1

The importance of uncertainty in science and technology

We live with uncertainty every day. Will the weather be fine for a barbecue at the weekend? What is the risk to our health posed by a particular item of diet or environmental pollutant? Have we invested our money wisely?

It is understandable that we would like to be able to eliminate, or at least reduce, uncertainty. If we can reduce it significantly, we become more confident that a desirable event will happen, or that an undesirable event will not. To this end we seek out accredited professionals, such as weather forecasters, medical researchers and financial advisers.

However, in science and technology uncertainty has a narrower meaning, created by the need for accurate measurement. Accurate measurement, which implies the existence of *standards* of measurement, and the evaluation of uncertainties in a measurement process are essential to all areas of science and technology. The branch of science concerned with maintaining and increasing the accuracy of measurement, in any field, is known as *metrology*.[1] It includes the identification, analysis and minimisation of errors, and the calculation and expression of the resulting uncertainties.

Whether or not a measurement is regarded as 'accurate' depends on the context. Supermarket scales used for weighing fruit or vegetables need not be better than 1% accurate. By contrast, a state-of-the-art laboratory balance is able to determine the value of an unknown mass of nominal value one kilogram[2] to better than one part in ten million. These figures, 1% in one case and one part in ten million in the other, are numerical measures of the degree of accuracy: low in the first case and high in the second, but each of them fit for its particular purpose. Evidently, accuracy and

[1] This word derives from the Greek 'to measure'. It should not be confused with *meteorology*, the study of climate and weather. The need for accurate measurement, and for standards of length, weight and volume (for example), was recognised in many ancient societies with relatively primitive technology and hardly any 'science' in the modern sense.

[2] A 'nominal' value is the ideal or desired value of a particular quantity. Thus the nominal value of the mass of an object might be 1 kilogram, implying that its accurately measured value is close to 1 kilogram.

uncertainty are inversely related: high accuracy implies low uncertainty; and low accuracy implies high uncertainty.

When we say that the result of a measurement has an associated uncertainty, what, exactly, are we uncertain *about*? To begin to answer this question, we acknowledge that the result of a measurement is usually a *number* expressed as a multiple of a *unit* of measurement. As in the example of the laboratory balance above, we should refer to a number that results from a measurement as a *value*. For example, a person's mass may have a value of 73 kilograms, meaning that the mass is 73 units, where each unit is one kilogram. Similarly, the temperature of coffee in a cup may be 45 degrees Celsius, the length of a brick 231 millimetres, the speed of a car 60 kilometres per hour, and so on. The value that expresses the given quantity therefore depends on the unit. The same speed of the car, for example, could be expressed as 17 metres per second. There are cases where the value is independent of the unit. This happens when a quantity is defined as a ratio of two other quantities, both of which can be measured in terms of the same unit. The units then 'cancel out'. For example, the coefficient of static friction, μ_s, is defined as the ratio of two forces and therefore μ_s is a *dimensionless* number; for glass on glass, $\mu_s \simeq 0.94$.

A measurement whose result is characterised by a value holds more information than a measurement whose result is not characterised in this way. In the latter case we might hesitate to call the result a 'measurement'; it would be more in the nature of an opinion, judgment or assessment. In fact, this is how we tend to function in everyday life. When parking a car in a busy street, the driver estimates the available space in most – though not all – cases quite adequately without a rule or tape-measure. We may think a person handsome or beautiful, but it would be rash to attempt seriously to attach a numerical value to this. (If we drop the word 'seriously', then it is possible. A 'millihelen' may be defined as the amount of beauty required to launch exactly one ship![3])

The information-rich use of a value to characterise the result of a measurement comes at a price. We should also consider – particularly in pure and applied science, in medicine and in engineering – how 'uncertain' that value is. Is the length of the brick 231 millimetres, or more like 229 millimetres? What is the most appropriate instrument for measuring the length of the brick, and how can we be sure of the accuracy of the instrument? How, in any case, do we define the 'length' of a brick, which may have rough or uneven edges or sides? How much 'leeway' can we afford to allow for the length of a brick, before we must discard it as unusable?

