On the meaning of uncertainty of measurement

Fabien Grégis

Laboratoire SPHERE, Université Paris Cité

PREN Convention
June 20, 2022, Sorbonne Université

Journal of Physical and Chemical Reference Data ARTICLE scitation.org/journal/jpr

TABLE XXXI. The CODATA recommended values of the fundamental constants of physics and chemistry based on the 2018 adjustment

Quantity	Symbol	Numerical value	Unit	Relative std. uncert. u_r
		UNIVERSAL		
speed of light in vacuum	c	299 792 458	m s^{-1}	exact
vacuum magnetic permeability $4\pi\alpha\hbar/e^2c$	μ_0	$1.256\,637\,062\,12(19)\times 10^{-6}$	N A^{-2}	1.5×10^{-10}
$\mu_0/(4\pi\times 10^{-7})$		1.000 000 000 55(15)	N A^{-2}	1.5×10^{-10}
vacuum electric permittivity $1/\mu_0 c^2$	ε_0	$8.854\,187\,8128(13)\times 10^{-12}$	F m^{-1}	1.5×10^{-10}
characteristic impedance of vacuum $\mu_0 c$	Z_0	376.730 313 668(57)	Ω	1.5×10^{-10}
Newtonian constant of gravitation	G	$6.674\,30(15)\times 10^{-11}$	m^3 kg^{-1} s^{-2}	2.2×10^{-5}
	$G/\hbar c$	$6.708\,83(15)\times 10^{-39}$	$(\text{GeV}/c^2)^{-2}$	2.2×10^{-5}
Planck constant[a]	h	$6.626\,070\,15\times 10^{-34}$	J Hz^{-1}	exact
⋮	⋮	⋮	⋮	⋮
fine-structure constant $e^2/4\pi\varepsilon_0\hbar c$	α	$7.297\,352\,5693(11)\times 10^{-3}$		1.5×10^{-10}
inverse fine-structure constant	α^{-1}	137.035 999 084(21)		1.5×10^{-10}
Rydberg frequency $\alpha^2 m_e c^2/2h = E_h/2h$	cR_∞	$3.289\,841\,960\,2508(64)\times 10^{15}$	Hz	1.9×10^{-12}
energy equivalent	hcR_∞	$2.179\,872\,361\,1035(42)\times 10^{-18}$	J	1.9×10^{-12}
		13.605 693 122 994(26)	eV	1.9×10^{-12}
Rydberg constant	R_∞	10 973 731.568 160(21)	[m^{-1}][b]	1.9×10^{-12}
⋮	⋮	⋮	⋮	⋮

J. Phys. Chem. Ref. Data **50,** 033105 (2021); doi: 10.1063/5.0064853 **50,** 033105-46

U.S. Secretary of Commerce.

Journal of Physical and Chemical Reference Data ARTICLE scitation.org/journal/jpr

TABLE XXXI. The CODATA recommended values of the fundamental constants of physics and chemistry based on the 2018 adjustment

Quantity	Symbol	Numerical value	Unit	Relative std. uncert. u_r
		UNIVERSAL		
speed of light in vacuum	c	299 792 458	m s^{-1}	exact
vacuum magnetic permeability $4\pi\alpha\hbar/e^2c$	μ_0	$1.256\,637\,062\,12(19) \times 10^{-6}$	N A^{-2}	1.5×10^{-10}
$\mu_0/(4\pi \times 10^{-7})$		1.000 000 000 55(15)	N A^{-2}	1.5×10^{-10}
vacuum electric permittivity $1/\mu_0 c^2$	ε_0	$8.854\,187\,8128(13) \times 10^{-12}$	F m^{-1}	1.5×10^{-10}
characteristic impedance of vacuum $\mu_0 c$	Z_0	376.730 313 668(57)	Ω	1.5×10^{-10}
Newtonian constant of gravitation	G	$6.674\,30(15) \times 10^{-11}$	m^3 kg^{-1} s^{-2}	2.2×10^{-5}
	$G/\hbar c$	$6.708\,83(15) \times 10^{-39}$	$(\text{GeV}/c^2)^{-2}$	2.2×10^{-5}
Planck constant[a]	h	$6.626\,070\,15 \times 10^{-34}$	J Hz^{-1}	exact
⋮	⋮	⋮	⋮	⋮
fine-structure constant $e^2/4\pi\varepsilon_0\hbar c$	α	$7.297\,352\,5693(11) \times 10^{-3}$		1.5×10^{-10}
inverse fine-structure constant	α^{-1}	137.035 999 084(21)		1.5×10^{-10}
Rydberg frequency $\alpha^2 m_e c^2/2h = E_h/2h$	cR_∞	$3.289\,841\,960\,2508(64) \times 10^{15}$	Hz	1.9×10^{-12}
energy equivalent	hcR_∞	$2.179\,872\,361\,1035(42) \times 10^{-18}$	J	1.9×10^{-12}
		13.605 693 122 994(26)	eV	1.9×10^{-12}
Rydberg constant	R_∞	10 973 731.568 160(21)	[m^{-1}][b]	1.9×10^{-12}
⋮	⋮	⋮	⋮	⋮

U.S. Secretary of Commerce.

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$

What are we reading?
What does it mean?

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$

“measurement uncertainty”

Measurement uncertainties

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$

"measurement uncertainty"

accuracy

error

Measurement uncertainties

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$

“measurement uncertainty”

accuracy

error

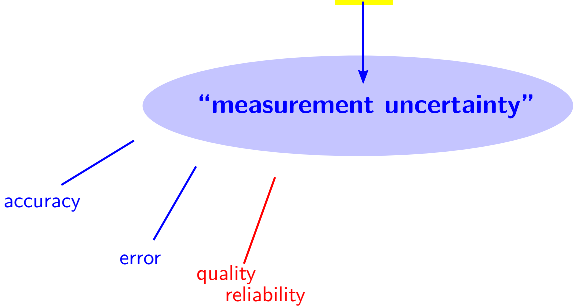
International Vocabulary of Metrology **(VIM), 1st edition, 1984**

Uncertainty of measurement : an estimate characterizing the range of values within which the true value of a measurand lies.

