figure 5). The human scientific response to other than ‘agreement’ is to rethink and expand the scope of the error budget until the right number results. All of these factors (big *G* is small, scientists *are* human, and error budgets *are* flawed) make determining the precise value of the Newtonian constant of gravitation exceedingly difficult.

As previously mentioned, one problem facing scientists today is image. This is, however, not a new problem. Science and scientists have long been tagged with images such as that described by Charles Dickens in *The Lamplighter* (1838), ‘He was dressed all slovenly and untidy, in a great gown. . . , with a cap. . . ; and a long old flapped waistcoat; with no braces, no strings, very few buttons. . . Tom knew by these signs, and by his not being shaved, and by his not being over-clean, and by a sort of wisdom not quite awake. . . that he was a scientific gentleman’. And, more recently, by Richard Rhodes in his book *The Making of the Atomic Bomb* (1986), where he writes ‘The bomb was a weapon created not by the devilish inspiration of some warped genius but by the arduous labor of thousands of normal men and women working for the safety of their country. *Yet they were not normal men and women, they were scientists*’ (emphasis mine).

Contrary to the above-stated beliefs, the scientific personality is human; scientists want to be accepted by society and respected by their scientific peers. In the latter case, this means they want to be seen getting the ‘right’ answer. However, it is precisely this human aspect of the scientific personality which compromises their ‘objectivity’ in carrying out precision measurement experiments; for scientists are deathly afraid of being seen as having missed the boat (figure 6).

Monwhea Jeng, in his 2006 *American Journal of Physics* article, ‘A selected history of expectation bias in physics’ [4], describes how well-known scientists such as Galileo, Roemer, Einstein, etc., all showed the effects of an ‘expectation bias’ in the research that they did. This expectation bias can be described as follows, ‘Because we are aware of earlier results (or predictions) we tend to look for and find systematic errors which permit us to correct our result until it stands at least in the shadow of those measurements—at this point, we stop looking, fold up our equipment,



Figure 6. Scientists (and unicorns) are unhappy when they are seen missing the boat. (© 1956 Charles Addams. Reproduced with permission from Tee and Charles Addams Foundation.)

and publish our “new” result in substantial agreement with. . .’ [5]; and, as has recently been put more succinctly, ‘There is an almost irresistible pressure **to stop** when the result is about what one expects it **to be**’ [6, p. 455].

You might want to believe that grown men and women—especially scientists—would never fall prey to such things. Let me give you one particular example, of many that I am aware of, to think about. It comes from the 1961 *JOSA* paper by Mandelberg & Witten [7], who redid the Ives–Stilwell experiment of 1938 [8,9] (an experiment which was also performed by Otting in 1939 [10]) involving a measurement of the second-order Doppler shift. At the beginning of the paper, Mandelberg & Witten justify why they are redoing this experiment stating, ‘An analyses of the experiments of Ives and Stilwell and Otting indicates that although their reported experimental points seem to fit the curve with an accuracy of about 2 to 3% the experimental uncertainty is more nearly 10–15%’ [7, p. 529]. The above quotation is from the first page of their paper; now here is what they say on the final page of their paper. ‘In the region described approximately by $0.0045 < \beta < 0.0065$, lie eight consecutive points we have measured, each of which lies below the theoretical line. However, in this region most of the points measured by Ives and Stilwell lie above the line; taking the experiments together gives a reasonable spread of measured points about the theoretical line in this velocity range’ [7, p. 529]. I hope the readers of this paper will look at this *JOSA* paper and think very carefully about those two pages. I doubt very much that either Mandelberg or Witten realized the bias evidenced in what they had written.

There is also a very human concern (and therefore a hesitation before proceeding) not to ‘rock the scientific boat’. A. G. Shenstone (~1958, personal communication), who was the Chair of Princeton’s physics department for many years, told me that in 1938 Houston saw deviations from Dirac theory for H_α but attributed it to a grating ghost and thereby missed the Lamb-shift. Pound & Rebka [11] delayed publishing the results of their red shift experiment, because it initially disagreed with the Einstein’s prediction (R. H. Dicke ~1959, personal communication) until they found that their speaker-driven Mossbauer foils were breaking up into Chladni figures rather than parallel translating as they had assumed. Correcting for this brought their number into agreement with the expected result, and they published shortly thereafter. The LURE (Lunar Laser Ranging Experiment) Team initially had a non-Einstein-predicted result in its test of the strong principle of equivalence. LURE experimenters had used lunar laser ranging to measure

the differential free-fall of the Earth and the Moon towards the Sun and from their analysis gravitational energy appeared not to fall at the same rate as does mass and other forms of energy. Accordingly, they delayed by some months the publication of this result. Then—having been told by the MIT group that their analysis of the same data gave agreement with Einstein—on, once again, rechecking their analysis they found that the inclusion of a ‘small’ term, which they had initially considered small enough to neglect, removed the apparent discrepancy ... at which point they and the MIT group simultaneously published [12,13].