This book considers measurement, uncertainty in measurement and, in particular, how uncertainty in measurement may be quantified and expressed. International

[3] This refers to a story from ancient Greece, as recounted by Homer in the *Iliad* around the eighth century BC. The beautiful Helen of Sparta, in Greece, had been taken to Troy (in what is now Turkey), and that started the ten-year Trojan War. The Greeks launched a fleet of one thousand ships to reclaim her.

guidelines exist to assist in these matters. The guidelines are described in the *Guide to the Expression of Uncertainty in Measurement*, published by the International Standardisation Organisation (corrected and reprinted version in 1995), abbreviated as ‘the GUM’. Before discussing and illustrating these guidelines in detail, we highlight the importance of measurement and uncertainty by considering some examples.

1.1 Measurement matters

Just how important are measurement and uncertainty? Careful measurement with properly identified and quantified uncertainties could lead to a new discovery and international recognition for the scientist or scientific team that made the discovery. To the engineer it may lead to improved safety margins in complex systems such as those found on the space shuttle, and to the police it could contribute to the successful prosecution of a driver who exceeds the speed limit in a motor vehicle. In biochemical metrology, accurate measurement is needed for the reliable estimation of (for example) trace levels of food contaminants such as mercury in fish. In medical metrology, high accuracy in blood-pressure measurements reduces the risk of misdiagnosis. We now give some examples of advances in measurement accuracy. At the end of this chapter we indicate where further information on these topics may be found. We use the SI (Système International)[4] units of measurement, which include the metre (m) for distance, the kilogram (kg) for mass and the second (s) for time.

1.1.1 Measurements of the fundamental constants of physics

Theories of the physical world incorporate fundamental constants such as the speed of light, c, the Planck constant, h, the fine-structure constant, α, and the gravitational constant,[5] G. As far as we know, these are true constants: they do not change with time or location and have the same values on Earth as anywhere else in the Universe. In many cases their numerical values are accurately known, and in a few cases the constants have been exactly defined. For example, the speed of light, c, in a vacuum is defined as $c = 299\,792\,458\ \mathrm{m \cdot s^{-1}}$. The Planck constant, h, which is the ratio of the energy of a photon of radiation to its frequency, is accurately known: $h = 6.626\,069 \times 10^{-34}\ \mathrm{J \cdot s}$ (joule-second) with an uncertainty of less than one part in a million.

[4] The French acronym is universally used in recognition of the central role played by France, during the late eighteenth century and later, in introducing and establishing the uniform system of units of measurement that came to be known generally as the ‘metric’ system and that later evolved into the SI.

[5] ‘Big G’ is not to be confused with g, ‘little g’, which is the acceleration due to gravity near the Earth’s surface and varies with location.

The gravitational constant, G, appears in the equation that describes the inverse-square law of gravitation discovered by Isaac Newton in the seventeenth century: $F = Gm_1m_2/r^2$, where F is the gravitational force of attraction between two masses m_1 and m_2 a distance r apart. To calculate the force using this equation, we must know the value of G. With the practically available masses in a laboratory this force is tiny because G is very small: about $6.68 \times 10^{-11} \mathrm{m}^3 \cdot \mathrm{kg}^{-1} \cdot \mathrm{s}^{-2}$. For example, two uniform spherical bodies, each of mass 200 kg, whose centres are separated by 1 m (these could be two solid steel spheres each of approximate diameter 36 cm) would attract each other with a gravitational force of about 2.7×10^{-6} N. This is roughly one-tenth the weight of a small ant (mass $\simeq$3 mg).