(p.16)

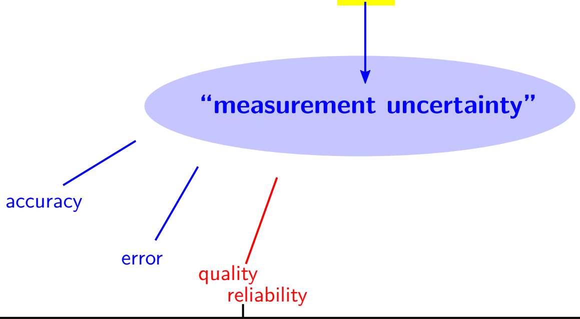
$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$



$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$



Guide to the expression of uncertainty in measurement (GUM), 1993

When reporting the result of a measurement of a physical quantity, it is obligatory that some quantitative indication of the quality of the result be given so that those who use it can assess its reliability. (p.vii)

Introductory supplement to the GUM (GUM supplement 4), 2009

The dispersion of the indication values would relate to how well the measurement is made. (p.2)

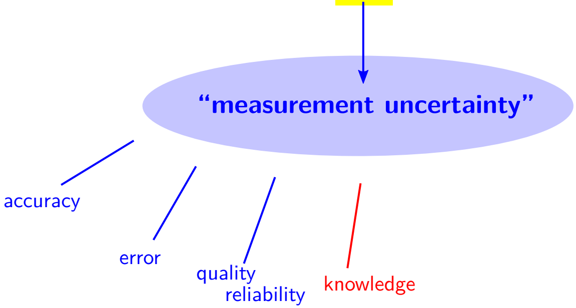
S. Bell, _A Beginner's Guide to Uncertainty of Measurement_, NPL 1999

The uncertainty of a measurement tells us something about its quality. (p.1)

Measurement uncertainties

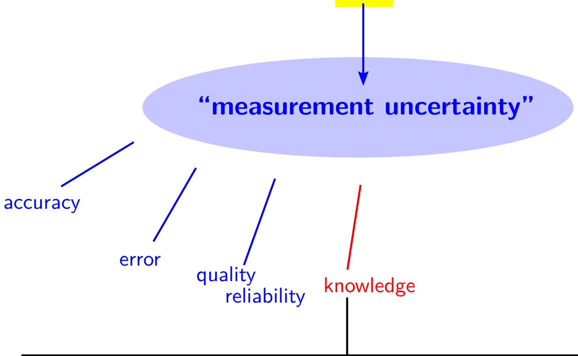
$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m}^3.\mathrm{kg}^{-1}.\mathrm{s}^{-2}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m}^{-1}$$



$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$

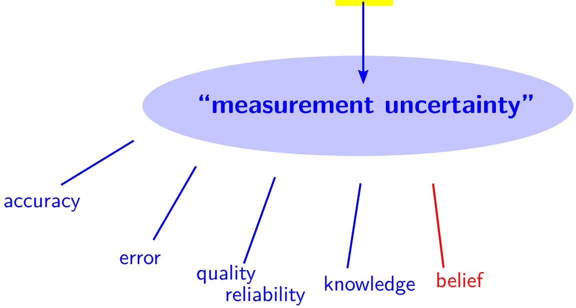


W. Bich, *From Errors to Probability Density Functions*, 2012

Uncertainty of measurement can be viewed as the logical reciprocal of state of knowledge. (p.2155)

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

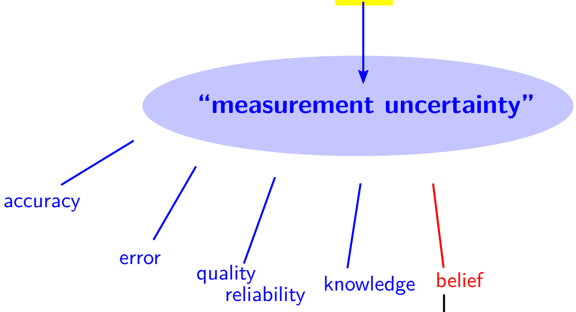
$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$



Measurement uncertainties

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$



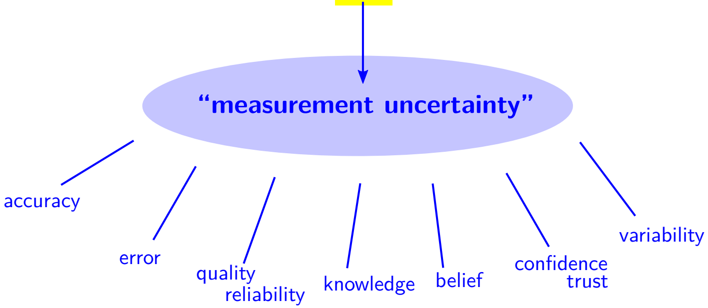
Introductory supplement to the GUM (GUM supplement 4), 2009

Measurement uncertainty can (...) be described as a measure of how well one believes one knows the essentially unique true value of the measurand. (p.3)

Measurement uncertainties

$$G = 6.674\,30\,(15) \times 10^{-11}\ \mathrm{m^3.kg^{-1}.s^{-2}}$$

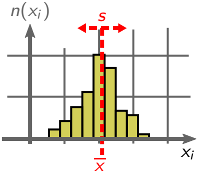
$$R_\infty = 10\,973\,731.568\,160\,(21)\ \mathrm{m^{-1}}$$



1– Statistical models in metrology

Statistical uncertainties

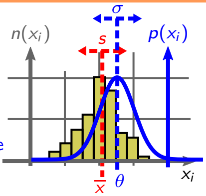
Experimental data: sample $\{x_1, \ldots, x_n\}$



