Collective Delusions, Blindness and Limitations of Imagination*

Greg Kochanski

http://kochanski.org/gpk

January 12, 2006

1 Introduction

The whole point of a statistics course, like this one, is to put limits on our knowledge. It helps answer broad scientific questions like:

- If I write this paper, will other people be able to get the same results, or will I look like a fool a few years down the line?
- Am I doing something with lasting value, or am I wasting my life?

Statistics helps answer broad questions like that by answering narrow questions like:

- How likely is it that I'm wrong?
- How well can I really measure that?
- Do I need more data?
- Does this theory plausibly explain that data?

Statistics is completely reliable within its own domain. If you understand the possible errors and you do the right tests, then you can trust the statistical results absolutely. The *art* is in understanding the possible errors. There are sources of uncertainty and doubt in science that are hard to quantify and where statistics can't save us.

The first is the "systematic error." Systematic errors happen when we have reached the limits of our understanding of a particular experiment or calculation. If we're lucky, we know that they exist, and may have some crude estimate for which way they might skew our results. However, by definition, we don't understand the possible errors well enough to apply statistical techniques.

 This work is available under http://kochanski.org/gpk/teaching/0401Oxford.

The second source of error can be called "Pathological Science." The term covers a lot of different breakdowns of the normal scientific process, typically involving honest self-delusion or wilful blindness.

Third, there is scientific fraud.

And finally, there are all possible degrees of sloppiness, laziness, confusion, and competence.

Pathological science and fraud are outside of science. Their practitioners ignore some of the basic rules that make science an interesting and useful description of our universe. Ultimately, the resulting papers are worthless. If they happen to get something right, it is by accident.

2 Systematic Errors

Some errors are built into any experiment. For instance, if you measure phone durations in a reverberant room, you will end up with a longer duration for the final phone because the sound will keep echoing around the room. You won't get a systematically lengthened duration for the medial phones, because the echoes of each phone get covered over by the direct sound of the next phone.

Thus, if you didn't understand this effect, you might believe that people lengthen final sounds more than they actually do. That would be a systematic error. Now, in practice, we know how to make this particular error small: we put absorbing materials on the walls and we record with a microphone fairly close to the subject, so that the echoes are faint compared to the direct sound. However, if you study speech in cathedrals, this artificial acoustic final lengthening might be more important than the linguistic final lengthening.

The problem is that systematic errors are a Catch-22 [Heller, 1961]: if you understand the problem, you can make it go away or correct for it, but if you don't understand the problem, you either won't think of it at all, or you will be wrong when you try to estimate how big an error it might cause. It's only at the edge of understanding where systematic errors are meaningful: you understand enough to realise it might be a problem, but not enough to fix it.

Note that theorists recognise the concept of systematic errors too. Calculations of what a theory means or what it would do often involve some approximations, which lead to errors in the final predictions. The systematic error of a calculation is built up from estimates of errors that different approximations cause. So, even if a theory is precise, applying that theory can be an imprecise process.

People often hope and assume that systematic errors are small, but that is not always the case. One dramatic example is shown in Figure 1. It shows various published values of the Hubble constant, one of the most important numbers in the field of astrophysics. The Hubble constant is a measurement of the rate at which the universe is expanding. For comparison with the figure, the current rate (circa 2004) is 71 km/s·Mpc [W.L. Freedman, 2001, Spergel et al., 2003] [1].

The result has dropped by a factor of ten since the first measurement. Despite what are now known to be huge systematic errors in the early values, the early error bars, as published, were not very large. The early workers completely missed some error sources, or at least vastly underestimated their magnitude.[2]

[1] The current value looks fairly solid because of the agreement of several techniques, two of which are completely independent.