6. Gravitational physics

Though gravitational Physics is one of the oldest areas of science, it is nevertheless a subject which has come of age in recent times. Today’s experimental techniques and capabilities are for the first time capable of measuring with the levels of precision needed to see some of gravity’s more subtle effects, e.g. gravitational waves. However, gravity, ‘A property of bodies perceptible to the vulgar when things fall to the ground, but long acknowledged by this [the Royal] Society to be a quality impressed by the Creator on all matter. ...’ [14, p. 4], is still far from being ‘understood’, in part because it remains disconnected from the rest of Nature’s forces. Though it is often said that the devil is in the details, in the case of gravitational waves, the detection prospects (not just the devil) are also in the details and depend critically on creativity and experimental cunning being applied to the myriad details which are associated with these gravity wave detectors, which constitute nothing less than a precision measurement experimental *tour de force*.

In the case of the experiments which have been carried out to determine a value for big G (many of which are described in some detail in this special edition of the *Philosophical Transactions A of the Royal Society*), the really quite large scatter in these results either reflects ‘new physics’ (an easy way out) or more likely the presence of one or more unrecognized systematic errors which are biasing the results. This discussion will come up again later.

7. The Newtonian constant of gravitation

The title of this Theo Murphy International Scientific Meeting is, ‘The Newtonian constant of gravitation, a constant too difficult to measure?’ An alternative title for this meeting might have been, ‘The Newtonian constant of gravitation, a constant too small to measure’. The value of the gravitational constant, G , remained unknown for over half a century after Newton. A rough estimate of G from guesses like Newton’s of the average density of the Earth showed that the attractions between objects in a laboratory would be almost hopelessly small. The familiar forces of gravity are large; but they are due to the Earth’s huge mass. The first experimental attempts to measure big G were experiments of desperation. Scientists used various mountains to deflect pendulum masses which were hung adjacent to their particular mountain, and this produced a rough (at best 10%) measure of G . Experimentally, they had to estimate (geologically) the mass of their mountain and its ‘distance’ from the hanging mass—not things which could be measured accurately.

To get a feeling for the ‘in-a-laboratory smallness’ of big G , one need only contrast the laboratory-felt strength that big G produces with the strength of the other gravitational quantity, little g . The differences are stunning. Little g , the acceleration due to our (large) Earth’s gravity causes a dropped object to fall roughly 5 cm in a time of 0.1 s...and if we only knew (but we do not) the exact mass of the Earth and its effective radius...our task of determining an accurate value for big G would be done. On the other hand, the additional free-fall distance because of the Newtonian gravitational attraction caused by a 5 lb sack of sugar pulling down (in addition to little g) on this dropped object from a distance of something like 1 ft, would cause this same object to fall an additional 5×10^{-10} cm during this same 0.1 s—not an effect that can be readably observed or, for that matter, a measurement which can be easily made. Though most measurements of G have involved the use of a torsion balance, a pan balance of some kind, or a vertical pendulum supported at its mass centre, in recent times this direct approach to measuring

G has been tried [15–17] with respective accuracies of 1.4, 5.0 and 1.7×10^{-3} —none fully up to the accuracy reported in 1896 by C. V. Boys, whose work is discussed later in this section. (At this meeting G. M. Tino told us about his group’s latest free-fall measurement of G that again used atom interferometry and which had an accuracy of 1.5×10^{-4} ! A paper about this now-becoming-interesting determination appears in these proceedings.)

While the measurement accuracy of little g has increased by eight orders of magnitude during its 400-year measurement history, the measurement accuracy of big G has only increased by three orders of magnitude during its 300-year measurement history. Why, you might ask, has big G ’s accuracy not benefitted from a similar sort of increase? I think the reason rests with the fact that the concomitant ‘noise sources’ have stood their ground (not gotten any smaller) and as such, they continue to hide the really small ‘big’ G signal. On the other hand, little g ’s sea of systematic problems does not appear until one reaches today’s obtained accuracies of a few parts in 10^9 . Accordingly, from this time forward, further increases in the accuracy with which little g can be measured will be hard won.

One other factor is that the measurement of little g has many practical uses and as such, has enjoyed considerable ‘research support’ because of these applications. Big G on the other hand, has essentially no practical uses (knowing its value to 1% satisfies the few demands on knowing its actual value) and accordingly, the kind of continuing support which would be required to make a focused frontal attack on its accuracy has been missing. I know, and many readers of this paper will also know, that this measurement is one of the toughest (and therefore most useful) training courses for precision measurement science. . .but its value as a source of ‘experimental pushups’ is a hard sell to a management whose horizon is ‘now’ or to funding agencies who have the task of ‘feeding the 5000’—supporting all of science.

Let me now detail a few things relating to several historical G experiments and then move on to today’s experimental situation—a situation which is not good; three of the most recent G determinations differ by nearly 5 parts in 10^4 !

