Scientific Method, Groundless Guesswork – Same Thing?

Victor Aguilar www.axiomaticeconomics.com

The Decline Effect

Most statisticians are hacks who readily admit that their job is to find data that supports the agenda of the journal they intend to publish in. But a few still have their souls and it bothers them that their amazing results, published in the most reputable of peer-reviewed journals, cannot actually be replicated. One such man was J. B. Rhine, the Goldilocks of the psychology profession: just flaky enough to be testing for E.S.P., but just serious enough to actually care.

Johan Lehrer of *The New Yorker* reports:

> *Joseph Banks Rhine, a psychologist at Duke, had developed an interest in the possibility of extrasensory perception, or E.S.P. Rhine devised an experiment featuring Zener cards, a special deck of twenty-five cards printed with one of five different symbols: a card was drawn from the deck and the subject was asked to guess the symbol. Most of Rhine's subjects guessed about twenty per cent of the cards correctly, as you'd expect, but an undergraduate named Adam Linzmayer averaged nearly fifty per cent during his initial sessions, and pulled off several uncanny streaks, such as guessing nine cards in a row. The odds of this happening by chance are about one in two million. Linzmayer did it three times.*
>
> *Rhine documented these stunning results in his notebook and prepared several papers for publication. But then, just as he began to believe in the possibility of extrasensory perception, the student lost his spooky talent. Between 1931 and 1933, Linzmayer guessed at the identity of another several thousand cards, but his success rate was now barely above chance. Rhine was forced to conclude that the student's "extra-sensory perception ability has gone through a marked decline." And Linzmayer wasn't the only subject to experience such a drop-off: in nearly every case in which Rhine and others documented E.S.P. the effect dramatically diminished over time. Rhine called this trend the "decline effect."*

This is clearly the result of publication bias: Journals only publish results that are interesting; that is, results that fly in the face of people's expectations. Initially, expectations are based on

logic. In Rhine's case, the expectation is that the subject will be accurate 20% of the time. But once some prestigious journal has given the statistician's results the weight of authority, those riding his coat tails can only be published if their results are different enough to be interesting but not so different that they embarrass the journal for having published the initial study.

This process will be investigated here and then an alternative methodology will be suggested.

The Culture of Neophilia

Alok Jha of *The Guardian* calls the publishing of only interesting results the culture of neophilia. He quotes Chris Chambers:

> *We have a culture which values novelty above all else, neophilia really, and that creates a strong publication bias. To get into a good journal, you have to be publishing something novel, it helps if it's counter-intuitive and it also has to be a positive finding. You put those things together and you create a dangerous problem for the field.*

This process can be simulated by a computer program that enforces two assumptions:

1) In the absence of any previous statistical studies, publication requires that one's results be – with 95% significance – more than double what logic would suggest.

2) Once a statistical result has been published, further publication requires that one's results be – with 95% significance – different but not less than 80% of the previous result.

Set the sample size at some arbitrary number larger than thirty (to avoid having to use Student's *t* distribution) and generate that many "observations" with your pseudo-random number generator. Verify that $z > 2.807$ where z is the average minus twice the expected (logical) value, all divided by the sample standard deviation. If it is not, then just try, try again with another batch of "observations." This try-try-again procedure is known as data dredging, which will be the topic of a later section.

But the culture of neophilia does not alone explain the decline effect, only the sudden appearance of positive results. The subsequent decline is explained by the Manchurian effect.

The Manchurian Effect

For our initial study we wanted an average more than double the expected value; logically, it seems like it would have been just as interesting if Linzmayer had guessed less than 10% of the Zener cards but, psychologically, it was guessing over 40% that struck researchers as interesting. However, for the follow-up studies, different can be in either direction. Sameness is boring, but if there is a statistically significant difference in the averages, people focus on exactly what that number is and ignore the dubiousness of the whole endeavor.

I will here coin a name for this: the Manchurian effect. Recall in the movie *The Manchurian Candidate* how, by getting people to debate exactly how many communists there were in the Defense Department (57; a number inspired by the Heinz sauce), the hen-pecked Senator Iselin got them to overlook the fact that it was his own wife who was the communist agent.

For our computer simulation, just check that the difference between the averages of the two samples divided by $\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}$ exceeds 2.807 where $\frac{s_x^2}{n}$ is the variance of the previous study divided by its sample size and $\frac{s_y^2}{m}$ is the variance of the current study divided by its sample size.