We have a healthy respect for the Earth's gravitational force, but this is largely due to the enormous mass of the Earth, about 6×10^{24} kg (this mass has to be inferred from a known value of G). In measuring G, the tiny gravitational forces that exist between bodies in a laboratory make an accurate measurement of G very difficult. These tiny forces must somehow be measured against a background of competing gravitational forces, including the much larger ordinary gravity due to the Earth as well as the gravity exerted by the mass of the scientist doing the experiment! At the time of writing (2005), the accepted fractional uncertainty in G is about one part in ten thousand. This is much larger than the fractional uncertainty with which other fundamental constants are known. Previous attempts to measure G made in the 1990s yielded results that were mutually discrepant by several parts per thousand, even though much smaller uncertainties were claimed for some of the individual results.[6] Experiments to measure G accurately are evidently beset by subtle *systematic* errors (systematic errors will be discussed later in this book).

When G or any other particular quantity is measured, it is important to know the uncertainty of the measurement. If two values are obtained for the same particular quantity, and these values differ by significantly more than the uncertainty attached to each value, then we know that 'something is wrong': the quantity has perhaps undergone some change in the interval between the two measurements, or systematic errors have not been properly accounted for. The latter interpretation is evidently the more likely one with respect to the determination of G.

Painstaking measurements of G, and of other fundamental constants, yield new insights into our physical world. In applied physics and engineering, seeking reasons for discrepancies often leads to better understanding of materials or of laboratory techniques. In the case of G, where several experiments have been based on the twisting of a strip of metal (a 'torsion strip') in response to the gravitational field of nearby masses, it has been found that such torsion strips are not perfectly elastic (that is, the amount of twist is not exactly proportional to the torque), and the

[6] Figure 4.2 in chapter 4 illustrates this.

amount of this so-called 'anelasticity' is significant. This finding is a contribution to knowledge in its own right. In theoretical physics, high-accuracy measurements of G will eventually contribute usefully to current speculation as to whether there may be some small but detectable violation of the inverse-square law over laboratory distances and even at the sub-millimetre level. Such violations would have profound implications for our understanding of the Universe. It is only through careful measurement and realistic estimates of uncertainty that we can have confidence in any conclusions drawn from results of studies designed to establish a value for G.

1.1.2 Careful measurements reveal a new element

At the end of the nineteenth century, Lord Rayleigh showed the benefits that accrue from close scrutiny of the results of measurements that appear at first glance to be consistent and to contain nothing very surprising. Rayleigh used two methods to measure the density of nitrogen.[7] In one method, the nitrogen was obtained wholly from the atmosphere, by passing air over red-hot copper that removed all the oxygen. In the other method, the nitrogen was obtained by bubbling air through ammonia and then passing the air–ammonia mixture through a red-hot copper tube. This also removed the oxygen (which combined with hydrogen from the ammonia to form water), but partly 'contaminated' the nitrogen from the air with nitrogen from the ammonia itself. The nitrogen obtained by the second method (the 'chemical' method) was about 0.1% less dense than that given by the first method (the 'atmospheric' method). Despite the close agreement, Rayleigh was uncomfortable with the 0.1% discrepancy and resisted his instinct to find ways to downplay or ignore the difference. Instead, he undertook a detailed study in which he tried to exaggerate the difference by varying the experimental conditions. He replaced the air in the chemical method by pure oxygen, so that all the collected nitrogen originated from the ammonia. This modified chemical method now provided nitrogen that was 0.5% less dense than that obtained by the atmospheric method. Thus Rayleigh had strong evidence that nitrogen derived from the atmosphere had a (very slightly) greater density than nitrogen derived from 'chemical' sources (for example, ammonia).

The inescapable conclusion of Rayleigh's careful measurements was that his atmosphere-derived 'nitrogen' was in fact nitrogen mixed with another gas. The gas that Rayleigh had discovered was argon, a new element, and for this discovery Rayleigh was awarded the Nobel prize in physics in 1904. While 78% of the atmosphere is nitrogen, only about 1.2% is argon, but argon is denser than nitrogen

[7] Rayleigh measured a mass of nitrogen. Since this was done at a standard temperature and pressure, the volume of nitrogen was fixed, so effectively its density was measured.

by a factor of about 1.4. So atmosphere-derived nitrogen, containing unidentified argon, appeared to be denser than chemical-derived nitrogen.

