1 What's wrong with consulting experts?

I came into the kitchen this morning and the sink was blocked. My wife, who is handier than me, tried the usual things: some drain-cleaning fluid and then a plunger. Despite her enthusiasm for the task, the drain remained blocked. So I rang a plumber. An expert. Someone who has done it before, has the right equipment, and has solved much more difficult drainage problems than were created by my inattention to coffee grounds.

I'm all for the division of labour and specialisation. It's an especially wise strategy for me because I'm one of the world's least handy people. If I need to clear a drain, I call a well-regarded plumber. If I wanted to build a sturdy bridge, I'd contact an engineer who has built lots of bridges. When I needed my knee reconstructed following a bad football tackle, I asked around and found an experienced surgeon with a good reputation.

I'll refer to those expert abilities as skills. Skills are abilities to execute particular tasks efficiently and effectively, acquired through training, concrete practice and feedback.[1] In all these cases, when their actions don't succeed, it's hard for the expert to blame someone else. Their failures are unambiguous and personal. Ideally, you'll be able to look at their records of efficient drain clearing, bridge building and surgical outcomes. Of course, there are shoddy plumbers, reckless engineers and incompetent surgeons. But in the main, they will do a much better job of these things than me.

This book is not about such skills. We also rely on experts for advice when we need to make decisions and we don't have enough information. In government, business and elsewhere, our reliance is greatest when circumstances are unique, the consequences of the decision are significant, the decision is imminent and the future is

uncertain. This book deals with the experts on whom we rely for estimation and prediction. Typically, such experts are defined by their qualifications, their experience[2] – and importantly, by their status among their peers. We find someone with the right training and experience, someone whom we trust and can understand.[3] Often, expert judgement is all we have. However, our propensity to turn to experts and accept their judgements uncritically, even when we don't need to, often appears to be automatic, or at least not sufficiently cautious.

This book takes experts to include engineers, political scientists, economists, military and police officers, lawyers and financial analysts, together with the more usual chemists, physicists, geologists, biologists and medical scientists. These people may have many skills. However, their expertise for estimation and prediction is not necessarily supported by relevant, concrete actions and verifiable outcomes. When I use the term 'expert' I will refer to people who are considered by their peers and society at large to have specialist knowledge and who are consulted to make an estimate or prediction. I will show that, in many situations, non-skill-based expertise may not be worth the time and expense involved in using it. While superficially, such expertise may appear to have the same foundations as skill, often it does not.

There is a continuum between skill-based judgement and expert predictions. An engineer's skill may be to design a particular kind of bridge. Circumstances may be such that we consult them on related matters in which they have no direct experience, such as building other kinds of bridges. Beyond that, they may also appear to be expert in more distantly related topics, such as other structures, but have had no exposure to them beyond the things they've seen in textbooks or heard from colleagues. At what point does their ability to estimate or predict become no better than that of a random person from the street? Do they know, themselves, when their knowledge becomes too thin? Do their peers know? We will answer these questions in the chapters that follow.

WHAT DO EXPERTS DO?

Broadly, experts help with three kinds of questions.[4]

They estimate clearly defined, verifiable facts such as:

How prevalent is this disease in the population?
What is the maximum weight this bridge can carry?

They predict events, such as:

Will the President still be in office next year?
How much rain will fall next week?

Quite often, they advise on questions about the best course of action, such as:

What is the best way to manage this problem?
Is this the best portfolio of investments for me?

These are variations on what decision theorist Simon French calls the 'expert problem'.[5] Someone facing a decision consults an expert. The decision-maker alone is responsible for the decision. The emphasis in this book is on how the decision-maker should learn from experts. Sometimes, the experts are also the decision-makers, and sometimes, experts provide information without a decision in mind. I will touch on these issues briefly in the final chapter.

Experts may create models of underlying processes to help them make predictions. For example, many atmospheric scientists, physicists, glaciologists, earth scientists, oceanographers and biologists have been developing models for many years to forecast the outcomes of increasing carbon dioxide in the earth's atmosphere.[6] Their work is the basis for global policy decisions.

