Assessing uncertainty in physical constants

Max Henrion
Carnegie-Mellon University, Department of Engineering and Public Policy, Pittsburgh, Pennsylvania 15213
Baruch Fischhoff
Decision Research, A Branch of Perceptronics, 1201 Oak, Eugene, Oregon 97401

(Received 7 March 1985; accepted for publication 12 September 1985)

Assessing the uncertainty due to possible systematic errors in a physical measurement unavoidably involves an element of subjective judgment. Examination of historical measurements and recommended values for the fundamental physical constants shows that the reported uncertainties have a consistent bias towards underestimating the actual errors. These findings are comparable to findings of persistent overconfidence in psychological research on the assessment of subjective probability distributions. Awareness of these biases could help in interpreting the precision of measurements, as well as provide a basis for improving the assessment of uncertainty in measurements.

I. INTRODUCTION

Accurate estimates of the fundamental constants of physics, such as the velocity of light or the rest mass of the electron, are central to the enterprise of science.[1] Like any measurements, they are subject to uncertainties from a variety of sources. Reliable assessments of this uncertainty are needed (a) to compare the precision of different measurements of the same quantity, (b) to assess the accuracy of other quantities derived from them, and, most crucially, (c) to evaluate the consistency of physical theory with the current best measurements. Thus, as Eisenhart has pointed out, "A reported value whose accuracy is entirely unknown is worthless."[2]

It is not unusual to encounter individual examples of errors in measurements of physical quantities that turn out to be disconcertingly large relative to the estimated uncertainty. One well-known case was in R. A. Millikan's oil-drop experiment in 1912 to determine e, whose result turned out 15 years later to be off by 0.6% or three standard deviations due to reliance on a faulty value for the viscosity of air.[3] A more recent example concerns measurements of $|\nu_{+-}|$, the parameter that measures the degree of violation of CP (charge-conjugation-parity) invariance. The six measurements prior to 1973 agreed reasonably, but more accurate measurements since then differ consistently by about seven standard deviations from the pre-1973 mean, a discrepancy that remains unexplained in terms of experimental procedure.[4] Such extreme cases may be exceptions, but they raise the more general question of how well on the average reported uncertainties reflect actual errors, an issue on which there has been little systematic study.[5] Here we will present evidence from historical measurements of a range of physical constants to illustrate the scope of the problem of underestimation of uncertainty. A wider awareness of such results may help in interpreting reported uncertainties, and may have some important educational implications.

A *comprehensive* assessment of uncertainty cannot rest solely on statistical analysis. Unavoidably, it involves a considerable element of subjective judgment. Therefore, we shall first review some recent findings of cognitive psychology from laboratory studies of human judgment under uncertainty. After examining evidence from measurements of physical constants, we will discuss possible explanations for these problems in the light of the psychological literature, and explore the prospects for alleviating them.

II. THE PSYCHOLOGY OF JUDGMENT UNDER UNCERTAINTY

The premise of laboratory studies of human judgment is that all judgments are governed by a set of core cognitive processes.[6] If those can be understood in experimental settings, then reasonable speculations can be made about human performance in the real world.[7] This literature has revealed both strengths and weaknesses. Where people have the explicit training or where there has been the opportunity to receive clear, prompt feedback, people can assess many aspects of uncertain processes. For example, weather forecasters in the U. S. provide assessments of the probability of precipitation that are probabilistically *well calibrated*: It rains on about 70% of the occasions on which they forecast a 70% probability of rain.[8] They have developed this ability through years of hands-on experience, with guidance from computer models, with ample feedback, and within an institution that rewards them for candor (rather than, say, for exuding confidence or avoiding firm commitments). In less favorable circumstances, however, people often lack an intuitive feel for probabilistic processes, relying instead on mental "heuristics" (deterministic rules of thumb) to guide their judgments. Although often useful, these rules can lead to predictable biases.

In these studies, the intutive asessment of uncertainty has proven to be especially problematic.[9] People seem insufficiently sensitive to how much they know, so that changes in knowledge are accompanied by inappropriate changes in confidence. The most common problem is *overconfidence*. A common way to assess the precision of an uncertain quantity is by a subjective confidence interval, indicating a range within which the assessor believes the true value has, say, a 98% chance of falling. The *probabilistic calibration* of a set of such judgments for different quantities may be measured by comparing the assessed probability for the interval with the fraction of times the true value lies within it. Cases in which the true value (once known) turns out to fall outside the assessed confidence interval, may be termed "surprises." The *surprise index* is the percent of 98% confidence intervals for which the true value is a "surprise." Having significantly more than 2% surprises indicates overconfidence, in the sense of underestimating the range of possibilities. Conversely, too few surprises means underconfidence. Another measure of calibration is the *interquartile index*, the percent of judgments for which the true value lies between the assessed 25th and 75th percentiles. An interquartile index that is much less than 50% also indicates overconfidence. The overwhelming result in laboratory studies with nonspecialist assessors has been intervals that are too tight, reflecting overconfidence: Typically the surprise index is 20%–40% instead of the ideal 2%, and the interquartile index is 30%–40% instead of 50%.[9]

The important role of probabilistic judgment is coming to be recognized in several areas involving risk analysis, including medicine, toxicology, and nuclear safety.[10] The few studies of judgments in such real-world contexts outside the psychologist's laboratory suggest that the laboratory findings of overconfidence may generalize to situations of practical importance.[11] However, such evaluations have been rare. An evaluation of the uncertainty estimates for physical constants should therefore be of interest to the study of probabilistic judgment in general, as well as to the practice of physics.

