European Review, Vol. 21, No. S1, S32–S39 © 2013 Academia Europæa. The online version of this article is published within an Open Access environment subject to the conditions of the Creative Commons Attribution license <http://creativecommons.org/licenses/by/3.0/>.

doi:10.1017/S1062798713000112

From Evidence to Understanding: A Precarious Path

DAVID J. HAND

Department of Mathematics, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK. E-mail: d.j.hand@imperial.ac.uk

Falsifiability is the cornerstone of science. However, Rutherford notwithstanding, almost by definition science functions at the limits of measurement accuracy and theoretical grasp, so that statistical analysis is central to scientific advance. This applies as much to physics as it does to psychology, as much to geology as to biology. I look at some of the potholes in the path of scientific discovery, showing how easy it is to stumble, and at some of the consequences for the scientific endeavour.

1. The Causes of Unreason

The late Stuart Sutherland began Chapter 22 of his book *Irrationality* by saying that: 'At a rough count, about a hundred different systematic causes for irrational thinking have been described.'[1] In his book *Why People Believe Weird Things*, Michael Shermer enumerates 25 of these.[2] But it seems to me that there are a few high-level categories of sources of unreason. The dividing lines between these categories are not sharp – they intersect and overlap to some extent – but nevertheless, as is so often the case in human discourse, a taxonomy can be useful. My categories are ideology, ignorance, gullibility, biology, behavioural economics, misunderstandings of science, and chance.

1.1. Ideology

Ideology will usually have been imbibed when young (recall the Jesuits' proud boast) but in some cases will have resulted from a conversion. Clearly, education is at the root of tackling this. But education cannot dispel the problem entirely. Our politicians are (mostly) educated, but they sometimes give the impression of working on the basis of ideology-based policy, rather than evidence-based policy.

1.2. Ignorance

Ignorance, in some sense, underlies it all. One would like to think that, as knowledge advances, so unreason arising from ignorance retreats. But perhaps that would be hoping

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

for too much: one cannot be knowledgeable about everything and despite the Royal Society's motto (*nullius in verba*: take nobody's word for it), sometimes one just has to take somebody's word for it.

1.3. Gullibility

There is a long history of people relaxing their critical faculties, from nineteenth century mediums, through a belief that Uri Geller could actually perform magic, to pseudo-religious cults. A combination of ignorance and gullibility provides a particularly powerful force. Instructive examples are Alan Sokal's meaningless spoof article 'Transgressing the boundaries: toward a transformative hermeneutics of quantum gravity',[3] which was accepted and published in the journal *Social Text*, and the study described by Peters and Ceci,[4] in which 12 papers by eminent authors that had already been published in psychology journals were rekeyed and resubmitted to the same journals, using fictitious authors from imaginary non-prestigious organisations, and which were then almost all rejected by the journals (nine of the 12 were not recognised as having been published before, and eight of these were rejected, although none on the grounds that it added nothing new).

1.4. Biology

Here I have in mind irrationality induced by psychotropic drugs or brain-damage. One is most familiar with this in the context of psychiatric illness, but studies have shown that this can be more pointed. If an epileptic focus develops in a particular part of the brain, it can induce religious experiences. And now we also have discussions of a 'God gene', which predisposes people to have mystic experiences.

1.5. Behavioural Economics

Researchers such as Daniel Kahneman and Amos Tversky have challenged the rational man of classical economics by drawing attention to numerous ways in which people behave irrationally. Kahneman and Tversky's research makes entertaining, if somewhat unsettling, reading since it describes behavioural characteristics from which it is very difficult to escape. Many of their examples fall into the category of misunderstandings of chance and probability, which I will discuss later. But non-probabilistic examples include the immediacy effect, in which one's most recent exposures influence attitudes and decisions, the halo effect, in which one has a tendency to generalise about someone on the basis of little information, and availability errors, in which one's understanding is distorted by how easy it is to bring something to mind.

1.6. Misunderstandings of Science

My sixth category is how science is misunderstood by the lay public. This misunderstanding itself occurs in several ways.