Rayleigh's original measurements of the collected mass of nitrogen were made with an uncertainty of about 0.03% or less. A larger uncertainty might easily have obscured the small (0.1%) systematic discrepancy that compelled him to pursue the matter further. This story illustrates the need for accurate measurement, the benefit gained by measuring a quantity in more than one way and the importance of explaining any discrepancy thereby revealed.

Since Rayleigh's time, experimental methods and instruments have advanced significantly so that, for example, instruments under computer control can gather vast amounts of data in a very short time. With respect to measurement and uncertainty, this brings its own challenges.

1.1.3 Treat unexpected data with caution

In 1985 scientists doing atmospheric research in Antarctica announced that the ozone layer over the South Pole was being depleted at quite a dramatic rate. Their conclusion was based on ground measurements of ultraviolet radiation from the Sun that was absorbed by the atmosphere. For several years prior to this, other scientists had been 'looking down' on the ozone layer using satellites, though they had reported no change in the depth of the layer. A contributory factor to the inconsistency between the ground-based and satellite-based data could be traced to the processing of the satellite data. Natural variation in values of the thickness of the ozone layer was well known. Therefore it appeared reasonable, when processing the satellite-based data, to discard 'outliers' – that is, data that appeared not to conform with that natural variation. The problem with this approach was that, if the 'natural' variation were *itself* changing, one risked discarding the very data that would reveal such a change. When the satellite data were reanalysed with the outliers included, the conclusion of the Antarctica scientists was supported. The effect of this prominent work was to fuel international debate among scientists, industry and governments on the causes, consequences, extent and treatment of ozone depletion in the atmosphere.

Quantifying ozone depletion by investigating the absorption of ultraviolet radiation by the Earth's atmosphere is an example of the application of optically based measurement. Optically based methods of measurement are widely used, and many rely on that most versatile of devices, the laser.

1.1.4 The laser and law-enforcement

The laser ('light amplification by stimulated emission of radiation'), invented and developed in the early 1960s, offers very high accuracy in length measurement in research and industry. Laser interferometry is a standard technique used in industry to measure length to sub-micrometre precision. This is made possible by the monochromatic ('single-colour') nature of laser light, implying a single wavelength and therefore a natural 'unit of length'. The red light from an iodine-stabilised helium–neon laser has a wavelength of 632.991 212 58 nm (nanometres or 10^{-9} m), with an uncertainty of the order of a few parts in 10^{11}. A measurement of length can, therefore, be 'reduced' to counting wavelengths: more precisely, the counting of interference fringes that result from the interference of the beam of laser light with a similar reference beam.

Applications of lasers even extend to law-enforcement. The speed of a vehicle can be established by aiming a narrow beam of pulsed infra-red radiation emitted by an instrument containing a laser (the 'speed-gun') at the body of the moving vehicle. The pulses are emitted at an accurately known rate of the order of 100 pulses every second. The radiation is reflected by the body and returns to the instrument. If the vehicle is moving towards the speed-gun, the interval between successive reflected pulses is less than the interval between successive transmitted pulses. This difference is small, of the order of nanoseconds (or billionths of a second), but can be accurately measured. This difference and the known value of the speed of light enable the speed of the vehicle to be determined. Speeds recorded well in excess of the speed limit can lead to instant licence disqualification in some countries, and an appearance in court. Identifying and understanding the complications that may affect the value measured for the vehicle speed is the starting point for estimating the uncertainty of the measurement of speed. Such complications include the exact angle of the speed-gun relative to the direction of the vehicle, interfering effects of bright light sources and whether the speed-gun has been accurately calibrated and is not significantly affected by variations in ambient temperature. It is only when the uncertainty in the speed is known that it is possible to decide whether a vehicle is very likely to be exceeding the speed limit.