Statistical uncertainties

Experimental data: sample $\{x_1, \ldots, x_n\}$

Random variable



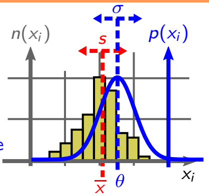
Statistical inference $(\overline{x}, s) \rightarrow (\theta, \sigma)$

Standard uncertainty $u(X) = \sqrt{s^2/n}$

Statistical uncertainties

Experimental data: sample $\{x_1, \ldots, x_n\}$

Random variable



Statistical inference $(\overline{x}, s) \to (\theta, \sigma)$

Standard uncertainty $u(X) = \sqrt{s^2/n}$

Probabilistic model: **frequencies**

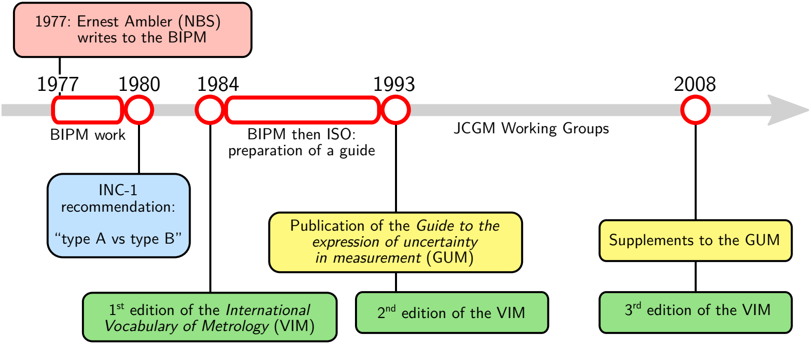
- $p(x_i)$ is a probability that characterizes the physical setup
- Relative occurrence frequency of possible measurement outcome

1970s: what about systematic errors?
→ If they induce an offset then it cannot be captured by frequentist probabilities
(no frequency ⇒ no probability)
⇒ Observation: **heterogeneity** of practices in the 1970s

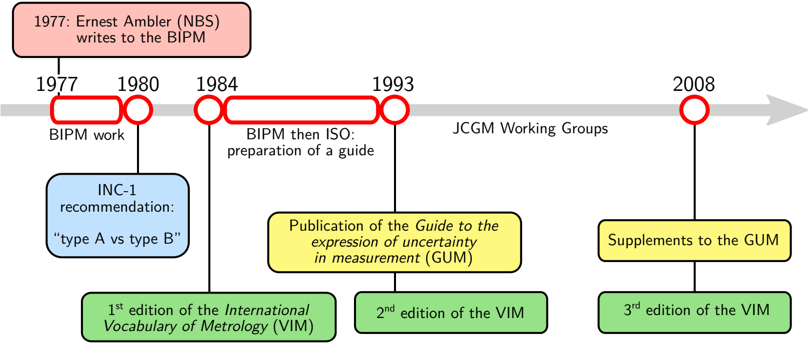
1970s: what about systematic errors?
$\rightarrow$ If they induce an offset then it cannot be captured by frequentist probabilities
(no frequency $\Rightarrow$ no probability)
$\Rightarrow$ Observation: **heterogeneity** of practices in the 1970s

1977: Ernest Ambler (NBS)
writes to the BIPM

1970s: what about systematic errors?
→ If they induce an offset then it cannot be captured by frequentist probabilities
(no frequency ⇒ no probability)
⇒ Observation: **heterogeneity** of practices in the 1970s



1970s: what about systematic errors?
→ If they induce an offset then it cannot be captured by frequentist probabilities (no frequency ⇒ no probability)
⇒ Observation: **heterogeneity** of practices in the 1970s



Cristallizes after the 70s around the interpretation of probabilities

→ type B methods imply **another interpretation of probabilities**
→ underlies debate on frequentism vs Bayesianism in metrological statistics
→ + some attempts to provide integrally Bayesian accounts of measurement uncertainty

JCGM 100:2008

E.3.5 In the traditional terminology, the third term on the right-hand side of Equation (E.6) is called a "random" contribution to the estimated variance $s^2(z)$ because it normally decreases as the number of observations n increases, while the first two terms are called "systematic" contributions because they do not depend on n.

Of more significance, in some traditional treatments of measurement uncertainty, Equation (E.6) is questioned because no distinction is made between uncertainties arising from systematic effects and those arising from random effects. In particular, combining variances obtained from a priori probability distributions with those obtained from frequency-based distributions is deprecated because the concept of probability is considered to be applicable *only* to events that can be repeated a large number of times under essentially the same conditions, with the probability p of an event ($0 \leqslant p \leqslant 1$) indicating the *relative frequency* with which the event will occur.

In contrast to this frequency-based point of view of probability, an equally valid viewpoint is that probability is a measure of the *degree of belief* that an event will occur [13, 14]. For example, suppose one has a chance of winning a small sum of money D and one is a rational bettor. One's degree of belief in event A occurring is $p = 0{,}5$ if one is indifferent to these two betting choices:

1) receiving D if event A occurs but nothing if it does not occur;

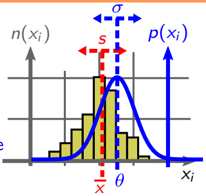
2) receiving D if event A does not occur but nothing if it does occur.