The first, and perhaps most fundamental, is the failure to recognise that science is a process, not a product. It describes a strategy for critically evaluating evidence, rather than taking it on trust. Science is all too often taught as the end product itself. Falsifiability is the cornerstone of science, and it is this notion that needs to be taught,

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

alongside the need for painstaking accumulation of evidence (although the conclusions themselves do also need to be taught).

A superficial illustration of how people fail to grasp that science is a process rather than a product is implicit in the idea that an experiment can 'fail'. Of course, an experiment can fail because the equipment breaks, or was not powerful enough to achieve its ends, and so on – but if the Large Hadron Collider 'fails' to detect the Higgs Boson it is not because the experiment itself has failed.

If evidence lies at the root of science, it is important to recognise that evidence accumulates, and science builds on that accumulated evidence. When a theory is found to be wanting, its replacement must explain both the new evidence, which casts doubt on the old theory, and the old evidence, which was consistent with the old theory.

One of the difficulties with which science has to contend is that, by definition, it functions at the frontier of knowledge. Very often, that means that it is working on the edge of detectability: it is surely a rare experiment in which a clear-cut result is found immediately. And this must be historically true as well, as scientific theory and measurement technology continue their leapfrogging of progress. In fact, I have formulated a sort of scientific version of the economists' efficient market hypothesis. However, given that the efficient market hypothesis has now been comprehensively discredited on various grounds, my efficient science hypothesis has a precautionary qualifier, which says: 'if it was easy, it would almost certainly already be known'.

I mentioned Lord Rutherford in my abstract. This was a reference to his comment 'if your experiment needs statistics, you ought to have done a better experiment'. I'm afraid, noting my comment about science being at the frontier of knowledge, I would respond by saying that 'if your experiment does not need statistics, then you have not been imaginative enough'. Of course, you will no doubt think it is exceedingly bold of me to question the wisdom and imagination of a scientific giant such as Lord Rutherford. But I take heart in the fact that he is also alleged to have said 'anyone who expects a source of power from the transformation of the atom is talking moonshine'.

I am sure that this location of science at the frontier between knowledge and ignorance accounts for much of the public misunderstanding of science, because it inevitably means that scientific results have a tendency to be positive and negative with equal probability. Far from the frontier, I can conduct an experiment and be confident that I can predict which way the outcome will go. At the frontier, the probability has to be almost a half. And this is one reason why we find reports one day saying that coffee is good for us, the next day it is bad; that one day we read we need an hour's vigorous exercise a day to keep healthy, and the next that ten minutes will do. And so on.

Furthermore, as science progresses, so it takes account of more subtle phenomena. The discovery that excess dietary fat was bad for us had to be modified in the light of the later recognition that there are different kinds of fat. Naturally this leads to scientific assertions changing as time passes: science accretes understanding gradually, and as evidence accumulates so people change their mind. Remember John Maynard Keynes, on being accused of inconsistency: 'when the facts change, I change my mind' (and his corollary question: 'what do you do?').

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

I will conclude my discussion of the public misunderstanding of science by telling you about an unexpected experience I had at the Dana Centre. The Dana Centre is a branch of the Science Museum, concerned with communicating science to the public. I had been asked to chair a discussion, in which I said just a few words of introduction and then invited questions from a lay audience of about 60 people. I summarised how science worked – and in particular the role of statistics in evaluating the match between theory and reality, the latter in the form of data collected from experiments. But I was taken aback when a woman in the audience asked why scientists should ever be trusted. After all, she said, scientists had to be funded by someone, and why would anyone ever pay unless they had a vested interest. I mumbled something about the disinterested nature of the research councils, but was uncomfortably aware of the historical record of commercial bodies funding research to reinforce a given position. My point is that this suspicion is justified. I was going to add that we need to overcome it, but that is not right. We need to encourage people like the questioner in their healthy scepticism, but not to the extent that they assume everyone is lying. Again, *nullius in verba* is all very well, but one must not take it too far.