Also – and this is very important – verify that the ratio of the new average and the old average is not less than 80% or greater than 125% to avoid embarrassing the journal. (The reciprocal of $\frac{4}{5}=80\%$ is $\frac{5}{4}=125\%$.) Logic is difficult to overcome, which is why we need our initial average to be a whopping double that of what logic would suggest. But once the initial groundbreaking work has been published, logic has been discarded in favor of statistics, so publication only requires something different, not something double. However, unlike logicians, statisticians have feelings, so we cannot be too different lest we step on any toes. Thus, the Manchurian Effect must be modified to assure that the results are different, but not *too* different.

Combining the culture of neophilia with the Manchurian effect modified to avoid stepping on toes fully explains the decline effect.

Data Dredging

Of course, using your pseudo-random number generator to produce "observations" that are publishable is fraud. Tsk. Tsk. Dirk Smeesters can inform us of what happens in cases where outright fraud is exposed; after being caught fabricating data, he can no longer pass GO and

collect $200. Much is made of the fact that he was caught by other statisticians noting how neat his data fit his conclusions; as though this lends the profession more credibility than if they wait for disenchanted students to snitch on their advisors. But he was caught only because he was a lousy cheater; if he had used the techniques taught here, he would have won this game.

Fraud cases such as that of Dirk Smeesters, Diederik Stapel, Marc Hauser, Karen Ruggiero and Naoki Mori get a lot of press, but they also leave people with the impression that the process is sound if it is not subverted by these few bad apples. But *all* statisticians engage in data dredging; that is, abandoning a study if it does not appear interesting and just starting over with new data. Example: Linzmayer was not Rhine's first subject; he was Rhine's first *interesting* subject. Because the data that actually appear in the published study are all real, this practice is not considered fraud. But the effect is the same. It is also more robust because people are not good at fabricating random data; the run test for randomness (look it up) pegs them every time.

John P.A. Ioannidis reports:

> *Usually it is impossible to decipher how much data dredging by the reporting authors or other research teams has preceded a reported research finding.*

Ioannidis' paper, provocatively titled *Why Most Published Research Findings Are False*, has been cited a lot, not because the title is provocative, but because the conclusion is not. After methodically describing the widespread problem of data dredging, Ioannidis concludes by limply assuring us that the scientific method is sound and all that is needed to fix the problem is bigger studies with more data. Statisticians everywhere breathed a sigh of relief on reading:

> *Better powered evidence, e.g., large studies or low-bias meta-analyses, may help… most research questions are addressed by many teams, and it is misleading to emphasize the statistically significant findings of any single team. What matters is the totality of the evidence.*

Statisticians make more money doing large studies than small ones and meta-analysis (combining the data from several teams) allows more statisticians to get their snouts into the same trough. When people say "every opinion counts," what they really mean is that nobody's opinion counts. This is where economics has gone; reporters poll a bunch of economists on their predictions for the percent change in some statistic and then take the arithmetic average. It should be the geometric average, but averaging at all is an open admission of blind ignorance.

As for Ioannidis advocating larger studies, we need only recall Hitler's famous dictum that simple people more readily fall victim to a big lie than a small lie. *Alok Jha* quotes Ferric Fang:

> *If someone says they did a 15-year clinical study with 9,000 subjects and they publish their results, you may have to take their word for it because you're not going to be able to run out and recruit 9,000 patients of your own and do a 15-year study just to try to corroborate something that somebody else has done.*

Ioannidis is suggesting no fundamental change; it is still the same scientific method that is so heavily promoted at Science Buddies, Science Bob, Science Made Simple and other kiddie sites. What the grown-ups need is not to throw more money at this childish game with the important-sounding name, but an alternative to the scientific method.

An Alternative to the Scientific Method

Ragnar Benson wishes to determine the correct mixture of ammonium nitrate and nitromethane to achieve complete detonation. A noble endeavor! But let us see how Mr. Benson goes about finding this correct mixture. Benson (1992, p. 136) writes:

> *Despite almost driving our family into poverty by my many costly experiments, I still do not feel I have all the answers pertaining to this process. My experiments indicate that one should use slightly less than one-third nitromethane by volume, but this seems to vary from one gallon of nitromethane to the next and from one bag of ammonium nitrate to the next.*
>
> *Unfortunately, I know of no formula that states precisely how much nitromethane to use. As a rough starting point, try one part nitromethane to three parts of ammonium nitrate by volume or two parts nitromethane to five parts ammonium nitrate by weight.*

In sharp contrast, real science is based on the axiomatic method. Real scientists do not just randomly mix reactants together, stick a blasting cap in the resulting glop and see what – if anything – happens. Calling such guesses "hypothesis" does not make this activity scientific.