[2] For the curious, the largest errors were in measuring the distance to other galaxies [Gingerich, 1996], and much of the problems seem to have been the result of the Malmquist bias [Mälmquist, 1920]. The Malmquist bias is simply

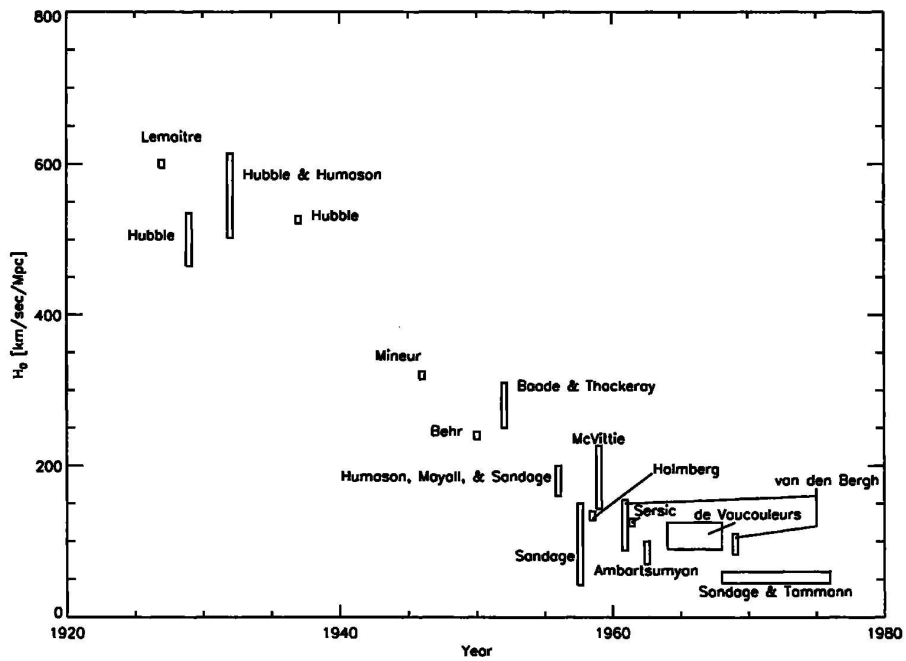

FIG. 1—Published values of the Hubble constant from Lemaître (1927) to the hardening of the battle lines. Rectangle dimensions are intended to suggest a range of values/uncertainties or a range of dates. Except where the errors listed below are larger, all uncertainties were claimed to be of order 10% or less (occasionally much less). A straight-line fit to the numbers from 1927 to 1965 or so would have suggested that the Hubble constant might have become negative within a decade or two (discovered by astronomy graduate students at Caltech in the 1960s and undoubtedly by many others). This did not actually happen. The numerical values represented are Lemaître 600, Hubble 465, 513, 535; Hubble and Humason 526; Mineur 320; Behr 240; Baade and Thackeray 280±30; Hubble, Mayall, and Sandage 180±20; Sandage 75 (+75,−40); Holmberg 134±6; McVittie 143–227; Sersic 125±5; van den Bergh 100 (+20, −12), 120 (+25,−20); Ambartsumyan 70–100; de Vaucouleurs 125, 100±10, 100±10; van den Bergh 95 (+15,−12); Sandage and Tammann 45–60.

Figure 1: The various published values of the Hubble Constant, from [Trimble, 1996].

It would not be completely misleading to say that the state of astrophysics in 1920 was similar to the state of linguistics today. People were just beginning to figure out what they were looking at, quantitative measurements were fairly new, and there were few theories that made real predictions.

By analogy, it's reasonable to believe that some linguistic facts that are thought to be well understood are, in fact, wildly wrong.

Another example is the age of the Earth around 1900. Various clever arguments were constructed which produced apparently solid numeric answers, all in the vicinity of 100 million years (see the discussion in Joly [c.1915]). For instance, one can measure how much sodium (salt) is carried into the sea by rivers. Then, assuming that the sodium accumulates in the oceans, one can compute how long it takes to make the oceans as salty as they are. One gets a maximum of 74 million years,

a selection effect: you see the bright stars and galaxies but not the faint ones, so you get a biased sample. A related error was the assumption that all Cepheid variables behave similarly: there are really two classes. The brighter class is more easily seen in external galaxies, but the fainter class is more common, and therefore dominated data in our galaxy.

even assuming the oceans started as fresh water. There seemed to be fairly good agreement among a variety of independent techniques.

Unfortunately, sodium does leave the ocean in various ways so that estimate is now known to be useless (E.g. water circulates through the sea-bed, especially near places where the sea floor is spreading like the mid-Atlantic ridge; some of the sodium is left behind). This turned out to be a substantial systematic error: It was an effect that the geologists of the day did not know about, and it was big. Effectively, they were interpreting good data in the light of a bad theory (the assumption that the salt just accumulates).