III. UNCERTAINTY IN PHYSICAL MEASUREMENT

It is universally accepted that every scientific measurement should be accompanied by a statement of its uncertainty, nowadays usually expressed probabilistically, as a standard error or confidence interval. However, there is less agreement about precisely what these statements are intended to represent. Typically, the terms "error" and "uncertainty" are used almost interchangeably.[12] We find it convenient to distinguish them thus: "Error" is the actual difference between a measurement and the value of the quantity it is intended to measure, and is generally unknown at the time of measurement. "Uncertainty" is a scientist's assessment of the probable magnitude of that error.

Another, conventional distinction is between *random* error due to uncontrolled variability among observations, and *systematic* error, equal to the difference between the value to which the observed mean converges and the true value. Systematic errors can arise from undetected biases in the experimental apparatus, approximations in computational procedures, errors in auxiliary variables, and deficiencies in theoretical assumptions.

Random uncertainty, being simply an estimate of the random error, may be obtained straightforwardly by statistical analysis of the unexplained variability in the measurements. Systematic uncertainty is the experimenter's estimate of the systematic error and requires the exercise of judgment. Because the random uncertainty can be reduced simply by taking more observations, the overall uncertainty often comes to be dominated by the systematic uncertainty,[13] at least in those experiments where the marginal cost of observations is relatively low.

Some have suggested that random and systematic uncertainty are qualitatively different, and so should be reported separately,[14] a suggestion followed in some recent articles. However, Mueller argues that the categorization of uncertainties is context dependent, varying from one experiment to another, and that it is appropriate to combine them forming a single expression of uncertainty, which is far more useful for evaluating each measurement.[12] Acknowledging that the terms random and systematic uncertainty are often used in ambiguous and inconsistent ways the International Committee on Weights and Measures have recommended replacing them by Category A uncertainty, evaluated by statistical means, and Category B uncertainty, evaluated by other means, including "subjective appreciation."[15]

Whatever one calls them, it is the latter that cause the most problems. Known systematic errors can usually be compensated for by experimental control or computational corrections, and much of the skill of a precision metrologist lies in such issues of experimental design. For example, Youden has urged that key features of the experimental design[16] be varied systematically, thus converting some systematic errors to random errors, which can be estimated statistically from variations over observations.

Unfortunately, such treatment is not possible for all sources of error. Innovations in experimental technique are particularly likely to produce surprising new errors. In such cases, there is no substitute for judgment in assessing the magnitude of unresolved systematic errors. Some recent discussions of error estimation recognize this, e.g., warning that estimates of systematic error "are somewhat subjective and are usually obtained from what can only be called educated guesses."[17] Where reported uncertainties are intended to be comprehensive and are expressed as standard errors, confidence intervals, or other probabilistic forms, we conclude they are equivalent to assessments of the parameters of subjective probability distributions. Hence they may appropriately be analyzed for calibration and over- or underconfidence.

IV. MEASURES OF CALIBRATION

One can obtain some empirical insights into the calibration of a set of assessments by comparing reported uncertainties with the variability among the measurements. Suppose each experiment i, for $i = 1,2,...N$, reports an estimate x_i with standard error σ_i, and $\bar{x}$ is the group mean, weighted inversely by the variances. The sum of the squares of the normalized residuals, $h_i = (x_i - \bar{x})/\sigma_i$, should be distributed as the chi-squared statistic with $(N-1)$ degrees of freedom, assuming the errors are independent and normally distributed with the reported standard deviations. This statistic can then be used to test the appropriateness of these assumptions.

A related measure, the Birge ratio,[18] R_B, assesses the compatibility of a set of measurements by comparing the variability among experiments to the reported uncertainties. It may be defined as the standard deviation of the normalized residuals:

$$R_B^2 = \sum_i h_i^2/(N-1). \tag{1}$$

Alternatively, the Birge ratio may be seen as a measure of the appropriateness of the reported uncertainties. If the uncertainty assessments are independent and perfectly calibrated, the expectation of R_B is one. If R_B is much greater than one, then one or more of the experiments has underestimated its uncertainty and may contain unrecognized systematic errors. Such insensitivity to systematic error is a sign of overconfidence. If R_B is much less than one, then