2. Chance

There are quite a few instances of unreason arising from a misunderstanding of chance. Sometimes this is because chance phenomena can often be curiously counter-intuitive, but often it is simply because people do not have a good understanding of probabilistic notions.

2.1. The Prosecutor's Fallacy and Base Rate Ignorance

A series of recent papers, notably by John Ioannidis, have argued that 'most current published research findings are false'.[5,6] The argument, and the reasoning behind it, is not new, but it makes good headline material, which is why articles such as that in the *New Yorker*[7] caught on to it, and asked 'is there something wrong with the scientific method?' The answer to that question is, of course, 'no'. Ioannidis states (Ref. 5, p. 696) 'There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims' and he gives three references attesting to this 'increasing concern'. Now, there may indeed be such increasing concern but two of Ioannidis's three references are to his own papers, while the third is to a 2003 paper, so on the basis of that his assertion of 'increasing concern' would appear to be stretching things. I am reminded of Bjørn Lomborg's highly selective reporting of evidence in his book *The Skeptical Environmentalist*.[8]

The phenomenon Ioannidis has spotted is one of a number of biases with which statisticians are familiar, and which they take steps to ameliorate.

One of these is what is called the Prosecutor's Fallacy, or the error of the transposed conditional. This is the mistake of confusing the probability that the observed data would have arisen, given that a particular scientific hypothesis is true, with the probability that the scientific hypothesis is true, given the observed data has arisen. Now it is easy to get from one of these conditional probabilities to the other, via Bayes theorem (whether you are a Bayesian or not). But when you do this, not surprisingly, you need to take into account how many true and false null hypotheses there are. If I test thousands of drugs,

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

the vast majority of which really have no effect and just a handful of which have an effect, then amongst those I flag as apparently effective most will be useless. They will be scientific 'discoveries' which are in fact chance events, not of any real interest. This implies, in particular, that we should expect them to vanish on replication.

Ioannidis goes on to point out that smaller studies are less powerful (so there is less opportunity for false hypotheses to be rejected), that the same is true if the effect sizes are small (both of these points are obvious and well-known), and also that there are a number of conditions likely to lead to preferential selection of spuriously significant effects. These are discussed below.

2.2. Regression to the Mean

If I throw 3600 standard cubic dice, I might expect about 600 of them to show a 6. Let's focus our attention on those c.600. One-hundred percent showed a 6. Now what is likely to happen if I throw those 600 again? I'd expect about 100 of them, that is about 17% to show a 6. Does this mean that the characteristics of those 600 have changed: that previously they had a propensity to come up 6, and that this propensity has now vanished. Does it mean that 'cosmic habituation' has occurred? This, along with the term 'decline effect', is a term invented by Jonathan Schooler,[9] partly in jest, to describe the fact that it looked as if 'the cosmos was habituating to [his] ideas'[7] in that previous 'discoveries' he had made no longer seemed to be replicable. It occurs to me that Schooler is not the first to suggest hypotheses like this – I am reminded of Rupert Sheldrake's morphic resonance, which hit the headlines a few years ago.[10] And I note with some satisfaction, that morphic resonance seems itself to have been subject to the decline effect, as it appears to have vanished from the popular press.

In fact, of course, what has happened to my dice is an extreme example of regression to the mean. My selection process has chosen those that, by chance, produced values much higher than the mean. When I throw them again, they are equally likely, by chance, to produce high or low values. The apparent effect vanishes.

This phenomenon is ubiquitous. It explains why film sequels rarely do as well as the original, why sportsmen often deteriorate after an outstanding performance, and why an exceptional value observed in a scientific experiment may not be replicated when the experiment is repeated. But the important thing about it is that it is not causal in any sense. There is no mechanism relating the value of the first observation to the value of the second. It is simply that both are independently drawn from the same distribution.

2.3. Selection Bias

In my example of 3600 dice there was in fact nothing unusual about the c.600 dice that initially showed a 6. But in other situations there may be something distinctive about those chosen for inclusion in a study, and this distinctive aspect may be related to the aims of the study, to the extent that the conclusions are misleading. This phenomenon, generically called selection or selectivity bias is, like regression to the mean, ubiquitous. And it occurs in many guises. In science in particular, it occurs in the form of publication bias: the tendency to preferentially publish papers that appear to demonstrate an effect.