The other estimates for the age of the earth also turned out to be flawed; their apparent agreement was apparently a mixture of coincidence and wishful thinking. Better techniques for dating rocks, using the radioactive decay of atoms (incidentally pioneered by the same John Joly)[Off, 2005] eventually showed the Earth to be about 45 times older, $4.5 \cdot 10^9$ years.

3 Pathological Science

About every ten years, something goes wrong in the physical sciences. Recently (2002), it was faked nanotransistors [Newswire, 2003], and in 1999, it was a faked discovery of Element 118 [Schwarzschild, 2002]. Ten years before, in 1989, the world briefly watched the alleged phenomenon of Cold Fusion appear, perhaps cause a wild spike in world Palladium prices, and disappear [Carnell, 2001, Taubes, 1993, Huizenga, 1993]. Twenty-three years before that, it was Polywater [Franks, 1982]. Neither cold fusion nor polywater was a fraud; instead they were bad science perpetrated by people who were somehow blind to evidence against them.

These events are impossible to miss for someone in the field. Some, like Cold Fusion, are impossible to forget. The atmosphere of midnight experiments, of theorists racing explanations to publication [D'Angelo, 1989] then retracting as the experiments are refuted is intoxicating. A blizzard of claims and rebuttals fly by: it is scientific excitement at it's finest.

Yet, somehow, these dramatic events don't seem to happen in Linguistics. Why not? Are linguists less confused or self-delusive than physicists or chemists? Are linguists more honest? Probably not. Absent other evidence, one must assume the people are similar. One of several possibilities is that since there are fewer linguists than physicists, these events simply happen less often. Perhaps linguistics has just been lucky so far.

Of course, the cultures of linguistics and the hard sciences differ in several ways that might explain the difference. Primarily, the hard sciences have a strong belief that only one theory can be right in a given realm. A new theory thus has a large impact: if it exists and conflicts with existing work, a new theory or new data must invalidate something else. The culture of Linguistics has more tolerance for conflicting theories and data, so perhaps when these novel bad ideas appear, they are less disruptive. Perhaps linguists are used to big ideas with weak support, and take them with a pinch of salt. Another, and more disturbing possibility is that the bad science doesn't get corrected rapidly, so the bad ideas are never dramatically refuted. The answer is unclear.

The best and most concise survey of pathological science is the oldest, a lecture by Irving Langmuir in 1953 [Langmuir and Hall, 1953, 1989]. It's well written and doesn't require a deep understanding of the physics involved (though the events he describes may seem more pathological to the specialist). Langmuir studies several cases and lists some indicators of pathological science. Here's my summary of his list, modified in the light of the last fifty years.

- The maximum effect is produced by a barely perceptible cause, and the effect doesn't change much as you change the magnitude of the cause.
- The effect only happens sometimes, when conditions are just right, and no one ever figures out how to make it happen reliably. The people who can make it happen are unable to communicate how they do it to the people who can't.
- The effect is always close to the limit of detectability, or averages of many measurements[3] are needed to find it.
- There are claims of great accuracy, well beyond the state of the art or what one might expect.
- Fantastic theories contrary to experience are suggested. Often, mechanisms are suggested that appear no where else.
- Criticisms are met by *ad hoc* excuses thought up on the spur of the moment.
- Supporters are unable or unwilling to think about testing or disproving the effect. Tests that could lead to definitive disproof are never actually done.
- The implications of a theory or experiment are never extended outside its original domain. Supporters don't ask what implications it might have for neighbouring fields.

A single hit does not mark an idea as pathological, but multiple hits should serve to raise one's suspicions. This is a list primarily aimed at experiments, but many of the characteristics can also apply to theories.

Typically in pseudoscience, there is a hard core of believers. The number of supporters rises, peaks, and slowly declines, but is only brought to zero by retirement and/or death. Other relevant references include Park [2000], Turro [1999], Asimov [1962], Feynman [1985].

Thanks to Chilin Shih and Peet Morris for comments.

References

Isaac Asimov. My built-in doubter. In *Fact and Fancy*. Doubleday, New York, 1962. Also Pyramid, 1963 and Avon, March 1972. Originally in "The Magazine of Fantasy and Science Fiction," April 1961.

Brian Carnell. Giving pseudoscientists the benefit of the doubt. *Skepticism.Net*, April 2001. URL `http://www.skepticism.net/articles/2001/000003.html`. http://www.skepticism.net/articles/2001/000003.html.