When answering questions about verifiable facts, we want the expert to draw on the storehouse of data they have accumulated through training and experience. In the case of predicting outcomes of events, we want them to use models together with their treasure trove of data and experience. We are especially demanding when asking about a course of action because we expect the expert to have

data and models on hand and to understand our context and sensitivities. We trust them to have our best interests in mind. We will see, however, that often this is not the case.

The need for experts is felt keenly when it comes to making decisions about the guilt or innocence of people in trials. John Lawson, a lawyer from the University of Missouri, wrote the foundation rules for expert and opinion evidence for the US legal system in 1900. In these rules, opinion is not admissible in evidence except '*on questions of science …persons instructed therein by study or experience may give their opinions. Such persons are called experts*'.[7] This definition is reiterated in the US Federal Rules of Evidence that state that a witness may qualify as an expert by possessing '*knowledge, skill, experience, training, or education*'.[8] New Federal Rules of Evidence and subsequent decisions broadened the definition to include opinions '*of a type reasonably relied upon by experts in the particular field*'.[9] They must be scientifically reliable (accounting for procedural care and predictive reliability) and grounded in scientific principles and appropriate methodology.[10] All jurisdictions allow expert opinion to inform courts about facts that might be otherwise unattainable because they are future probabilities, contingencies or facts '*not within positive knowledge*'.[11]

Society generally accepts that scientific and technical experts provide a unique, valuable resource. The US National Research Council, for instance, asserted that scientific experts have indispensible knowledge, methodological skills and experience.[12] And scientists themselves believe it. For example, a review of expert veterinary epidemiologists stated '*[e]xperts can be excellent reservoirs, integrators and interpreters of knowledge. In many settings, their ability to generate accurate predictions is a critical function of their profession and a key measure of their success: for example a stock broker's ability to forecast performance of a market, or a physician's ability to triage and assess a patient's need for hospitalization*'.[13] I will examine this general claim as well as the specific performances of

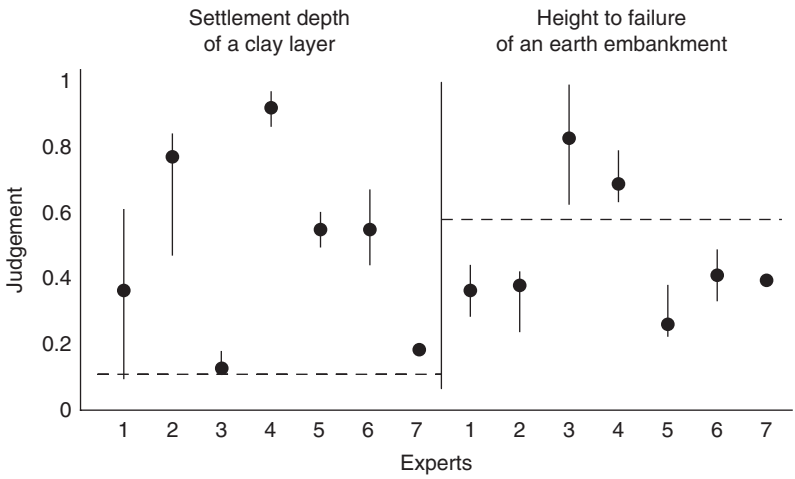

FIGURE 1.1 The correct (measured) value for settlement depth was 1.5 cm and for height to failure was 4.9 m. The y-axis for both was rescaled so the maximum value was 1. Correct values are shown as dashed horizontal lines. The intervals join the 'minimum' and 'maximum' values reported by the experts.

stockbrokers and physicians. In many circumstances, such optimism is misplaced.

SO, WHAT'S THE PROBLEM?

Geophysicist Ellis Krinitzsky spent many years working on earthquake risk, a notoriously difficult scientific problem. In an early review on the reliability of experts, he described an experiment in which seven geotechnical experts predicted the height of fill at which an embankment would fail, and the depth to which sediment would settle.[14] These questions were typical of the kinds of problems geotechnical experts were expected to assess reliably. The experts were provided with the data to make calculations. They used a variety of methods.