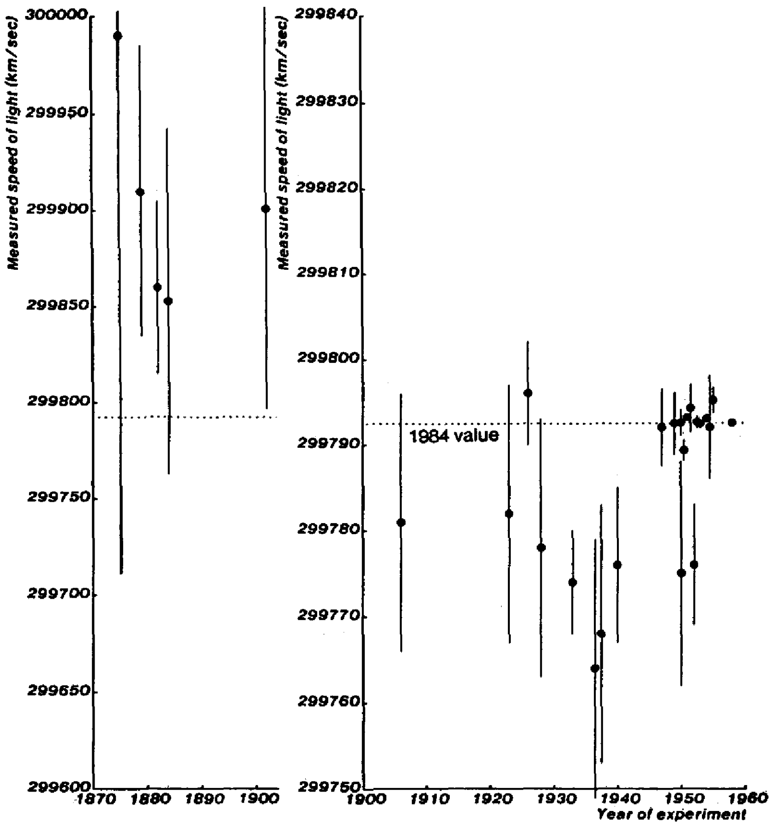

Fig. 1. Measurements of the velocity of light; 1875–1958. Results are as first reported, with correction from air to vacuum where needed. The uncertainties are also as originally reported, where available, or as estimated by the earliest reviewers. Error bars show standard error (s.e. = 1.48 × probable error).

either the uncertainties have, in the aggregate, been overestimated or the errors are correlated.

V. UNCERTAINTY IN *c*

The velocity of light, c, is perhaps the most measured fundamental physical constant,[1] starting with Galileo's unsuccessful attempts, using assistants flashing lanterns on neighboring hills. Figure 1 displays the results of all measurements between 1875 and 1958 with reported uncertainty from several major surveys.[13,17–20] The vertical bars represents the standard error according to the original experiment or earliest reviewer.[21] The horizontal dashed line is the 1984 value of 299 792.458 km/s.[22]

The Birge ratio for the entire set of 27 measurements with errors relative to the 1984 value is 1.42. The probability of finding such large variability by chance is less than 0.005, assuming that the error were normally distributed with the reported standard deviations. If the standard deviations of the discrepant studies fully express the respective investigators' uncertainties regarding their estimates, then, on average, those uncertainties must be significantly underestimated.

Over time, both the reported uncertainty and the actual error of these estimates have been reduced enormously. There was, however, no significant corresponding improvement in the Birge ratio, which was 1.47 for measurements up to 1941 and 1.32 since 1947 (top of Table I). For both periods, the variability was significantly greater than one would expect, indicating unduly tight uncertainty estimates.

These changes in accuracy over time were accompanied by changes in the direction of error. From 1976 to 1902, the measurements overestimated the value by about 70 km/s on the average. From 1905 to 1950, they underestimated it by about 15 km/s on the average. This change in the mean caused deBray to suggest that the speed of light was not constant but decreasing by about 4 km/s/year. In 1934, Edmondson proposed that it might be varying sinusoidally

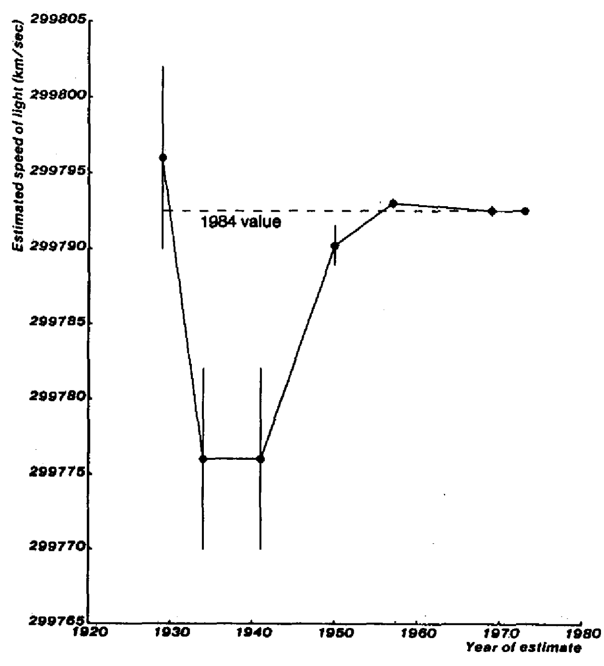

Fig. 2. Recommended values for the velocity of light; 1929–1973.

Table I. Calibration statistics for measurements of physical constants.