Recommendation INC-1 (1980) upon which this *Guide* rests implicitly adopts such a viewpoint of probability since it views expressions such as Equation (E.6) as the appropriate way to calculate the combined standard uncertainty of a result of a measurement

⋮

Type A

Experimental data: sample $\{x_1, \ldots, x_n\}$



Statistical inference $(\overline{x}, s) \rightarrow (\theta, \sigma)$

Standard uncertainty $u(X) = \sqrt{s^2/n}$

Probabilistic model: **frequencies**

- $p(x_i)$ is a probability that characterizes the physical setup
- Relative occurrence frequency of possible measurement outcome

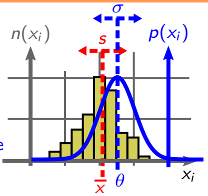
Type B

$\neq$ frequentist model: no random process
distribution **constructed** out of available info

Type A

Experimental data: sample $\{x_1, \ldots, x_n\}$

Random variable



Statistical inference $(\bar{x}, s) \rightarrow (\theta, \sigma)$

Standard uncertainty $u(X) = \sqrt{s^2/n}$

Probabilistic model: **frequencies**

- $p(x_i)$ is a probability that characterizes the physical setup
- Relative occurrence frequency of possible measurement outcome

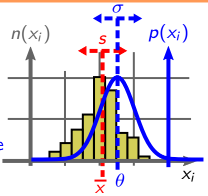
Type B

$\neq$ frequentist model: no random process
distribution **constructed** out of available info

ex. *measurement with a graduater ruler*
If information is: $5.a < x < 6.a$
where a is the graduation step

Type A

Experimental data: sample $\{x_1, \ldots, x_n\}$



Random variable

Statistical inference $(\overline{x}, s) \rightarrow (\theta, \sigma)$

Standard uncertainty $u(X) = \sqrt{s^2/n}$

Probabilistic model: **frequencies**

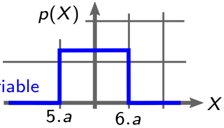
- $p(x_i)$ is a probability that characterizes the physical setup
- Relative occurrence frequency of possible measurement outcome

Type B

$\neq$ frequentist model: no random process
distribution **constructed** out of available info

ex. *measurement with a graduater ruler*
If information is: $5.a < x < 6.a$
where a is the graduation step

$\Rightarrow$ knowledge modeled as a rect. distribution



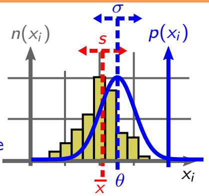
Random variable

Std uncertainty $\hat{=}$ std deviation
Here: $u = a/\sqrt{12}$

Type A

Experimental data: sample $\{x_1, \ldots, x_n\}$

Random variable



Statistical inference $(\overline{x}, s) \rightarrow (\theta, \sigma)$

Standard uncertainty $u(X) = \sqrt{s^2/n}$

Probabilistic model: **frequencies**

- $p(x_i)$ is a probability that characterizes the physical setup
- Relative occurrence frequency of possible measurement outcome

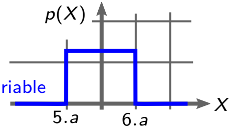
Type B

$\neq$ frequentist model: no random process
distribution **constructed** out of available info

ex. *measurement with a graduater ruler*
If information is: $5.a < x < 6.a$
where a is the graduation step

$\Rightarrow$ knowledge modeled as a rect. distribution

Random variable



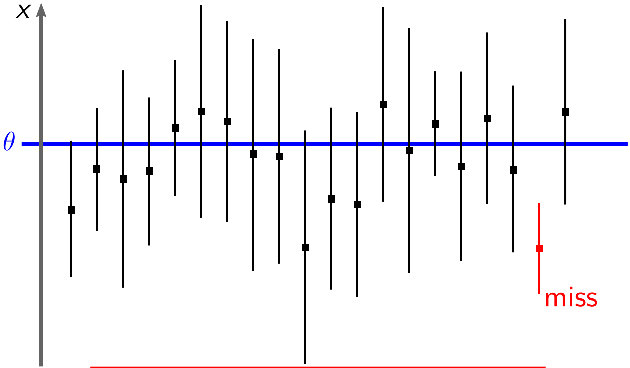
Std uncertainty $\hat{=}$ std deviation
Here: $u = a/\sqrt{12}$

The x-axis and y-axis are different!

- Probabilities are epistemic: $p(X)$ is a **degree of belief/state of knowledge**
- About a constant parameter (the value of the measured quantity)

Result expressed as interval $[x - U(x); x + u(X)]$ **with probability** 95%
$\rightarrow$ *What is this probability?*

Frequentist confidence interval: refers to counterfactual situations
$\rightarrow$ the interval *does or does not* contain the true value θ
$\rightarrow$ There is no probability!



The confidence level $p = 95$ %
is the expected long-term success rate

Renewed success of Bayesian statistics: has triggered a reconsideration of the interpretation of measurement uncertainty

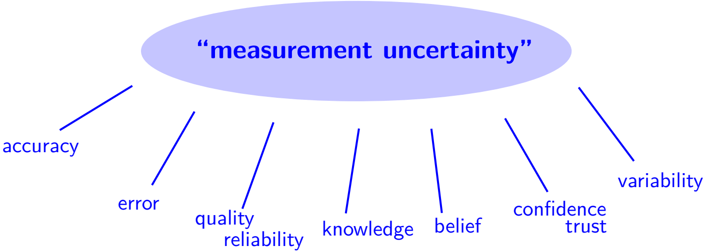
→ Emphasis put on the *subjectivity* of a measurement result

ex. W. Bich: "the published uncertainty is *my* uncertainty"

Renewed success of Bayesian statistics: has triggered a reconsideration of the interpretation of measurement uncertainty

→ Emphasis put on the *subjectivity* of a measurement result

ex. W. Bich: "the published uncertainty is *my* uncertainty"



Renewed success of Bayesian statistics: has triggered a reconsideration of the interpretation of measurement uncertainty

→ Emphasis put on the *subjectivity* of a measurement result

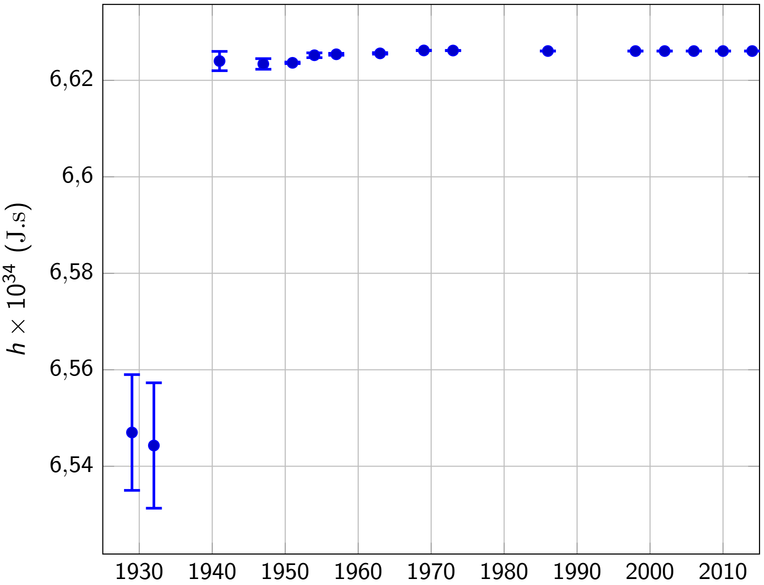
ex. W. Bich: "the published uncertainty is *my* uncertainty"