Following the early experiments carried out by the ‘mountain men’ came the experiment of Cavendish. I want to point out several aspects of this experiment. Whereas the focus on the first page of Cavendish’s ‘Determining the Density of the Earth’ paper [18] (In those day’s no one was measuring G , rather they were all weighing the Earth.) is usually on the opening words of the second paragraph, ‘The apparatus is very simple’. I think much more interesting—and a better measure of Cavendish the man—are the words in his first paragraph where he credits Michell with the idea (an idea which independently occurred to Coulomb some 27 years later) of using a torsion balance to measure very small forces. Were I ever to teach a course on ‘100 great ideas in physics’, Michell’s torsion balance would be one of them. The beauty of a torsion balance is that it gets rid of (or, in practice, mostly gets rid of?) little g , whose effects, as we have seen, are many orders of magnitude larger than the big G effects which one is trying to sense and measure.

The other thing which I found particularly interesting in Cavendish’s paper appears on pages 484 and 485 where he describes (though he did not know what to call them) the anelastic effects which he observed in his measurements of the deflections of his torsion balance:

These experiments are sufficient to shew [*sic*], that the attraction of the weights on the balls is very sensible, and are also sufficiently regular to determine the quantity of this attraction pretty nearly, as the extreme results do not differ from each other by more than 1/10 part. But there is a circumstance in them, the reason of which does not readily appear, namely, that the effect of the attraction seems to increase, for half an hour, or an hour, after the motion of the weights as it may be observed, that in all three experiments, the mean position kept increasing for that time, after moving the weights to the positive position; and kept decreasing, after moving them from the positive to the midway position. [18, pp. 484–485]

This observation of Cavendish was made nearly 200 years before Kuroda’s *Physical Review Letter* [19] in which he describes the importance of taking into account anelasticity in all torsion

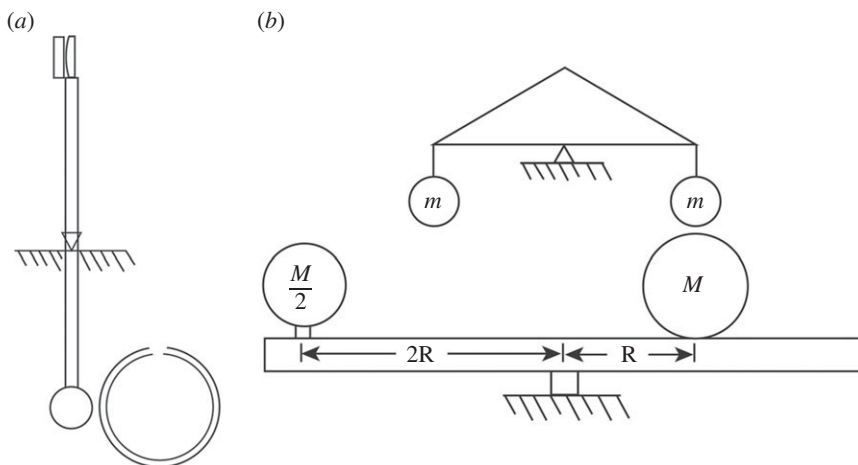


Figure 7. (a) Suggested method for determining G , Dr W. Laska, 1889. No results were ever published. (b) Prof. J. H. Poynting, 1891. At first, no counterweight ($M/2$) was used.

balance-based big G determinations. And interestingly, if one ‘corrects’ Cavendish’s value for G (having transcribed his value for the density of the Earth to a value for G) using a reasonable guess for the anelasticity of the approximately 11 mil diameter Cu wire which he used for his fibre, one arrives at a value for G which agrees with modern measurements!

After Cavendish’s determination, there are three experiments which I find particularly interesting. The first of these was an experiment which was never done. In 1889, Dr W. Laska proposed to use ‘Newton’s rings’ to sense the deflection of a vertical dumbbell pendulum (figure 7a) which (to the best of my knowledge) was the first ‘suggested use’ of optical interferometry to measure the very small resultant (sideways) deflection which often needs to be measured in precision measurement experiments. (Incidentally, a vertical pendulum of the type shown also gets rid of little g provided the bar is perfectly balanced with its mass centre on the knife edge.) A second experiment which I find interesting was performed by Prof. J. H. Poynting (of ‘the’ vector fame), who measured the resulting tilt of a two-beam balance when the gravitation force of a mass acted first on a small mass on one side and then on the small mass on the other side of the balance (figure 7b). This double pan balance approach also serves to remove little g from interfering with the measurement. At first no counterweight ($M/2$) was used with the result that 1 year of work was lost in measuring little but the tipping of the floor. Why do I find this both interesting and important? Because instrumental tilt is a problem which has again and again had to be addressed in virtually every precision measurement experiment that has been performed, and missing it puts one in good company.