Quantity	Date	Statistic					Refs.[f]
		N[a]	R_B[b]	Pr[c]	IQ[d]	SI[e]	
c	1875–1941	13	1.47	0.01	23%	8%	13, 17–20
c	1947–1958	14	1.32	0.05	57%	14%	
c	1875–1958	27	1.42	0.005	41%	11%	
G	1798–1983	14	1.38	0.025	41%	29%	24
μ_p'/μ_n	1949–1967	7	1.44	0.05	14%	14%	17
α^{-1} High accuracy		24	2.95	(0)	21%	38%	17
α^{-1} Low accuracy		14	1.26	0.10	64%	7%	17
$\Omega_{ABS}/\Omega_{NBS}$	1938–1968	7	0.40	0.995	100%	0%	17
Particle lives		92	1.26			9%	5, 25
Particle masses		214	1.24			6%	
Particle total		306	1.24			7%	
$S_{10}(h/1.1)$			1.24		44%	6%	26
Recommended values	1928–1973	40	7.42	(0)	22%	57%	3,17 19, 28, 29
Well-calibrated normal distribution			1.00		50%	2%	

[a] N Number of measurements analyzed.
[b] R_B Birge ratio (standard deviation of normalized residuals).
[c] Pr Probability of getting Birge ratio that large by chance if the normalized residuals have independent unit normal distributions.
[d] IQ Interquartile index: Percent which fall between assessed 25%ile and 75%ile, or within 0.675σ.
[e] SI Surprise index: Percent that fall outside assessed 98% confidence interval, or outside 2.33σ.
[f] Refs. Sources of data.

with a period of 40 years. Discounting these claims, Birge[19] argued c has a constant value, that the overconfident uncertainty estimates too frequently failed to include. In 1941, he adjusted a set of recent estimates with untreated systematic error, producing corrected values with a Birge ratio of 0.544. Birge concluded, "Thus, after a long and, at times, hectic history, the value for c has at last settled down into fairly satisfactory 'steady' state."[19]

Unfortunately, as illustrated in Fig. 2, his confidence turned out to be premature. Just nine years later, the recommended value had shifted by 2.4 of his 1941 standard deviations.[2] This 1950 value, too, was soon supplanted, by a value different by over 2 of its standard deviations. Once again, shifting estimates prompted the suggestion that c might be changing, this time increasing.[20] One can only speculate whether better calibrated assessments of uncertainty (showing that the varied estimates were compatible with one another) would have forestalled such paradigm-shattering proposals. Table I summarizes these data in terms of various calibration statistics, including the Birge ratio, interquartile index, and surprise index. All indicate overconfidence. For measurements of c, there are too few values between the 0.25 and 0.75 quartiles (41% instead of 50%), and too many surprises (11% instead of 2%).

VI. UNCERTAINTY IN OTHER CONSTANTS AND RECOMMENDED VALUES

The remainder of Table I summarizes similar analyses for other physical constants, including the gravitational constant G, the magnetic moment of the proton μ_p'/μ_n' two sets of measures of the inverse fine structure constant, α^{-1}, and the ratio of the absolute ohm to the ohm maintained by the National Bureau of Standards, $\Omega_{ABS}/\Omega_{NBS}$. Table I reveals a similar pattern of overconfidence in all cases but the last. This shows a Birge ratio of significantly less than 1.0 and an interquartile ratio of 100%, results we shall return to later.

The "particle lives" and "particles masses" entries in Table I represent 92 measurements of Kaon and hyperon lifetimes and 214 measurements of meson resonance masses taken from the Particle Data Group's biennial *Review of Particle Properties*.[5,25] Their thorough scrutiny of all published experimental measurements results in the rejection of about 40% on grounds of suspect assumptions, poor quality work, unreported errors, or gross inconsistency with other results. Nonetheless, the Birge ratio of 1.24 and surprise index of 6% of the 306 combined results indicate a significant degree of overconfidence in the re-