"Betting odds" (ex. Thomsen & Franken 1971)

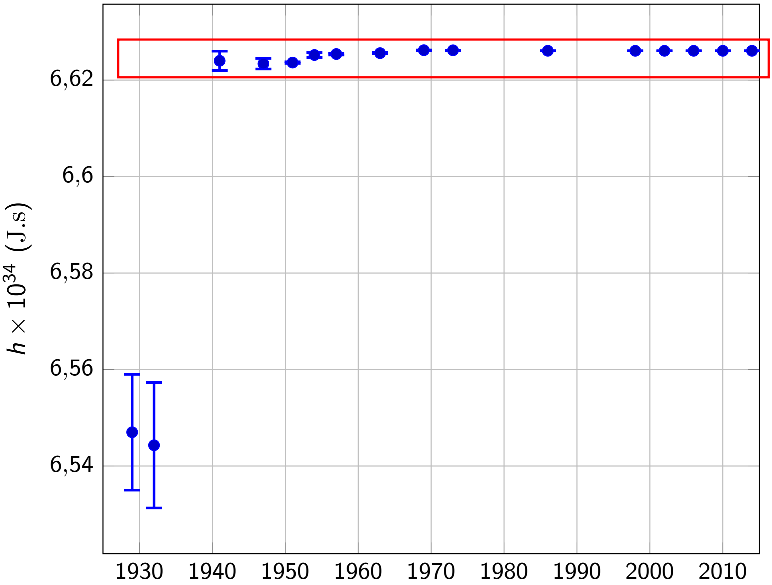
[...] the experimenter must recognize that he is quoting betting odds. Thus, if he states a result as (100 $\pm$ 1) cm (probable error) he is asserting that there is a 50 percent probability that the true value lies between 99 cm and 101 cm. If he has formed his error estimate honestly, avoiding both overoptimism and undue conservatism, *he should be willing to take either side of the bet.* This is the essence of an honest error estimate.