One hundred years after Cavendish, C. V. Boys (one of my scientific heroes; he wrote papers which are truly enjoyable to read) improved our knowledge of big G by a factor of 10 which resulted in the value for this constant being known to 1×10^{-3} ! He also saw himself as measuring G —not the density of the Earth. In his talk at the Weekly Evening Meeting of the Members of the Royal Institution of Great Britain on 8 June 1894, Boys strongly objected to his work being characterized as measuring the density of the Earth as opposed to measuring a constant of universal import:

In spite of the courteously expressed desire of your distinguished and energetic secretary, that I should indicate in the title that, to put it vulgarly, I had been weighing the earth, I could not introduce as the object of my work anything so casual as an accidental property of an insignificant planet... The earth has no more to do with it than the table has upon which the apparatus is supported. [20, p. 356]

Further of interest to precision measurers was his noting that his introduction of fused-silica fibres, which made possible his 0.1% measurement accuracy, meant that big G was then known with accuracy comparable to that with which the electrical and magnetic units were known:

...knowing that by my discovery of the value of the quartz fibre, and my development of the design of this apparatus, I had, for the first time, made it possible to obtain the value of Newton's Constant with a degree of accuracy as great as that with which electrical and magnetic units are known; though I have up to the present succeeded to an extent which is greater, I believe, than was expected of me, I am not yet entirely satisfied. I hope to make one more effort this autumn, but the conditions under which I have to work are too difficult; I cannot make the prolonged series of experiments in a spot remote from railways or human disturbance; I cannot escape from that perpetual command to come back to my work in London; so after this I must leave it, feeling sure that the next step can only be made by my methods, but by some one more blest in this world than myself. [20, p. 372]

This was the first *and last* time that such an 'accuracy' parity amongst the various standards—including G —had been established!

The most important big G determination made between Boy's and today's more recent determinations was the experiment of Luther and Towler which was done at what was then the National Bureau of Standards. This 7.5 parts in 10^5 experiment, which was published in 1982, was the gold standard of big G determinations for some time and served as the principle basis of CODATA's 1986 recommended value for this constant.

At this point, I would like to take advantage of my 'author's privilege' and use the experiment Harold Parks and I did as a vehicle to talk more about errors including systematic errors and also to comment on today's (embarrassing) experimental situation with respect to G . Let me first point out that had we started our experiment 3 or 4 years earlier, we would have published **at that time** our new result in substantial agreement with the 1986 CODATA value which was—to a large extent—based on the 1982 measurement of Luther & Towler [21]. However, in 2000 before we finished our experiment, Jens Gundlach & Steve Merkowitz [22] at the University of Washington published their big G determination which was five times more accurate than the determination of Luther and Towler. CODATA may have gotten carried away by the concern of there being no anelasticity correction made in the Luther–Towler experiment (Kuroda's *PRL* [19] on this subject appeared in October 1995), since they removed the Luther–Towler experiment as a contributor to the 2002 CODATA-recommended value for G , and essentially adopted the Gundlach–Merkowitz value as their new recommended value for the Newtonian constant of gravitation.

Accordingly, as we then felt that we had gotten the 'wrong number', we spent some 5 years looking under every experimental rock and stone, checking and rechecking all of our calculations, and at the same time we rethought our various assumptions, remeasured and measured anew various things, and consulted with 'experts'—only to uncover nothing that would change our result by more than 1 or 2 parts in 10^5 ! What to do? We simply (and finally) published our new result which disagreed with CODATA's recommended value by some parts in 10^4 but which agreed with the Luther–Towler value at the one-sigma level. (I have to confess that early on I had not thought carefully through the implications of not correcting the Luther–Towler result (had the fibre Qs all been measured) for anelasticity; in fact, given the fibres used, a correction would have only slightly **lowered** the Luther–Towler published value.) I think that CODATA may have also rethought the implications of this as they again included the Luther–Towler number in their subsequent 2006 and 2010 recommendations. Figure 8 shows the history of CODATA's recommended values for G going back to 1973 where, as plotted, the position of the recommended G value moves to the right or to the left proportional to the recommended increase or decrease for this number. Note that the accuracy associated with the CODATA-recommended value for this constant has not improved in the past 25 years!

CODATA RECOMMENDED VALUES FOR G		
(SI Units)		
1973	6.6720 (41) x 10 ⁻¹¹ (6.1 x 10 ⁻⁴)	
1986	6.67259 (85) x 10 ⁻¹¹ (1.28 x 10 ⁻⁴) [Luther-Towler]	
1998	6.673 (10) x 10 ⁻¹¹ (1.5 x 10 ⁻³) [PTB]	
2002	6.6742 (10) x 10 ⁻¹¹ (1.5 x 10 ⁻⁴) [G&M]	
2006	6.67428 (67) x 10 ⁻¹¹ (1 x 10 ⁻⁴)	
2010	6.67384 (80) x 10 ⁻¹¹ (1.2 x 10 ⁻⁴)	

Figure 8. In parentheses are the recommended uncertainties; in the brackets are the ‘influencing’ measurements.

8. Reviewers and comments

Let me firstly conclude with a scientist’s version of ‘lessons and carols’, namely ‘reviewers and comments’. I believe that much is to be learned from the comments of the unusually thoughtful reviewers of the big *G* determination of Harold and myself—an experiment which is described in some detail in the proceedings of this meeting. So what can be learned?