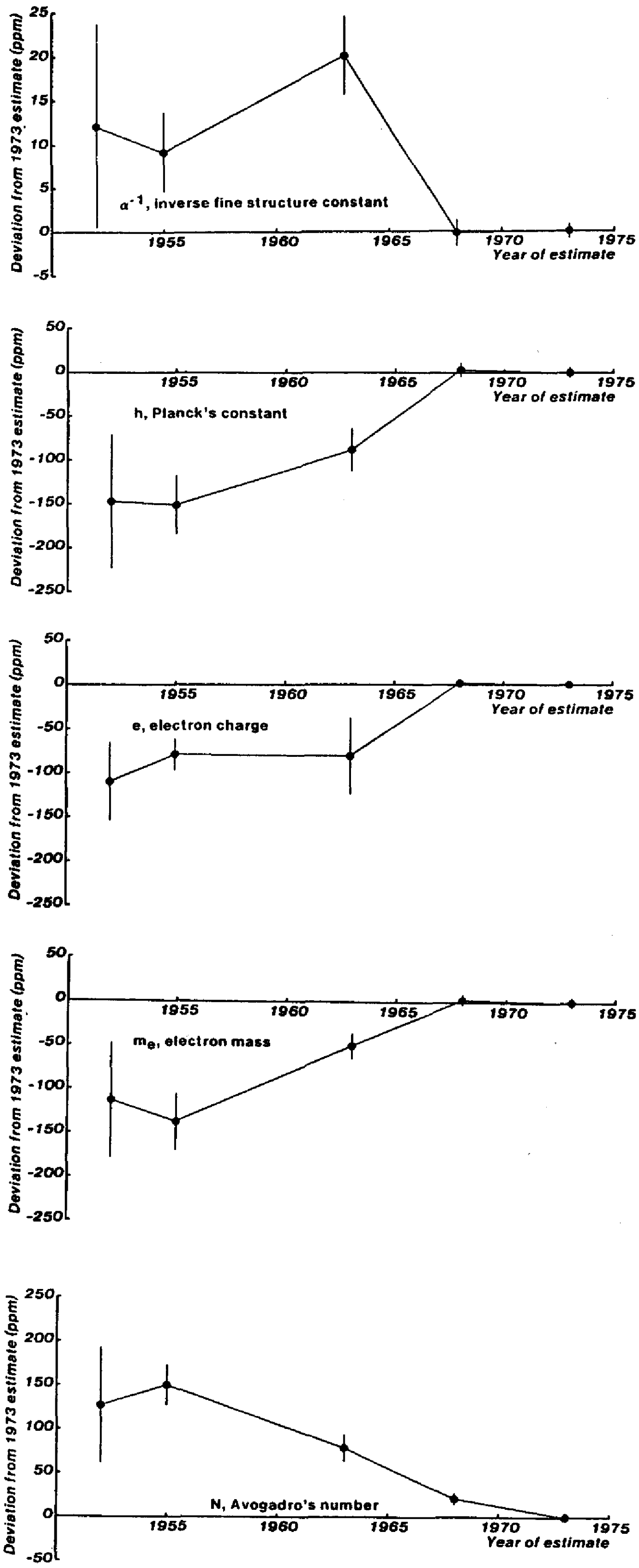

Fig. 3. Recommended values for fundamental constants; 1952–1973.

maining uncertainty assessments.

Periodically, reviewers propose "recommended values" for physical constants after careful consideration of all the most precise measurements. The diligence of these analyses and the incumbent obligation to confront the inconsistencies among studies should make them particularly sensitive to uncertainty. Figure 2 showed the residual overconfidence in such reviews for the velocity of light. Figure 3 shows comparable results for five other fundamental constants derived from a series of six reviews between 1952 and 1973.[3,17,28] In most cases, the best-guess value at each revision was well outside the range of possibilities defined by the uncertainty estimates for the preceding evaluation period. For example, the 1963 to 1969 shift was three to five 1963 standard deviations for all five of these constants. Table I (second to last line) gives calibration indices for these same five constants for eight reviews from 1929 to 1969,[3,17,19,28,29] with residuals calculated relative to 1973 estimates. The surprise index of 57% indicates that current estimates would have come as a surprise to earlier reviewers in more than half the cases. The 1969 review acknowledges these problems, and provides graphs similar to Fig. 3 to caution readers to be skeptical.[17]

VII. SOURCES OF OVERCONFIDENCE

In several sets of carefully analyzed measurements of physical constants, we have found consistent replication of a robust finding of laboratory studies of human judgement: Reported uncertainties are too small. How could this apparent overconfidence arise? Experimental studies of human judgment have shown that such biases can arise quite unintentionally from cognitive strategies employed in processing uncertain information. However, there are two possible ways in which such effects might be caused by deliberate decisions by the scientists:

One concerns the procedures chosen to assess the uncertainty. The recommended practice in physics is to consider all possible sources of systematic uncertainty when reporting results. However, without specific guidelines regarding what to consider and explicit recognition of the subjective elements in uncertainty assessments, one cannot be sure how comprehensively individual scientists have examined the uncertainty surrounding their own experiments. Conceivably, some of the apparent overconfidence reflects a deliberate decision to ignore the harder-to-assess sources of uncertainty.

A second possible source of bias is that, unlike laboratory experiments on judgment, which can take great care to ensure that subjects are motivated to express their uncertainty candidly, real-world settings create other pressures. Taylor *et al.* suggest that variations in attitude leads some investigators to

> ...cautiously assign unreasonably large errors so that a later measurement will not prove their work to have been "incorrect." Others tend to underestimate the sources of systematic error in their experiments, perhaps because of an unconscious (or conscious) desire "to have done the best experiments." Such variation in attitude, although out of keeping with scientific objectivity, is nevertheless unavoidable so long as scientists are also human beings.[17]

In principle, none of these problems should affect the compilations of recommended values. These analyses are intended to be comprehensive, to consider reporting practices in the field, and to capture the state of knowledge (not just the precision of particular studies).

The factors that seem to have encouraged overconfidence in laboratory studies of judgement also seem to be likely candidates for having encouraged overconfidence in the estimation of physical constants. One such factor is the difficulty of thinking of reasons why one's best guess might be wrong.[31] Supporting reasons typically come to mind more readily than contradicting ones. If ease of recollection is taken as an indicator of frequency,[32] the preponderance of reasons supporting the best guess will tend to be overestimated.

A second factor contributing to overconfidence is the unequal treatment of such confirming and disconfirming evidence as is discovered. When discrepant results are encountered, it could mean that either the new data or the old database contain undetected systematic errors. Unfortunately, people have a considerable ability to "explain away" events that are inconsistent with their prior beliefs.[33] The data elimination and adjustment procedures that precede estimation of recommended values are natural places for disproportionate skepticism regarding unexpected results to emerge. One documented example of "trimming," i.e., excessive zeal in eliminating outliers, emerged in a recent examination of Millikan's laboratory notebooks for his oil-drop experiment. Of the 107 oil drops he observed, he excluded from publication 49 which seemed discrepant, despite his claim that he reported all his observations.[34] This had little effect on the best estimate of e, but increased the apparent precision of the experiment.