2– The adjustments of the fundamental physical constants

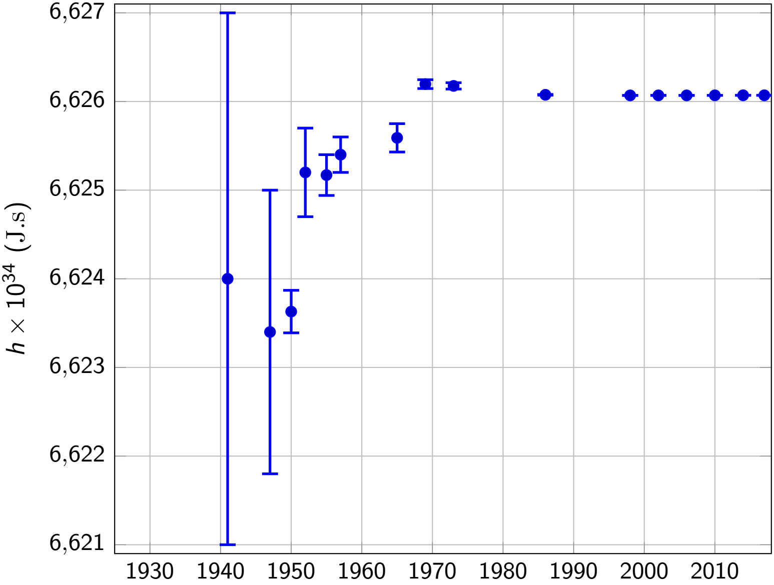
Planck's constant *h* from 1929 to 2017



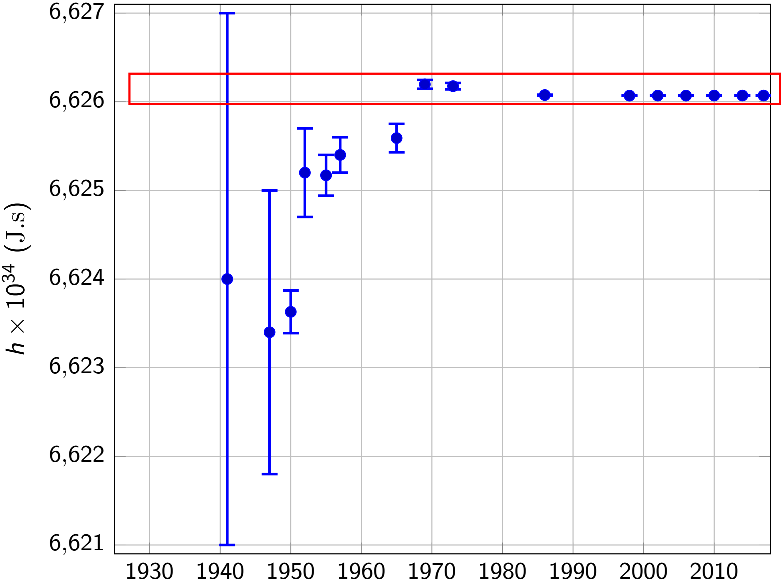
Planck's constant *h* from 1929 to 2017

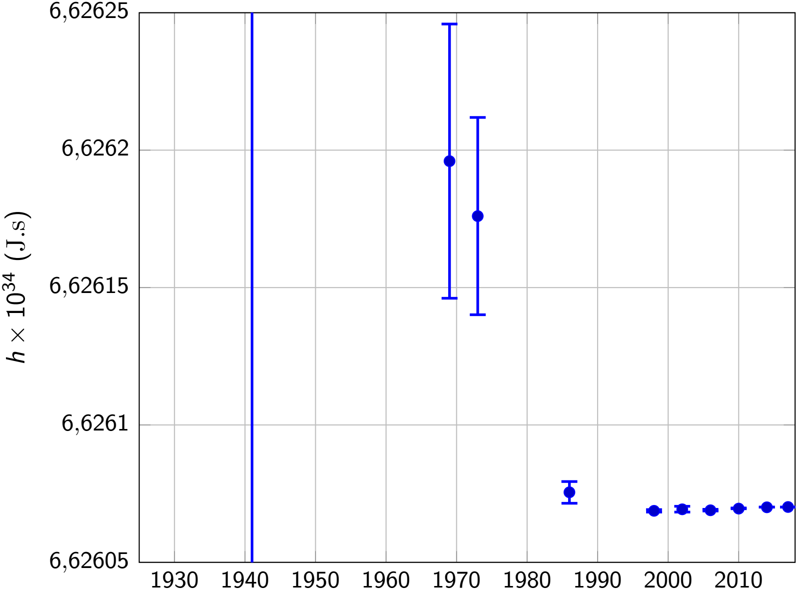


Planck's constant *h* from 1929 to 2017

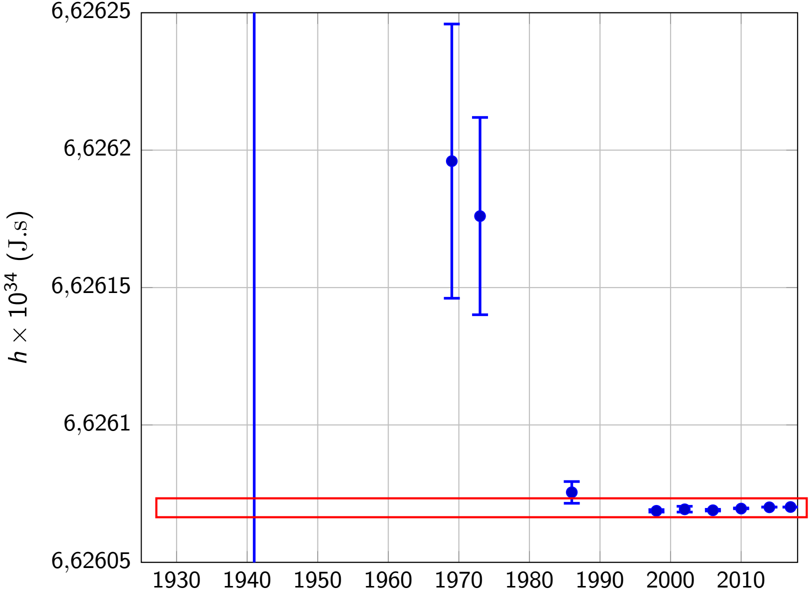


Planck's constant h from 1929 to 2017

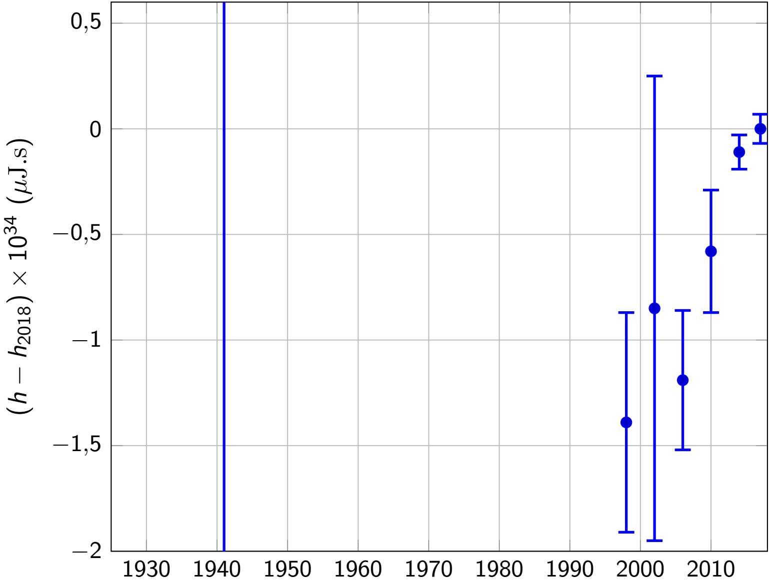



6,62625
6,6262
6,62615
6,6261
6,62605
$h \times 10^{34}$ (J.s)
1930
1940
1950
1960
1970
1980
1990
2000
2010

Planck's constant *h* from 1929 to 2017



Planck's constant *h* from 1929 to 2017



Repeated leaps across time

$\Rightarrow$ **why should we have any confidence in today's values?**

Question: confidence in recommended values

Repeated leaps across time
⇒ **why should we have any confidence in today's values?**

Henrion & Fischhoff, *Assessing uncertainty in physical constants* (1986)

"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. [...] the most common problem is overconfidence"
(*American Journal of Physics* **54** p.791)

→ **"surprise index":** "the surprise index is the percent of 98% confidence intervals for which the true value is a 'surprise.'"

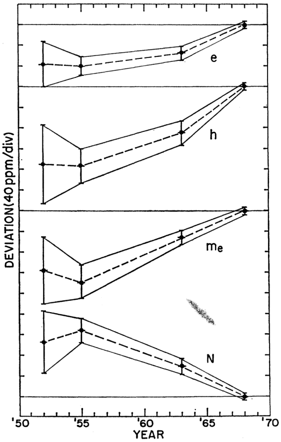
Precision Measurement and Fundamental Constants

Proceedings of the International Conference
held at the National Bureau of Standards
Gaithersburg, Maryland, August 3–7, 1970

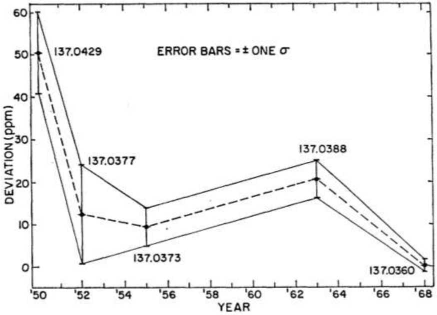
PANEL DISCUSSION:

SHOULD LEAST-SQUARES ADJUSTMENTS OF THE FUNDAMENTAL CONSTANTS BE ABOLISHED?