1.1.5 The Global Positioning System (GPS)

A GPS receiver can determine its position on the Earth with an uncertainty of less than 10 metres. This is made possible by atomic clocks carried on satellites orbiting the Earth with an approximate half-day period and at a distance of about 20 000 kilometres. The atomic clocks are stable to about one part in 10^{13} (equivalent to gaining or losing one second in about 300 000 years). Atomic clocks of this degree of stability evolved from research by Isador Rabi and others in the 1930s and later on the natural resonance frequencies of atoms. The receiver contains its own clock (which can be less stable) and, by comparing its own clock-time with the transmitted satellite clock-times, the receiver can calculate its own position. The comparison of clock-times must take into account the first-order Doppler shift, of about one part in 10^5 in the case of the GPS, of the frequency of a clock moving towards or away from a fixed clock.[8] A further requirement for the accuracy of the GPS is the relativity theory of Albert Einstein. Two of the relativistic effects that must be taken into account are the slowing (time-dilation) of satellite clocks moving transversely relative to fixed clocks (this is also known as the second-order Doppler shift) and the speeding up of clocks far from the Earth's surface due to the weaker gravitational field. These two effects act in opposition and have magnitudes of about one part in 10^{10} and five parts in 10^{10}, respectively. So two major branches of theoretical physics have made possible timekeeping metrology of extremely high accuracy and have revealed subtle properties of time and space. As a result, inexpensive devices that accurately determine the location of aircraft, ships and ground vehicles, and help with the safety of explorers and trekkers, are now available.

1.1.6 National measurement institutes, international metrology and services to industry

It is obvious that industrial products must perform reliably. This implies something that is perhaps not so obvious: the relevant physical properties of their components must be certified against local and, ultimately, international standards of measurement. Such standards are very precisely and meticulously manufactured objects, for example steel rules and tape-measures, standard weights, standard resistors and standard lamps. If component A of a motor-vehicle (for example) must fit or be compatible with component B, this certification will ensure that, if A is made in country X and B in country Y, A will fit B in country Z where the motor-vehicle is assembled. International certification depends on the existence of standards of measurement in every field of science and technology. Research into, and the development and maintenance of, standards of measurement at the highest possible level of accuracy are the function and responsibility of a country's national measurement institute (NMI).

For a physical property of a component to be certified, it must be compared with or *calibrated against* the relevant standard. If the component is (for example) a 1000-Ω resistor, its resistance will be compared with a local 1000-Ω *standard* resistance, which may, however, be of relatively low accuracy. This standard, in turn, must be calibrated against a higher-accuracy standard, generally maintained by industrial calibration laboratories, and so on until the top of the comparison chain is reached. This would normally be the national standard of resistance maintained

[8] This first-order Doppler effect is familiar to us in its acoustic analogue as the raised pitch of the sound made by an approaching object, and the lowered pitch when it recedes.

by an NMI, and would itself be validated by frequent international comparisons by various NMIs of such national standards (or of very stable and highly accurate standards directly *traceable* to such national standards). The degree to which the participating NMIs' standards 'agree with one another' or, more formally, have the essential property of 'mutual equivalence' is a decision made by the BIPM (Bureau International des Poids et Mesures, or International Bureau of Weights and Measures, in Paris). Such international comparisons are a routine feature of international metrology, and serve to maintain the reliability and underpin the quality control of a huge variety of industrial products in day-to-day trade and commerce.[9]

Progress in metrology – namely, permanently improved standards and reduced uncertainties – is usually made by an NMI, although occasionally by other institutions. This happens through a major change in method inspired by a novel application of existing knowledge, or by use of an advance in physics or other science. There are many such examples; two will be described here, while other cases will be mentioned later in the book.

1.1.6.1 Standards of electrical resistance and capacitance

The history of the standard of resistance provides a good example of the kind of research, often in seemingly unrelated areas, that informs progress in metrology. For about the first half of the twentieth century the 'international ohm' was defined as the resistance of a specified length and volume of mercury at a specified temperature. The complicating factors here are the inevitable uncertainties in the measurements of the length, volume and temperature of the mercury, and uncertainty regarding its purity.