Having a preexisting recommended value may particularly encourage investigators to discard or adjust unexpected results, and so induce correlated errors in apparently independent experiments. The result would be initially small Birge ratios, unduly tight confidence intervals, and the exclusion of discrepant data that later prove to be more accurate than included ones. This can explain what Franklin has termed the "bandwagon effect,"[4] e.g., the tendency for the measurements of the speed of light (in Fig. 1) to cluster around particular values at different periods. The large Birge ratio (2.95) for high accuracy measurements of α^{-1} (Table I) may be due to the fact that those measurements were actually derived from experiments designed to measure other quantities, so that discrepancies in the implied value of α^{-1} would have been less obvious to the original investigators. On the other hand, measurements of the ratio of the absolute ohm to the as-maintained ohm ($\Omega_{ABS}/\Omega_{NBS}$) have a natural anchor at unity (i.e., assuming they are equal), which could explain the unusual degree of agreement, evidenced by the Birge ratio of 0.40.

VIII. IMPROVING UNCERTAINTY ASSESSMENT

In order to reduce the need for judgment in selecting data for producing recommended values, several attempts have been made to develop mathematical procedures for selecting and averaging measurements. Statisticians have long sought more robust estimators, such as trimmed means, that use all useful information without being unduly sensitive to outliers. However, in tests with real data, more sophisticated estimators have provided only marginal improvements over simple averages.[35] Roos *et al.*[5] describe a maximum likelihood technique based on their initial observation that the adjusted residual errors for particle properties had a broadtailed Student distribution. This was used in the *Review of Particle Properties* from 1976 to 1980, but was dropped after subsequent distributions of residuals differed in form.[27] There have also been attempts to develop algorithms for estimating the fundamental physical constants that deal better with apparently discrepant data. However, after tests with historical data, Taylor concluded that thoughtful and conscientious judgment about which data to exclude is still much more important than the choice of algorithm.[30]

It is no easy matter to eliminate judgmental biases.[9] A categorical recommendation to treat all results equally would, on the average, give undue credence to inaccurate observations and bad research. Nor is simple exhortation to "think harder" likely to prove any more useful here than it has in laboratory experiments on judgment. What are needed are ways to think more effectively. If the judgmental aspect of assessing uncertainty in physical measurements is explicitly acknowledged, then several techniques based on insights from psychological research[36] may be helpful.

One approach is to encourage a broader search for relevant considerations, both supportive and disconfirming, and an unbiased evaluation of those considerations. For example, referees might be asked to scrutinize the account of an experiment "blind," before knowing its outcome and whether it affirmed prior expectations.[37] To overcome the problem of anchoring on an initial best guess, investigators can be pressed to search for reasons why they might be wrong.[31] One device to encourage that search is to focus attention on extreme possibilities, for example, asking "imagine that, ten years hence, today's best estimate proves to be off by four standard deviations; how could you explain it?" Additional prompts may come from reading "horror stories," case studies describing major unsuspected errors in past experiments.

A second approach is to decompose the holistic judgment of overall uncertainty into its component sources of potential systematic errors, each to be estimated separately with the best available elicitation procedure.[36] This can be aided by checklists covering each component of the measurement apparatus, each auxiliary quantity, and each theoretical assumption and approximation employed in the calculations.

Wherever analysis of reported uncertainties reveals a systematic bias, users of the measurements may use this information to *readjust* the original reports so as to improve calibration. Along these lines, the Particle Data Group expands the standard deviation whenever the Birge ratio is greater than unity among measurements of the same quantity.[25] Their procedure usually produces stable recommended values for particle properties; however, it somewhat exaggerates the uncertainty by ignoring the possibility that some fluctuations of R_B above one are purely by chance. Roos *et al.*[26] describe another procedure for readjusting the standard errors, based on the empirical distribution of the residuals thay had observed for the particle properties.

The basis for such readjustments must be secondary analyses, such as those given here, showing probabilistic miscalibration in the estimates of uncertainty for previous similar measurements. In using such data, it is important to bear in mind that miscalibration reflects, not properties of the physical quantities themselves, but the nature of the procedures and judgmental processes used in selecting data and assessing uncertainty. Thus they are likely to vary with the training of the experimenter, the familiarity of the experimental techniques, and the maturity of the field, among other things. Compiling such data for wider classes of measurements can permit analysis of the importance of these factors. Even if uncertainty estimates are not explicitly readjusted, empirical calibration information could be helpful for anyone who uses measurements and recommended values for physical constants in their work. Even a rough estimate of overconfidence could help one interpret the significance of apparent discrepancies between measurements.

Another approach to improving uncertainty assessment is education designed to get better estimates in the first place. For example, in teaching experimental methods, greater exposure to the kind of results presented here should be helpful, together with attention to the role of judgment in the assessment process. Although one might hope that instructions in the processes and pitfalls of judgment would, by themselves, improve performance, the evidence to date suggests that people have difficulty in integrating an understanding of general principles with their own cognitive processes.[38] What *has* proved more effective in other domains is task-specific training with personal feedback.[9] For example, in laboratory classes, where students are required to measure the same quantities, it should be instructive to compile distributions of normalized residuals and measures of calibration for the class results, and to discuss together the reasons both for individual errors and for systematic miscalibration in reported uncertainties.