Taylor, Parker and Langenberg 1969



e, h, m_e and N
p.481



Fine-structure constant $1/\alpha$
p.379

Precision Measurement and Fundamental Constants

Proceedings of the International Conference
held at the National Bureau of Standards
Gaithersburg, Maryland, August 3–7, 1970

PANEL DISCUSSION:

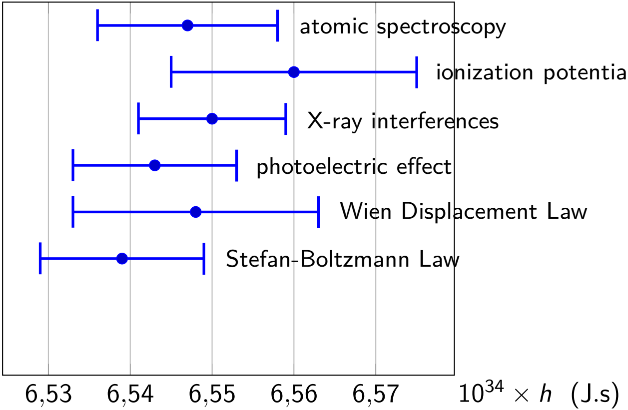
SHOULD LEAST-SQUARES ADJUSTMENTS OF THE FUNDAMENTAL CONSTANTS BE ABOLISHED?

- Are the adjusted values *safe*?
- *Should* they be safe/made safer?

⇒ dilemma between *safety* and *precision*.

- **Safety**: uncertainties should not be understimated / too optimistic
- **Précision**: results are not intended to be safe

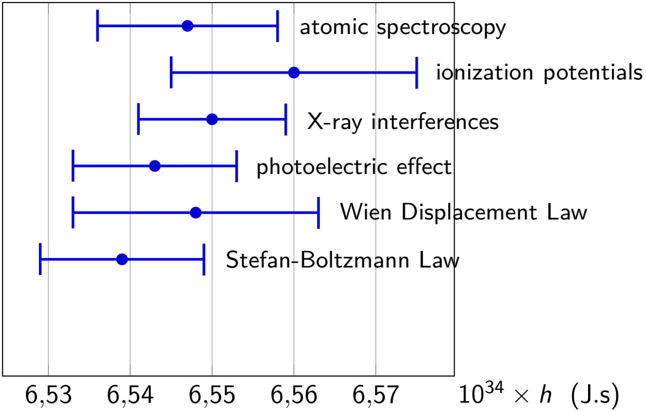
Birge 1929: Planck constant h



Raymond Thayer Birge (1887-1980)

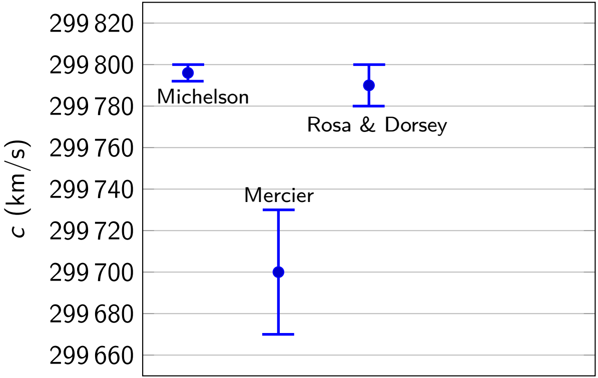
Probable values of the general physical constants
-- *Reviews of Modern Physics*, 1929, **1**, 1-73

Birge 1929: Planck constant h

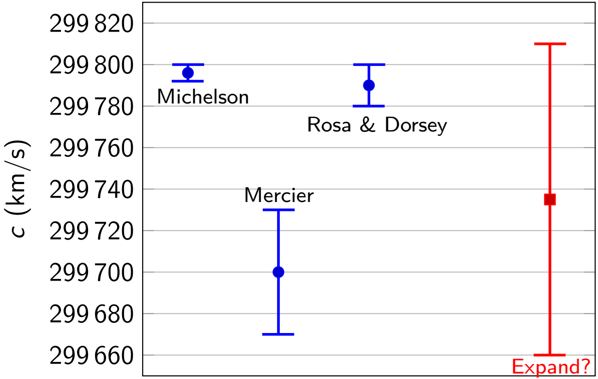


Values mutually consistent, $\boxed{h_{\mathrm{adj}} \propto \sum_{i=1}^{6} \frac{h_i}{{u_i}^2}}$

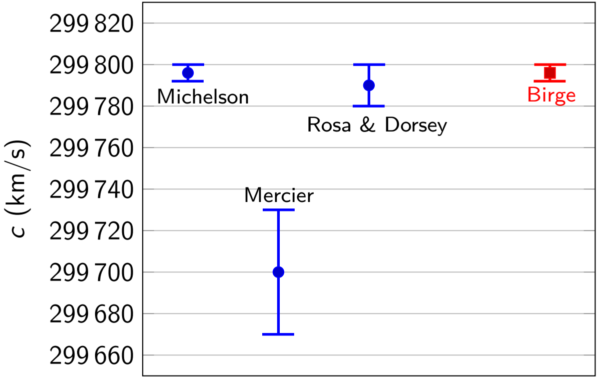
Birge 1929, speed of light c



Birge 1929, speed of light c

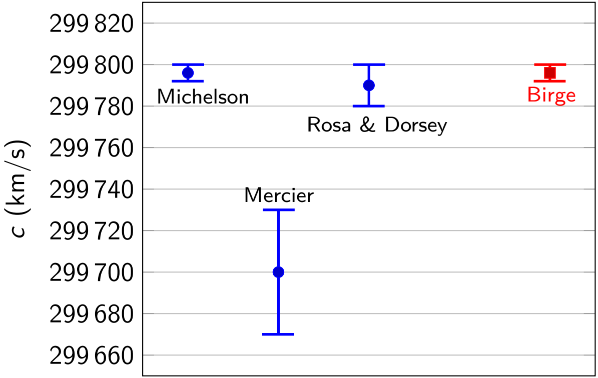


- Inconsistent values: expand the uncertainty?