Referee A wrote ‘Although the number of people who worry about the value of *G* is probably pretty small, discrepancies of these magnitudes are serious business, and it is important to bring this result to the notice of the standards community. The fact that the value reported here lies within one sigma of the value reported by Luther *et al.* is also interesting, although it might be a coincidence’.

I have known Gabe Luther for many years and during my not infrequent trips to NBS/NIST in Washington, DC, I always made a point of spending at least half a day in what always turned out to be useful discussions with him. And given my high respect for Gabe, I would like to think that the agreement between our two experiments is more than a ‘coincidence’.

Referee B thought the paper was well written and clear for the general reader but believed ‘that the reported discrepancy in the value of *G* is due to some unidentified systematic effect, and the (*PRL*) letter did not contain enough detail to independently confirm that all sources of error were identified. (Hopefully the Parks-Faller paper in the proceedings of this meeting will provide the “enough detail” that this reviewer wished for.) Nevertheless (this reviewer continues) it does appear the authors have done due diligence in investigating all the major sources of error in this very challenging measurement’.

The real problem with errors, including systematic errors, was put succinctly by Lennert Robinson of the Bureau International des Poids et Mesures (BIPM), who once said, ‘*You don’t include what you haven’t thought of*’. In a big *G* determination, one must eliminate all extraneous effects from the experiment or *completely* understand them (so that a correction can be made). An example of a systematic error may be useful at this point. To do this, let us return to Galileo’s inclined plane ‘motion’ experiments where he rolled various shapes (cylinders and balls) down slopes ranging between 5° and 10°. The choice of these slopes was indicative of his good experimental technique. (Use of much greater slopes would have removed most of the timing advantage while use of somewhat less of a slope would probably have yielded little except an unfavourable statement about the straightness of the board he was using).

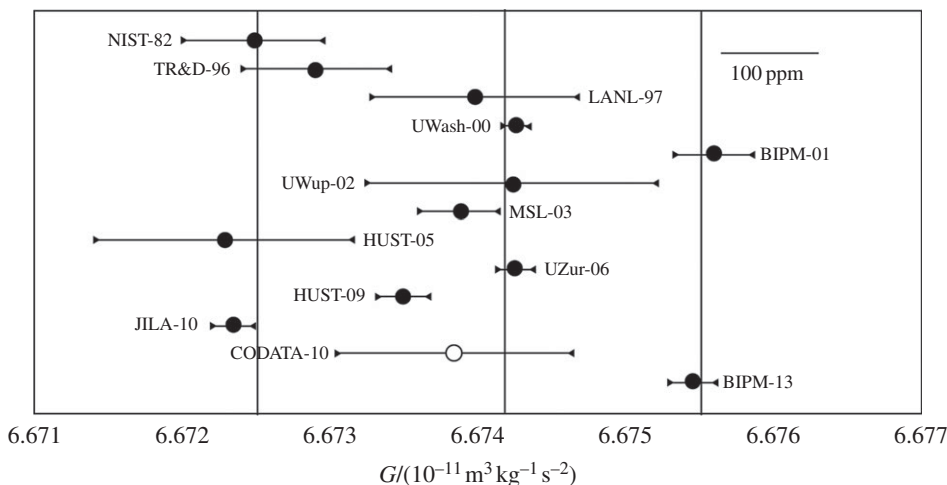


Figure 9. Author's view of 'three groups of data'.

Had Galileo actually tried to measure little g , instead of simply studying the accelerated motion of 'falling objects', his neglect of taking into account the rotational inertia of the rolled object would have resulted in a little g value which was low. Were it a solid ball, his measured value for g would have been something like 70% of g 's true value. However, each time he did the experiment he would have gotten the same (wrong) answer. One might also imagine that other scientists of Galileo's day would then have also made inclined plane measurements of g , using balls of different sizes and a variety of slopes, in the expectation that by thus varying the conditions of the measurement they would discover possible errors in the measurement procedure. However, making repeated experiments with different balls and using inclined planes of differing slopes—but neglecting (because they did not think or know about it) rotational inertia—would only have served to confirm a wrong (low) value for little g . This is why systematic errors, in particular, are a real problem in the determination of any constant.

Finally, Referee C said, 'This simple-pendulum measurement seems like a good one, and it is a nice alternative to the usual torsion balance measurements, also utilized by the UW people. I recommend publication. I leave it up to the community and CODATA to sort all this out. I have no recommendation on a value of G , nor most importantly its realistic error bar'. Here, I think the message the reviewer is trying to send is contained in his last seven words: '*nor most importantly its realistic error bar*'. (Assigning errors is easy if you know what they are... but what is one to do about 'the elephants in the room' which you are either unaware of or uncertain as to the significance (or insignificance) of their contributions?)

When I eyeball the big G determinations dataset which appears in the BIPM 2013 publication, I see the data as falling into three groups (figure 9). And though the centre group has a larger subscribing membership, I think it is unwise not to give at least some *non-zero* credence to the values which are associated with the measurements which fall in the two outer groups. One or the other of them just might possibly be right; and this would justify an (even larger) error bar than that which CODATA assigned to the value of G in its 2010 revision of the fundamental constants (figure 10). One additional observation: the distribution of the various measured values for G that is shown here is not even close to being Gaussian, which makes it difficult to 'interpret' this dataset.