IX. CONCLUSIONS

The underestimation of uncertainty in measurements of physical constants and compilations of recommended values seems to be pervasive. This evidence extends previous findings of overconfidence in laboratory studies of human judgment to a task domain of great practical importance. If reported uncertainties do not reflect the magnitude of actual errors, whether due to incomplete analysis or to judgmental biases, the usefulness of those measurements is significantly diminished. Measurements are then hard to compare, and are unlikely to produce "enduring values,"[16] as we have illustrated in the repeated contradiction of accepted values by subseqent measurements. Recognizing that subjective judgment is an essential element in the assessment of systematic uncertainty enables us to use findings from cognitive research to help understand how these biases arise, and to suggest approaches for dealing with these problems more effectively.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation under Grant No. SES-8213452 to Perceptronics, Inc. Any opinions or findings expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are grateful to George Sparling for helpful suggestions.

[1]F. M. Pipkin and R. C. Ritter, Science **219**, 913 (1983).

[2]C. Eisenhardt, Science **160**, 1201 (1968).

[3]E. R. Cohen, K. M. Crowe, and J. W. M. Dumond, *Fundamental Constants of Physics* (Interscience, New York, 1957).

[4]A. D. Franklin, Am. J. Phys. (to be published).

[5]M. Roos, M. Hietanen, and J. Luoma, Phys. Fenn. **10**, 20 (1975).

[6]A. Tversky and D. Kahneman, Science **211**, 453 (1981); *Judgment Under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky (Cambridge, New York, 1982).

[7]B. Fischhoff and C. Whipple, Environmental Professional **3**, 277 (1981); M. G. Morgan, M. Henrion, and S. C. Morris, BNL No. 51358, (Brookhaven National Laboratory, 1981).

[8]*Probability, Statistics and Decision-Making in the Atmospheric Sciences*, edited by A. H. Murphy and R. W. Katz (Westview, Boulder, CO., 1983).

[9]S. Lichtenstein, B. Fischhoff, and L. D. Phillips, in *Judgment Under Uncertainty: Heuristic and Biases* (Cambridge, New York, 1982); B. Fischhoff, *ibid.*

[10]See, e.g., H. Raiffa, *Decision Analysis* (Addison–Wesley, Reading, MA, 1968); A. E. Green and A. J. Bourne, *Reliability Technology* (Wiley, New York, 1972); U.S. Nuclear Regulatory Commission, *Fault Tree Handbook* (NUREG-0492, NRC, Washington, D.C., 1981); National Academy of Sciences, *Priority Mechanisms for Toxic Chemicals* (National Academy, Washington, DC, 1983).

[11]See, e.g., M. Hynes and E. Vanmarcke, in Proceedings of the ASCE Engineering Mechanics Division Specialty Conference, Waterloo, Ontario, Canada (1976).

[12]See, e.g., J. W. Mueller, Nucl. Instrum. Methods **163**, 241 (1979).

[13]K. D. Froome and L. Essen, *The Velocity of Light and Radio Waves* (Academic, London, 1969).

[14]P. J. Campion, J. E. Burns, and A. Williams, *A Code of Practice for the Detailed Statement of Accuracy* (National Physics Laboratory, H.M.S.O., London, 1973).

[15]Recommendation (CI-1981) adopted by Le Comité International des Poids et Mesures, 1982.

[16]W. J. Youden, Technometrics **1**, 1 (1972).

[17]B. N. Taylor, W.H. Parker, and D. N. Langenberg, *The Fundamental Constants and Quantum Electrodynamics* (Academic, New York, 1969).

[18]R. T. Birge, Phys. Rev. **40**, 207 (1932).

[19]R. T. Birge, Rep. Prog. Phys. **8**, 90 (1942).

[20]J. H. Rush, Sci. Am. August, 62-67, (1955).

[21]The mean values are those originally reported, adjusted from velocity in air to velocity in a vacuum, where necessary. Where uncertainty was reported as "probable error" it was converted to standard deviation using a factor of 1.48.

[22]This value has been adopted by the General Conference on Weights and Measures as the *definition* of c; c is now used in defining the meter instead of vice versa.

[23]J. W. M. Dumond and E. R.Cohen, Phys. Rev. **82**, 555 (1951).

[24]From 14 measurements of G, from Pipkin and Ritter (Ref. 1) and G. T. Gillies, *The Newtonian Gravitational Constant: An Index of Measurements* (BIPM-83/1), Bureau International des Poids et Mesures, 1983), omitting three that were off by more than ten standard deviations from the current recommended value.

[25]A. H. Rosenfeld, Annu. Rev. Nucl. Sci. **25**, 555 (1975); Particle Data Group: M. Roos, F. C. Porter, *et al.*, Phys. Lett. B **111**, 1 (1982).

[26]Roos *et al.* (Ref. 5) fitted the normalized residuals, h_i, for the particle lives and masses to an adjusted Student distribution, $s_N(h/c)$, with $N = 10$ degrees of freedom, and an adjustment factor, $c = 1.11$. The calibration indices for this distribution are shown in Table I for comparison.

[27]T. Trippe, 1983 (personal communication).