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

It is not a newly discovered phenomenon, so I think a lack of understanding of it qualifies as 'unreason'. Francis Bacon, writing in 1605, described the response of Diagoras, who, on having his attention drawn to a picture of sailors who had prayed and had survived shipwreck, asked 'but where are they painted that were drowned' after their vows?[11]

2.4. Coincidences

One popular source of unreason arising from chance is that of the coincidence. Coincidences are concurrences of events apparently so improbable that one suspects a hidden causal agency. Examples of such agencies are personal gods, superstitions, magic, and Jung's synchronicity. In fact, however, coincidences have natural laws, paralleling familiar statistical laws, such as the central limit theorem and the law of large numbers.[12] These laws include:

(1) *The law of total probability*: one of an exhaustive set of possible events must happen.
(2) *The law of truly large numbers*: with a large enough random data set, any specified data configuration is likely to occur.
(3) *The law of near enough*: events that are sufficiently similar are regarded as identical.
(4) *The law of search*: if it's not one of those, how about one of these?
(5) *The law of the lever*: a slight adjustment to a distribution can dramatically alter probabilities.
(6) *The law of the tortoise*: all journeys take place one step at a time.
(7) *The law of selection*: paint the target round the arrow.

3. Conclusion

Let me return to the *New Yorker* article.[7] This suggests that the universe may be changing about us: 'But now all sorts of well-established, multiply confirmed findings have started to look increasingly uncertain. It's as if our facts were losing their truth'. It goes on to note that 'If replication is what separates the rigor of science from the squishiness of pseudoscience, where do we put all these rigorously validated findings that can no longer be proved?' But that's the whole point of replication. What is being described here is the very fact that the attempt to replicate is failing, so suggesting that the initial 'discoveries' were not actually truths at all. That is the very essence of science.

I particularly liked the line in the article saying 'it appears that nature often gives us different answers.' Indeed it does. This is known as *random variability*. It arises from the fact that circumstances and people differ, and that they respond differently at different times and under different conditions. That life and the universe simply are not deterministic. That's why we need statistics to tease out the answers.

To give Lehrer credit, the article does refer to regression to the mean, publication bias, selective reporting in the first place, small sample sizes aggravating selection bias, perception biases in measurement, and data dredging. But we also need to consider the relative balance between the number of facts that 'look as if they are losing their truth'

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

and the number of facts that still look as if they are true. Those that appear to be 'losing their truth' are the ones that attract attention. We don't get up in the morning and shout 'hey, guys, aspirin still works!'

I think the tone of Lehrer's article demonstrates the failure of education to convey the nature of the scientific process. Either that or it is a journalist's attempt to whip up a story. Perhaps it merits a sequel. The headline could be 'Has reporter got it wrong?', and the content could contain sentences such as: 'if things are uncertain you don't know for sure what will happen' and 'a coin that comes up heads on one toss and tails on the next hasn't necessarily been switched from a two headed to a two-tailed coin'.

> I want to conclude by quoting two passages.
> The first is from the eminent biologist E.O. Wilson:
>
> > Today the greatest divide within humanity is not between races, or religions, or even, as widely believed, between the literate and illiterate. It is the chasm that separates scientific from prescientific cultures. Without the instruments and accumulated knowledge of the natural sciences – physics, chemistry, and biology – humans are trapped in a cognitive prison. They are like intelligent fish born in a deep, shadowed pool. Wondering and restless, longing to reach out, they think about the world outside. They invent ingenious speculations and myths about the origin of the confining waters, of the sun and the sky and the stars above, and the meaning of their own existence. But they are wrong, always wrong, because the world is too remote from ordinary experience to be merely imagined. (Ref. 13, p. 48)