Some months ago, Eric Cornell of JILA sent me a 'you probably know all about this' email (The 2013 BIPM result had just appeared in *PRL*) which concluded with these words (reproduced here with Eric's permission): 'The "historical perspective" plot (figure 3) [figure 10*b* this publication] is a useful lesson for all would-be precision metrologists, such as myself. I do not know what the right answer is, but whatever it turns out to be, multiple, talented metrology groups are way

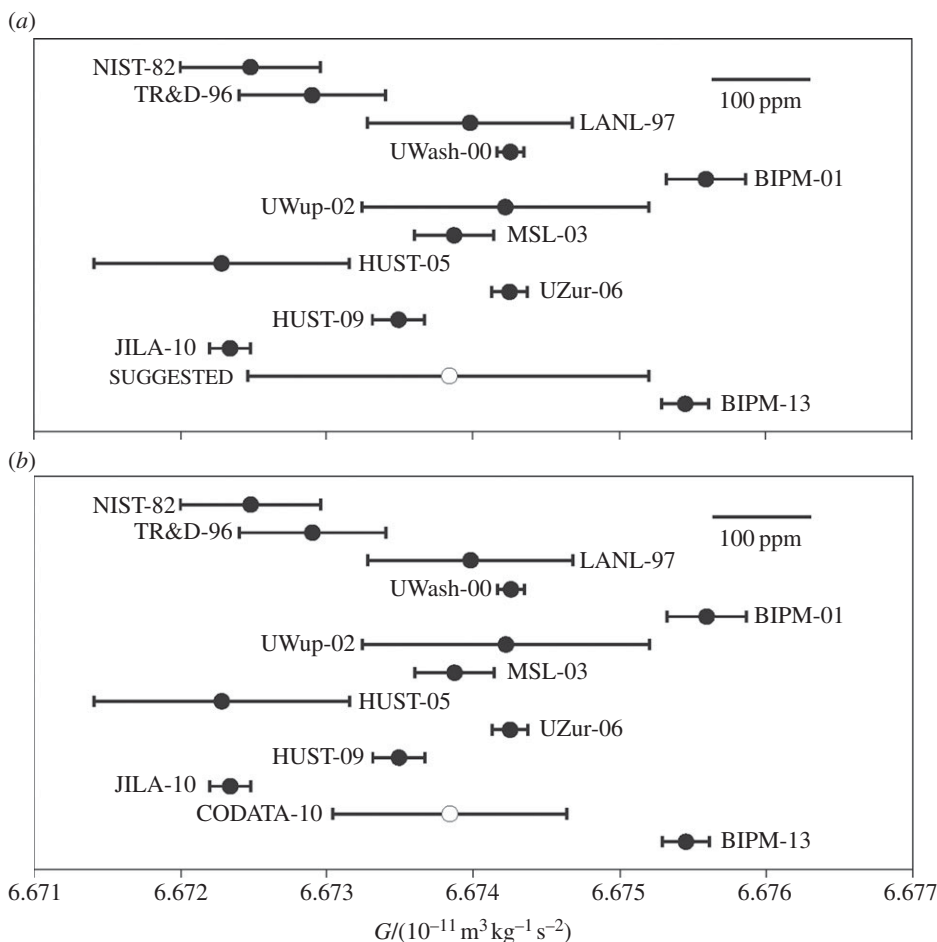


Figure 10. (a) Author's suggested error bar for CODATA-recommended value of G . (b) CODATA-recommended value of G and error bar.

off. The next CODATA value for G will likely need to have larger error bars'. This is exactly the point which I have been trying (unsuccessfully) to make now for some years. (I recognize that some might choose to dismiss my view on this issue claiming that I too am human, and that my opinion is probably clouded by self-interest.) Nevertheless, I do have quite a bit of respect for the precision measurement experience of those groups whose determinations fall in these 'outlying' G determinations. Accordingly, I believe that it would be quite reasonable to admit to the metrological community that things here are not nearly as tidy as we wish they were; and that G might in fact be somewhat higher (or lower) than that suggested by CODATA's 2010 recommended uncertainty for this particular fundamental constant which 'is too hard to measure'.

9. Benedictory remarks

I would lastly conclude with a few thoughts which I hope you will carry with you long after you have finished reading this paper. First, the subject matter at hand is that of precision measurement, not precision measurements with an s , because big G determinations and all other precision measurement experiments share a common knowledge base. Simply put, all of these experiments, which at first glance may appear different, are at heart similar. This allows one to draw on the over-a-lifetime generated and ever-evolving skill set which is needed to successfully work in this field.