[28]J. W. M. Dumond and E. R. Cohen, Rev. Mod. Phys. **25**, 691 (1953); E. R. Cohen *et al. ibid.* **27**, 363 (1955); E. R. Cohen and J. W. M. Dumond, *ibid.* **37**, 537 (1965); E. R. Cohen and B. N. Taylor, J. Phys. Chem Ref. Data **2**, 663 (1973).

[29]R. T. Birge, Rev. Mod. Phys. **1**, 1 (1929).

[30]B. N. Taylor, NBSIR 81-2426, National Bureau of Standards (1982).

[31]A. Koriat, S. Lichtenstein, and B. Fischhoff, J. Exp. Pysch.: Hum. Learn. Mem. **6**, 107 (1980).

[32]A. Tversky and D. Kahneman, Cognitive Psychology **4**, 207 (1973).

[33]R. E. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Printice–Hall, Englewood Cliffs, NJ, 1980).

[34]A. D. Franklin, Hist. Studies Phys. Sci. **11**, 2 (1981).

[35]S. M. Stigler, Ann. Stats. 5,6 1055 (1977).

[36]See notes (Refs. 7 and 8) and M. G. Morgan *et al.*, Risk Analysis 4, 4 (1984); C. S. Spetzler and C.-A. S. Stael von Holstein, Man. Sci. 22 (1975).

[37]P. Slovic and B. Fischhoff, J. Exp. Psych.: Hum. Percep. Perf. 3, 544 (1977).

[38]See, e.g., G. J. Gaeth and J. Shanteau, Org. Beh. Hum. Perf. 33, 263 (1984).

Inverse sprinklers: A lesson in the use of a conservation principle

A. Theodore Forrester

University of California, Los Angeles, Los Angeles, California 90024

(Received 10 June 1985; accepted for publication 3 September 1985)

When a common lawn sprinkler is hooked to a suction, rather than a pressure source the conservation of angular momentum can be invoked to show that no rotation will result. A recently reported experiment leading to a different conclusion is not the equivalent of simply changing from a fluid exhaust to a fluid intake.

I. THE PROBLEM AND A FORCE ANALYSIS

In a recent book, Feynman[1] discusses an interesting elementary physics problem and reaches a conclusion which he claims to have verified in an experimental test. His test was short-lived, ending in a self-destruct of the apparatus, but not before he felt his contention was verified. I shall return to a description of his experiment later.

The elementary physics problem referred to is the inverse lawn sprinkler problem. A lawn sprinkler of the type referred to is shown in Fig. 1. When water flows through in the usual direction, as shown in Fig. 1, the sprinkler head will rotate in the indicated direction. If the fluid pumped through the sprinkler were air rather than water flowing in the same direction, it is clear that the head would also rotate as shown. The question for which the answer is not quite so clear, concerns the direction of rotation when air (or water) is sucked in through the sprinkler.

For simplicity, examine the forces acting on a nozzle with the help of Fig. 2. The normal flow situation is shown in Fig. 2(a). As the fluid is bent around the curve it does so because the metal tube exerts an inward centripetal force on the fluid. The reaction to this is in the outward direction and has components away from the axis, which we can ignore, and a component F_c in the direction shown, which is responsible for the torque. Often neglected in these considerations is the additional term due to the fact that the pressure inside the tube is greater than that outside producing an additional force $F_p = A\Delta p$, where A is the aperture area. Since these forces are in the same direction there is little question as to the direction of the resultant force. The torque $r \times F$ on the sprinkler head will be out of the plane of the figure leading to a counterclockwise rotation, as viewed from above.

If the air is sucked in, then the force on the tube due to the fluid bending the corner F_c is in the same direction as before, but the force F_p is reversed and the direction of rotation no longer appears unambiguous.

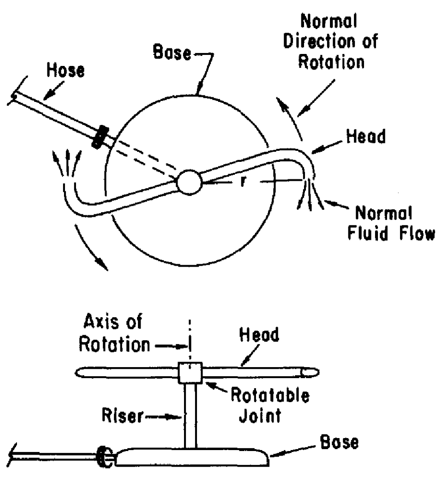

Fig. 1. A common lawn sprinkler.

II. CONSERVATION OF ANGULAR MOMENTUM

The detailed analysis given above could, in principle, lead to the actual torque on the head in terms of mass flow rate given sufficient details concerning the fluid, such as viscosity and density as a function of pressure, as well as the geometric details of the sprinkler—but it would be hard.

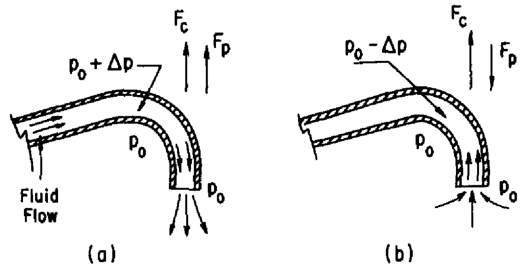

Fig. 2. Examination of forces F_c and F_p on a nozzle during (a) fluid exhaust and (b) fluid intake.