- Inconsistent values: expand the uncertainty?
- Birge: chose to **select** Michelson's value → judgment

Birge 1929, speed of light c



- Inconsistent values: expand the uncertainty?
- Birge: chose to **select** Michelson's value → judgment

The **filtering of the input data** has always been a topic of discussion throughout the history of the adjustments of the physical constants

Peter Bender, 1970 conference, "Handling of Discrepant Data in Evaluations of the Fundamental Constants"

[Most users] would rather use values which are at least consistent with each other and which someone who has studied the problem carefully feels are as close as one can come to the true ones at the time of their adoption. [...] The major question is how to make the value and uncertainty that are chosen as unbiased as possible.

Up until now people doing evaluations have felt compelled to make crisp yes-or-no decisions on how to handle discrepancies, and no allowance has been made in the final uncertainties for the possibility that the wrong choice was made.

[A] fair estimate of the actual uncertainty in the result [is] the most important goal. To summarize, the basic desire is to find a way to avoid having the quoted uncertainty in the results be systematically too small because of throwing out data. (p.493-494)

Against safety: **process** more important than outcome

Taylor, 1970 conference, "Comments on Least-Squares Adjustments of the Constants

True, it is worthwhile to have at any given epoch a consistent set of constants which can be used by all workers requiring them. But this is the least important result of a constants adjustment – the most valuable contribution of such studies to human knowledge is the information gained during the course of the critical review which necessarily accompanies the adjustment. (p.495)

The fact to keep in mind is that those scientists who really need to use the last decimal places will not be content just to take numbers out of a table but will go to the originating article. Those workers who are content to use the numbers as given without worrying about where they came from could use almost any number. (p.496)

Since the majority do not particularly care what numbers they use, adjustments should be geared to the small but more important minority who do and who need the most useful and stimulating numbers they can get their hands on. (p.497)

Our philosophy in the 1969 adjustment was to provide the best possible cutting tool for those workers who needed it most and who were prepared to use it in full knowledge that care was necessary to avoid cutting themselves. (p.497)

Cohen & DuMond (1965)

The idea that an overestimate of error "for safety" is somehow [...] laudable or virtuous [...] is somehow deplorably prevalent. We ask, for whom is such an overestimate "safe"? Certainly not for the general scientific community who wish to use the result. For them it is a concealment of the true facts regarding the results of the measurements.

Taylor (conférence 1970)

It is an admission of defeat [...] the only way out is to assume that all the data is bad. It therefore throws away information by making quantities more unknown that they actually are.

Taylor, Parker & Langenberg (1969)

Measurements of the fundamental physical constants to ever greater levels of accuracy are important, not just because they "add another decimal point" and provide us with a more consistent set of constants to work with, but because they may lead to the discovery of a previously unknown inconsistency or the removal of a known inconsistency in our physical description of nature.

Safety against precision (2)

Purpose?

Ensure a result as exact as possible

⇓

Safety

Measurement uncertainty must not be underestimated

Margin of security = lose precision but maintain accuracy

↓

Implicitly, characterizes measurement uncertainty in an objective way (success rate)

Account for the experimental outcomes as well as possible

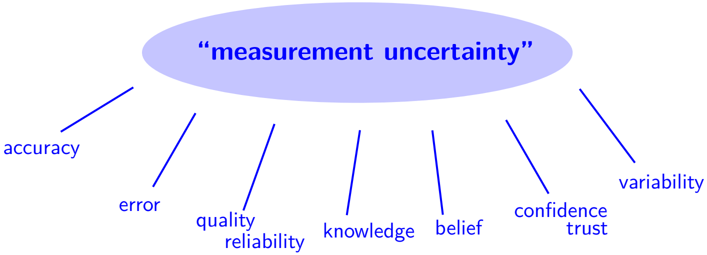
⇓

Precision

No particular worry about the long-term validity of the result on the long-run

↓

Incompatible with a characterization of measurement uncertainty leaning towards accuracy/error


"measurement uncertainty"
accuracy
error
quality
reliability
knowledge
belief
confidence
trust
variability

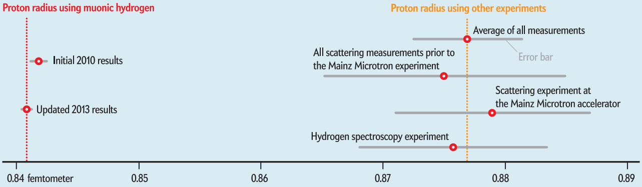
Jan C. Bernauer is a postdoctoral researcher in nuclear physics in the Laboratory for Nuclear Science at the Massachusetts Institute of Technology.

Randolf Pohl works on laser spectroscopy of hydrogen and hydrogenlike exotic atoms at the Max Planck Institute of Quantum Optics in Garching, Germany.

RESULTS

The Incompatible Measurements

The size of the proton should stay the same no matter how one measures it. Laboratories have deduced the proton radius from scattering experiments [*see box on opposite page*] and by measuring the energy levels of hydrogen atoms in spectroscopy experiments. These results were all consistent to within the experimental error. But in 2010 a measurement of the energy levels of so-called muonic hydrogen [*see box on page 38*] found a significantly lower proton radius. Attempts to explain the anomaly have so far failed.



SOURCE: RANDOLF POHL

Thank you for your attention!