The results were not heartening. In Figure 1.1, the dashed lines represent the correct answers to the two questions. The dots are the expert's best guesses. The vertical lines, their uncertainty intervals, connect their 'minimum' and 'maximum' estimates.

There are at least six important things to note about the results of this simple experiment. First, the experts were generally overconfident. They were reasonably sure that the truth lay within the interval shown by the lines connecting their minimum and maximum guesses. However, in the first case, only two people's intervals enclosed the truth. In the second case, no-one's interval enclosed the truth. If their estimates of uncertainty were generally reliable, we would expect most of the intervals to enclose the horizontal dashed lines. Because they did not, it means that, in both cases, the experts were overconfident when they assessed the reliability of their own knowledge.[15]

Second, geophysicists conducted the study in the 1970s. Therefore, technical experts have been aware of these kinds of phenomenon for at least 40 years.

Third, it's possible for everyone to be wrong in the same direction. In the left-hand panel, all the experts overestimated the truth. So, whole groups of experts may be biased.

Fourth, the fact that someone did well on one question does not mean that they will do well on another. Expert 4 did best in the right-hand panel and worst in the left-hand panel.

Fifth, the width of the intervals between the minimum and maximum values tells us how confident the experts were. In the left-hand panel, Expert 3 was confident (the interval was narrow) and accurate (the best guess was close to the truth), whereas Expert 5 was confident and inaccurate. Generally speaking, there was no clear relationship between confidence and accuracy.

Lastly, these were credible, socially accepted experts. They would have passed muster as expert scientists in a court or serving on a government panel dealing with the safety of earth embankments. All were qualified and respected members of scientific societies, attending an international scientific conference. No doubt each had a confident and plausible story to tell about how she or he arrived at an estimate and could defend the interval that she or he gave with the answers.

Misjudgements such as those reflected in these geophysicists' academic estimates may seem relatively benign, but experts'

mistakes may have more direct consequences. Fingerprint identification is based on expert judgement. Is it reliable?

In May 1997, an officer of the Boston Police Department was shot twice following a struggle with an assailant. The assailant ran, leaving behind the baseball hat he was wearing. He entered a nearby home, where he stopped to drink a glass of water. He then fled, leaving the gun, the sweatshirt he had been wearing and a thumbprint on the glass.[16]

The injured police officer later identified Stephan Cowans (Figure 1.2) from a photo array and then from a live line-up. The family in the house did not identify him. A fingerprint expert, however, matched the thumbprint to Cowans'. Having served six years in a Massachusetts prison, he was released in 2004 after the fingerprint evidence on which he had been convicted was contradicted by new DNA evidence.

Cowans' case is not unique. In March 2004, bombs exploded in the commuter train system in Madrid, killing 191 people. Brandon Mayfield, a US lawyer, was incorrectly identified from fingerprints taken from the crime scene.[17] Despite three FBI examiners and an

FIGURE 1.2 Stephan Cowans listening to testimony, prior to being convicted and spending five and a half years in Massachusetts prisons. Source: *Boston Herald*, Boston, MA.[18]

external expert agreeing on the identification, Spanish authorities eventually matched the prints to another suspect.

So how reliable is fingerprint evidence? The earliest large-scale study I found on the reliability of print examiners' decisions was published in 2011 in the prestigious journal *Proceedings of the National Academy of Sciences of the USA*.[19] The study gave pairs of fingerprints to 169 experts and asked them to determine whether the same person had made them, or not. A total of 32 per cent of the pairs were 'mated' pairs (from the same people) and 68 per cent were 'un-mated'.

The false-positive rate (the chance that fingerprint experts would falsely conclude two prints were the same) was satisfyingly low, at 0.1 per cent. The false-negative rate (the chance of falsely declaring two prints were different when in fact they were from the same person) was higher, at 7.5 per cent. A substantial number of the comparisons were judged by the experts to be 'inconclusive' or of no value. It is important to note that the experts knew they were being tested. We could reasonably assume that people unaware of such scrutiny may perform differently.