Linda D'Angelo. MIT professor may publish papers on cold fusion. *The Tech*, 109:1, 15, April 1989. URL `Tuesday,April25,1989,http://www-tech.mit.edu/V109/N20/fus.20n.html`. Tuesday, April 25, 1989, http://www-tech.mit.edu/V109/N20/fus.20n.html.

[3] This one is not completely satisfactory, in that much real science depends on statistical analysis to find small effects. However, much pseudoscience has been committed by improper selection of data and bad sampling techniques, then hidden by a layer of statistical mumbo-jumbo. I think Langmuir's intent here is to capture J. B. Rhine and his parapsychology card experiments.

Richard Feynman. *Surely You're Joking Mr. Feynman!*, chapter Cargo Cult Science, pages 308–317. Bantam, 1985. URL http://www.physics.brocku.ca/etc/cargo_cult_science.html.

Felix Franks. *Polywater*. The MIT Press, Cambridge, MA, 1982. 0-262-06073-6.

Owen Gingerich. The scale of the universe: A curtain raiser in four acts and four morals. *Publications of the Astronomical Society of the Pacific*, 108:1068–1072, December 1996.

Joseph Heller. *Catch-22*. Simon and Schuster, New York, 1961.

John R. Huizenga. *Cold fusion: The scientific fiasco of the century*. Oxford University Press, New York, 1993.

John Joly. *The Birth-Time of the World and Other Scientific Essays*. E. P. Dutton and Company, New York, c.1915. URL http://www.gutenberg.org/etext/16614.

Irving Langmuir and R. N. Hall. *Voodoo Science: The Road from Foolishness to Fraud*. Theo Pavlidis and Ken Steiglitz, Princeton University, http://www.cs.princeton.edu/k̃en/Langmuir/langmuir.htm, 1953. URL http://www.cs.princeton.edu/~ken/Langmuir/langmuir.htm. Colloquium at the Knolls Research Laboratory, (General Electric), December 18, 1953. Hall transcribed and edited Langmuir's talk.

Irving Langmuir and R. N. Hall. Pathological science. *Physics Today*, pages 36–48, October 1989. Hall transcribed and edited Langmuir's talk.

K. G. Mälmquist. Final results from the hubble space telescope key project to measure the hubble constant. *Lund Medd Ser. II*, 22:1–39, 1920.

AP Newswire. Star researcher fired after committee finds fradulent work. *USA Today*, pages September 25, 2003, September 2003. URL http://www.usatoday.com/tech/news/2002-09-25-researcher-fired_x.htm. http://www.usatoday.com/tech/news/2002-09-25-researcher-fired_x.htm.

John Joly – Offaly Scientist. Offaly Historical and Archaeological Society, Bury Quay - Tullamore - Co. Offaly - Ireland, 2005. URL http://www.offalyhistory.com/content/reading_resources/famous_people/joly_john.htm. Checked Jan 2006.

Robert L. Park. *Voodoo Science: The Road from Foolishness to Fraud*. Oxford University Press, New York, 2000. ISBN: 0-19-513515-6.

Bertram Schwarzschild. Search and discovery. *Physics Today*, September 2002. URL http://www.aip.org/pt/vol-55/iss-9/p15.html. http://www.aip.org/pt/vol-55/iss-9/p15.html.

D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L.Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *The Astrophysical Journal Supplement Series*, 148:175–194, September 2003.

Gary Taubes. *Bad Science : The Short Life and Weird Times of Cold Fusion*. Random House, New York, 1993. ISBN: 0-394-58456-2.

Virginia Trimble. h_0: The incredible shrinking constant 1925-1975. *Publications of the Astronomical Society of the Pacific*, 108:1073–1082, December 1996. Systematic errors/Hubble Constant/Astrophysics.

Nicholas J. Turro. Pathological science. *http://www.columbia.edu/cu/21stC/*, 3(4): http://www.columbia.edu/cu/21stC/issue–3.4/turro.html, 1999. URL `http://www.columbia.edu/cu/21stC/issue-3.4/special.html`.

et al. W.L. Freedman. Final results from the hubble space telescope key project to measure the hubble constant. *Astrophysical J.*, 553(1):47–72, May 2001.