Figure 11. Honoré Daumier, French, 1808–1879. *Monsieur Babinet informed by his porter of the arrival of the comet*. Plate 556 from the series *Actualités*, published in *le Charivari*, 22 September 1858. Lithograph, second state. Sheet 23 × 33 cm (9 1/16 × 12 in.) Museum of Fine Arts, Boston. Bequest of William Perkins Babcock, B4184.109. (Photograph © 2005 Museum of Fine Arts, Boston, MA, USA.)

I believe, as Daumier has pointed out (figure 11) in his *Monsieur Babinet Prévenu par sa Portière de la Viste de la Comète*, that it is sometimes useful and important in science to look in another direction (or try picking apples from another tree). This same look-in-a fresh-direction approach, or looking at a problem the other way around, can sometimes be helpful in overcoming one's cherished belief that everything was done correctly and that all possible extraneous effects have been accounted for.

Precision measurement physics is a field, however, in which 'something is usually not for nothing'. R.V. Jones said in this connection, '...for in contrast with much of humanity, Nature usually "plays fair", and he who wrestles with it to make a better instrument (or I would suggest measurement) although often exasperated by Nature's "sheer cussedness", will sometimes find his efforts rewarded' [23, p. 376].

I have sometimes put this same idea as follows: 'Though Nature holds all of the aces, she never deals off the bottom of the deck'.

It is also an area of physics where 'Much more is known than is actually true' (J. R. Pierce).

Precision measurement science is a field of physics in which John Wooden's, 'I would have been disappointed if mere physical strength would ever prove superior to finesse, dexterity and maneuverability' (said in describing UCLA's win in the finals of the 1965 NCAA tournament when they defeated Kentucky) also applies to how one should do experiments in physics—with grace.

I completely agree with Humphrey Davey who noted that, 'we're not smarter than those working in this field 300 years ago...but rather we enjoy a richer technology together with an expanded experimental experience which gives us "today's edge" in doing these experiments'.

Were I to ask what characteristics are both needed and helpful in carrying out precision measurement experiments, I would suggest in addition to faith, hope and pride (Dickens writing in *Nicholas Nickleby* (figure 12)) that on adding 'patience' (figure 13) one would then have all of the needed ingredients.

Some years ago, the American magician, Houdini, made a pact with his wife that when one of them died, the remaining partner would, once a year, attempt to make contact with the departed, via séances, etc. Houdini's wife died first so each year on 31 October, Houdini would arrange to attend a séance with the aim of 'contacting' his wife. However, before going to these séances, Houdini would sandpaper his legs raw—making his pants-covered legs extremely sensitive detectors of things which were going on beneath the table which might not otherwise be noticed.

Pride is one of the seven deadly sins; but it cannot be the *pride* of a mother in her children, for that is a compound of two cardinal virtues—*faith* and *hope*?

Charles Dickens in *Nicholas Nickleby*
(Chapter 44, last page)

Figure 12. *Pride, faith* and *hope*—helpful ingredients in precision measurement experiments.

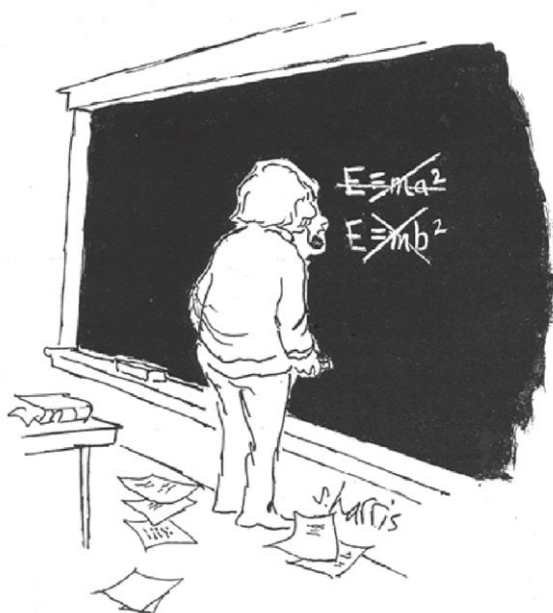


Figure 13. The remaining ingredient: *patience*. (Reproduced with permission of ScienceCartoonsPlus.com.)

Precision measurement science does not (fortunately) involve the sandpapering of anyone's legs; it does however—with each new experiment performed—sandpaper (i.e. increasingly sensitize) a scientist's mind to better deal with all of Nature's subtleties.

Finally, the determination of G is quite demanding of time and techniques. Einstein once said 'I have little patience with scientists who take a board of wood, look for its thinnest part, and drill a great number of holes where drilling is easy' [24, p. 350]. Precision measurement science, and particularly determining the value of G , is a 'thick board'.

Acknowledgements. I thank, in particular, Gwen Dickinson of JILA for her much needed help in working with me on this document. I would like to thank JILA for their continued support even though I am 'officially' retired. I also thank Harold Parks for his thoughtful reading of this manuscript. Many, if not most, of the people who 'trained' me, and who engendered in me my love of science, are now deceased. But I thank

them anyway in the hope that (from somewhere) they are looking down and reading this manuscript and my acknowledgement of them. A few of my friends have occasionally suggested that I ‘smile’ too much in my talks and in my writings, and, therefore, I may not be taken as a serious scientist. On the other hand, a fair number of people who know me and accept me as a serious person, when asked if I should change my speaking and my writing style have said, ‘Please don’t’. Maybe they are just being kind; but I thank them anyway for their support. Finally, I thank my wife, Jocelyne Bellenger, for her love and support even when I disappear into one of these writing journeys.