For the second, I want to turn to a physicist: James Clerk Maxwell. He was one of the world's greatest physicists, ranking alongside Newton and Einstein. What not everyone may know, however, is that he also wrote poetry. One of his poems had the title *Notes of the President's Address*, and was related to his presidency of the British Association in 1874. A passage from it seems particularly relevant to this conference. It says (Ref. 14, p. 639):

> In the very beginnings of science, the parsons, who managed things then,
> Being handy with hammer and chisel, made gods in the likeness of men;
> Till Commerce arose, and at length some men of exceptional power
> Supplanted both demons and gods by the atoms, which last to this hour.
> Yet they did not abolish the gods, but they sent them well out of the way,
> With the rarest of nectar to drink, and blue fields of nothing to sway.
> From nothing comes nothing, they told us, nought happens by chance but by fate;
> There is nothing but atoms and void, all else is mere whims out of date!
> Then why should a man curry favour with beings who cannot exist,
> To compass some petty promotion in nebulous kingdoms of mist?
> But not by the rays of the sun, nor the glittering shafts of the day,
> Must the gods be dispelled, but by words, and their wonderful play.

I have taken the liberty of interpreting the last lines as saying that our hopes that the dark mists of unreason can be dispersed by logic and rationality, are unrealistic. Instead, we will have to resort to emotional persuasion. We must take on board the recent discoveries by the behavioural economists I mentioned earlier, and acknowledge the fact that we are irrational human beings. Regrettable though it may be, we must tackle irrationality using our rationally acquired understanding of how irrational people behave.

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112

References

1. S. Sutherland (2007) *Irrationality* (London: Pinter and Martin).
2. M. Shermer (2007) *Why People Believe Weird Things* (London: Souvenir Press).
3. A. Sokal and J. Bricmont (1998) *Intellectual Impostures: Postmodern Philosophers' Abuse of Science* (London: Profile Books).
4. D. P. Peters and S. I. Ceci (1982) Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behavioural and Brain Sciences*, **5**, pp. 187–195.
5. J. P. A. Ioannidis (2005) Why most published research findings are false. *PloS Medicine*, **2**, pp. 696–701.
6. J. P. A. Ioannidis (2008) Why most discovered true associations are inflated. *Epidemiology*, **19**, pp. 640–648.
7. J. Lehrer (2010) The truth wears off: is there something wrong with the scientific method? *The New Yorker*, 13 December. http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer?currentPage=all
8. B. Lomborg (2001) *The Skeptical Environmentalist: Measuring the Real State of the World* (Cambridge: Cambridge University Press).
9. J. Schooler (2011) Unpublished results hide the decline effect. *Nature*, **470**, p. 437.
10. R. Sheldrake (2009) *Morphic Resonance: The Nature of Formative Causation* (Rochester, Vermont: Park Street Press).
11. F. Bacon (1605) *The Advancement of Learning*, *Book II*, Section XIV 1, para 9.
12. D. J. Hand (2010) The laws of coincidence. In: Y. Lechevallier and G. Saporta. (Eds) *COMPSTAT2010, Proceedings of the 19th International Conference on Computational Statistics* (Springer), pp. 167–176.
13. E. O. Wilson (1998) *Consilience: The Unity of Knowledge* (Alfred A. Knopf).
14. L. Campbell and W. Garnett (2010) *The Life of James Clerk Maxwell* (Cambridge: Cambridge University Press; originally published 1892).

About the Author

David Hand has held chairs in statistics at Imperial College, where he is now Emeritus Professor of Mathematics, and the Open University. He is a Fellow of the British Academy, a Fellow of the Institute of Mathematics and its Applications, and Honorary Fellow of the Institute of Actuaries. He has won various prizes and awards for his research, including the Guy Medal of the Royal Statistical Society and a Royal Society Research Merit Award. He was awarded the OBE for Research and Innovation in the 2013 New Year's Honours List. He has written over 300 scientific papers and 25 books, including *Information Generation: How Data Rule Our World* and *Statistics: A Very Short Introduction*. He has served twice as President of the Royal Statistical Society.

Downloaded from http:/www.cambridge.org/core. Imperial College London Library, on 19 Oct 2016 at 12:37:12, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S1062798713000112