Another metrological route towards a standard of resistance could be found if a standard of capacitance could be defined. These are two quite different electrical quantities measured in different units, but there is a simple relationship between them. Unfortunately, a capacitance, C, is normally physically constructed as two metal plates separated by an insulating gap (assumed here to be a vacuum), and so is calculated using an expression of the form $C = \epsilon_0 A/d$, where ϵ_0 is the constant permittivity of free space (or vacuum),[10] A is the area of the capacitor plates and d is their separation (figure 1.1(a)). The uncertainty in C will now result from the considerable uncertainties in the measurements of A and d.

[9] To maintain standards (of performance as well as measurement) private and government laboratories, and NMIs themselves, undergo regular review by assessors. Successful review is followed by accreditation of the laboratory for its particular area of expertise. The NMIs are accredited through international comparisons and by means of peer-review by visiting experts from other NMIs.

[10] The value of ϵ_0, a natural constant, in SI units to eight significant figures is $8.854\,187\,8 \times 10^{-12}\ \mathrm{F \cdot m^{-1}}$. Ordinary capacitors as used routinely in electronics have insulating material (a 'dielectric'), rather than a vacuum, between the plates. The effective permittivity is then larger than ϵ_0.

It is important to note that ϵ_0 in the expression $C = \epsilon_0 A/d$ has units farad per metre. The product $\epsilon_0 A/d$ consequently has units (farad per metre) $\times$ metre^2/metre, or farad, equal to the units of C. A major advance in capacitance and resistance metrology would therefore result if a geometry could be found in which C was given simply as ϵ_0 multiplied by a distance, since this product would also have the units of a capacitance: farad per metre $\times$ metre gives farad. In effect, the nuisance of having to measure an area and a length would be replaced by the convenience of measuring only a length.

Using both mathematical analysis (starting from Maxwell's equations of electrostatics) and experimental verification, such a geometry was found in 1956 by A. M. Thompson and D. G. Lampard of the National Standards Laboratory of Australia (now known as the National Measurement Institute). This discovery became known as the Thompson–Lampard Theorem of Electrostatics. The most common practical realisation of this theorem is shown in figure 1.1(b) and has come to be known as the 'calculable capacitor'. Four identical circular cylinders A, B, C and D, each centred at the corner of a square, are enclosed within a circular earthed shield E and are separated from one another and from the shield by narrow insulating gaps. There are two earthed central bars. Only one of these (F) is shown, and F is movable perpendicular to the plane of the diagram. If F is moved a distance d, it can be shown that the resulting change, C, in capacitance ('cross-capacitance') between A and C (with B and D earthed) or between B and D (with A and C earthed) is given by $C = \epsilon_0[(\log 2)/\pi]d$. This is a small change, approximately 2 pF per metre. The distance d can be very accurately measured using laser interferometry.

We therefore note the crucial geometry-independent property of the calculable capacitor: the capacitance depends only on d, not on (for example) the diameters of the cylinders. Figure 1.1(b) could be scaled up or down in the plane of the diagram by any factor, and C would still be given as stated above.[11] In electrical metrology, geometry-independence is a prized attribute of any measurement that strives towards the highest accuracy.

Standard resistors of nominal value 1 Ω can be calibrated against the calculable capacitance C by means of well-established procedures. The calculable capacitor has, therefore, provided a realisable 'absolute' ohm, a primary standard much superior to the 'international' ohm mentioned previously. The uncertainty of the resistances is of the order of a few parts in a hundred million, and these form the primary standards for disseminating the practical unit of resistance throughout the research and industrial communities.

[11] Figure 1.1(b) is a particular case of a more general configuration involving four surfaces separated by narrow gaps. For this general case, C is given by a formula that still involves only a single distance measurement d and that reduces to $C = \epsilon_0[(\log 2)/\pi]d$ for figure 1.1(b).