References

1. Rogers EM. 1960 *Physics for the inquiring mind, the methods, nature, and philosophy of physical science*. Princeton, NJ: Princeton University Press.
2. Pipkin FM, Ritter RC. 1983 Precision measurements and fundamental constants. *Science* **219**, 913–921. (doi:10.1126/science.219.4587.913)
3. Richtmyer FK. 1932 The romance of the last decimal place. *Science* **75**, 1–5. (doi:10.1126/science.75.1931.1)
4. Jeng M. 2006 A selected history of expectation bias in physics. *Am. J. Phys.* **74**, 578–583. (doi:10.1119/1.2186333)
5. Faller J. 2002 *Precision measurement with gravity*. Invited oral presentation at the 2002 *Conference on Precision Electromagnetic Measurement, Ottawa, Canada, 16–21 June 2002*.
6. Quinn T. 2014 Don’t stop the quest to measure big G. *Nature* **505**, 455. (doi:10.1038/505455a)
7. Mandelberg HI, Witten L. 1962 Experimental verification of the relativistic Doppler effect. *JOSA* **52**, 529–535. (doi:10.1364/JOSA.52.000529)
8. Ives HE, Stilwell GR. 1938 An experimental study of the rate of a moving atomic clock. *JOSA* **28**, 215–219. (doi:10.1364/JOSA.28.000215)
9. Ives HE, Stilwell GR. 1938 An experimental study of the rate of a moving atomic clock. II. *JOSA* **31**, 369–374. (doi:10.1364/JOSA.31.000369)
10. Otting G. 1939 Der quadratische Dopplereffekt. *Physikalische Zeitschrift*. **40**, 681–687.
11. Pound RV, Rebka Jr GA. 1960 Apparent weight of photons. *Phys. Rev. Lett.* **4**, 337–341. (doi:10.1103/PhysRevLett.4.337)
12. Williams JG *et al.* 1976 New test of the equivalence principle from lunar laser ranging. *Phys. Rev. Lett.* **36**, 551–554. (doi:10.1103/PhysRevLett.36.551)
13. Shapiro II, Counselman CC, King RW. 1976 Verification of the principle of equivalence for massive bodies. *Phys. Rev. Lett.* **36**, 555–558. (doi:10.1103/PhysRevLett.36.555)
14. Pringle J. 1775 *A discourse on the attraction of mountains, delivered at the Anniversary Meeting of the Royal Society, November 30, 1775*. London, UK: Royal Society.
15. Schwarz JP, Robertson DS, Niebauer TM, Faller JE. 1998 A free-fall determination of the Newtonian constant of gravity. *Science* **282**, 2230–2234. (doi:10.1126/science.282.5397.2230)
16. Fixler JB, Foster GT, McGuirk JM, Kasevich MA. 2007 Atom interferometer measurement of the Newtonian constant of gravity. *Science* **315**, 74–77. (doi:10.1126/science.1135459)
17. Lempore G, Bertoldi A, Cacciapuoti L, Prevedelli M, Tino GM. 2008 Determination of the Newtonian gravitational constant using atom interferometry. *Phys. Rev. Lett.* **100**, 050801. (doi:10.1103/PhysRevLett.100.050801)
18. Cavendish H. 1798 Experiments to determine the density of the Earth. *Phil. Trans. R. Soc. Lond.* **88**, 469–526. (doi:10.1098/rstl.1798.0022)
19. Kuroda K. 1995 Does the time-of-swing method give a correct value of the Newtonian gravitational constant? *Phys. Rev. Lett.* **75**, 2796–2798. (doi:10.1103/PhysRevLett.75.2796)
20. Boys CV. 1896 The Newtonian constant of gravitation, in *Notices of the Proceedings at the meetings of the members of the Royal Institution of Great Britain, vol. XIV (1896), 1893–1895*, pp. 353–377. London, UK: William Clowes and Sons Ltd.
21. Luther GG, Towler WR. 1995 Redetermination of the Newtonian gravitational constant G. *Phys. Rev. Lett.* **48**, 121–123. (doi:10.1103/PhysRevLett.48.121)
22. Gundlach JH, Merkowitz SM. 2000 Measurement of Newton’s constant using a torsion balance with angular acceleration feedback. *Phys. Rev. Lett.* **85**, 2869–2872. (doi:10.1103/PhysRevLett.85.2869)
23. Jones RV. 1988 *Instruments and experiences*. London, UK: John Wiley and Sons.
24. Frank P. 1949 Einstein’s philosophy of science. *Rev. Mod. Phys.* **21**, 349–355. (doi:10.1103/RevModPhys.21.349)