A UK-based study asked 27 experts to make a total of 2,484 judgements about pairs of fingerprints.[20] A quarter were controls, similar to the US-based study above. In the others, the experts were told fictitious emotional background stories that included murder and personal attacks, or they were shown disturbing photographs purportedly coming from the crime scenes from which the fingerprints were taken. Participants were not given the option of making inconclusive judgements. They had to decide either 'match' or 'no match'.

People were more likely to find a match between ambiguous fingerprints (an example is shown in Figure 1.3) if they had been exposed to emotional background stories or photographs. Participants found matches in 47 per cent of cases without emotional stimulus and in 58 per cent of cases when their emotions had been stirred.

This tells us that how we feel influences how we make technical judgements. In many circumstances, the people making judgements about fingerprints are aware of the background, and therefore may be susceptible to emotional distortions. Something as seemingly straightforward as judging fingerprints is error-prone and easily biased.[21] The prospects are not good for other situations in which experts are asked to make more difficult assessments in emotionally charged circumstances.

Expert misjudgements may have global consequences. As late as 2006, the International Monetary Fund (IMF) claimed that modern financial systems made the world a safer place. Their report on financial stability trumpeted the very instruments that led soon afterwards to global catastrophic financial collapse. It said, '*[i]n particular, the emergence of numerous, and often very large, institutional investors and the rapid growth of credit risk transfer instruments have enabled banks to manage their credit risk more actively and to outsource the warehousing of credit risk to a diverse range of investors. A wider dispersion of credit risk has "derisked" the banking*

FIGURE 1.3 Example of an 'ambiguous' pair of fingerprints.
Source: Dror et al. (2005). The author of the paper from which the image is sourced used a low-quality image to emphasise some of the real difficulties in matching prints.[22]

sector'.[23] Very few financial analysts saw the collapse coming. Most analysts at the time agreed with the IMF experts that the new financial systems were safe, well regulated and stable.

Less than two years later, in 2007, the system failed (Figure 1.4). Lest we forget, investment banks began to write down billions of dollars in mortgage-backed derivatives and other so-called toxic securities. In the US, Bear Stearns collapsed, Fannie Mae and Freddie Mac were taken over by the federal government, Lehman Brothers fell, Merrill Lynch was sold, AIG was saved, and a US$700 billion bailout bill was rushed into law.[24] The risks taken by the largest banks and investment firms in much of the Western world were so '*excessive and foolhardy*' that they threatened to bring down the financial system itself.[25]

Emotion and context may affect financial analysts and forensic scientists, but are other kinds of scientists immune? In the late 1990s, sociologist Lisa Campbell interviewed marine biologists and conservation experts and asked if they thought marine turtles could be

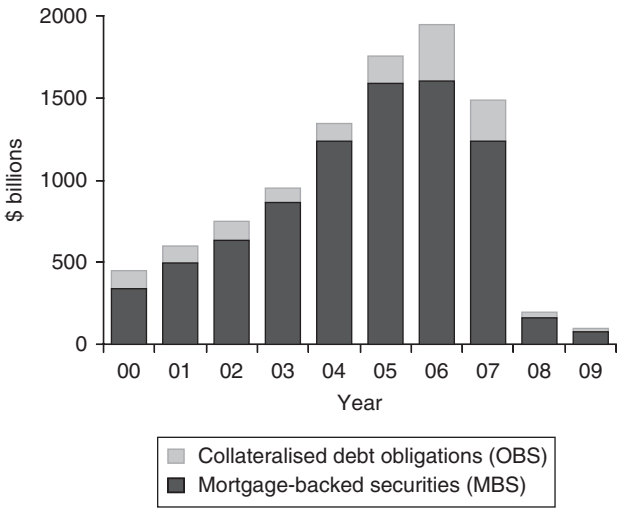

FIGURE 1.4 From economist Mark Zandi's 2010 testimony to the Financial Crisis Inquiry Commission. Securitisation occurred when banks and other financial institutions packaged various types of loan (including mortgages) into securities and sold them to global investors.[26]