Real scientists reason from axioms; in this case, the conservation of mass. Chemical reactions – even very energetic ones like the detonation of high explosives – do not destroy mass. This claim is an axiom; that is, a proposition that is assumed without proof for the sake of studying

the consequences that follow from it. An ammonium nitrate/nitromethane explosion just converts a solid and a liquid into three hot gasses; water vapor, carbon dioxide and nitrogen.

The important point here is that we must *assume* that there is exactly the same number of hydrogen, carbon, nitrogen and oxygen atoms in the reactants as there are in these hot gasses. Because water vapor, carbon dioxide and nitrogen are already abundant in the atmosphere, there is really no way that we could catch the products of the explosion, racing away from the blast at 7000 m/s or more, and weigh them. In some endothermic reactions that do not involve gasses, it may be possible to weigh the reactants and products, but even then we must remember that a mechanical scale is limited to only about three significant digits of accuracy. Yet I claim that mass is conserved *exactly*; to 23 significant digits if we could actually count every molecule in a mole of material. Such a bold assertion *must* be regarded as an axiom. There is no way to test it in the great majority of cases and even when it can be tested the accuracy of our measurements is 20 orders of magnitude short of proving our assertion.

So, having established the axiomatic nature of our line of reasoning, let us apply our axioms one-by-one and see what can be deduced from them, without recourse to any experiments or hidden assumptions. (And, as an added bonus, without blowing ourselves up or getting tossed into prison for possessing illegal explosives.)

NH_4NO_3	CH_3NO_2	→	H_2O	CO_2	N_2	(1)
x	y		?	y	?	

Let x be the number of ammonium nitrate, NH_4NO_3, molecules and let y be the number of nitromethane, CH_3NO_2, molecules. From the axiom that there are as many carbon atoms before as after the reaction, we can deduce that there are y carbon dioxide molecules, as shown in equation (1).

NH_4NO_3	CH_3NO_2	→	H_2O	CO_2	N_2	(2)
x	y		?	y	$\frac{2x+y}{2}$	

From the axiom that there are as many nitrogen atoms before as after the reaction, we can deduce that there are $\frac{2x+y}{2}$ nitrogen molecules, as shown in equation (2).

NH_4NO_3	CH_3NO_2	→	H_2O	CO_2	N_2	(3)
x	y		$\frac{4x+3y}{2}$	y	$\frac{2x+y}{2}$	

From the axiom that there are as many hydrogen atoms before as after the reaction, we can deduce that there are $\frac{4x+3y}{2}$ water molecules, as shown in equation (3).

$$3x + 2y = \frac{4x+3y}{2} + 2y \quad (4)$$

From the axiom that there are as many oxygen atoms before as after the reaction, we can deduce equation (4).

$$2x = 3y \quad (5)$$

By multiplying both sides of (4) by two and then subtracting $4x + 4y$ from both sides, we can deduce equation (5), which is all the information we need to deduce equation (6), our result.

$3NH_4NO_3 + 2CH_3NO_2 \rightarrow 9H_2O + 2CO_2 + 4N_2$ (6)

And so, using only deductive logic based on the axiom that mass is conserved, we have found the formula for an ammonium nitrate/nitromethane reaction.

But we still do not know the relative weights of the two reactants. For this we must introduce a second axiom: The atomic mass of an element is approximately equal to the sum of its protons and neutrons. This is the simple high school model of the atom without isotopes. Specifically:

Hydrogen has one proton and no neutrons.	The atomic mass of hydrogen is 1.
Carbon has six protons and six neutrons.	The atomic mass of carbon is 12.
Nitrogen has seven protons and seven neutrons.	The atomic mass of nitrogen is 14.
Oxygen has eight protons and eight neutrons.	The atomic mass of oxygen is 16.

Atomic mass is actually the weighted average of the isotopes minus a tiny mass deficit for what is converted into binding energy – the results found in this paper have only three significant digits of accuracy – but the important point is that both this basic high school model of the atom and the more complicated professional model are axiomatic systems.

Nobody has ever seen an atom and nobody ever will. It is smaller than the wavelength of light one would have to reflect off it onto the lens of one's microscope to observe it. What divides the post-doctoral researcher from the high school chemistry teacher is not that the former has a more powerful microscope for observing atoms – no such microscope exists – but that the former has a more powerful, though also more complicated, axiomatic system.

At this point in the argument, the empiricist will invariably start screeching that he has "refuted" our entire theory by carefully weighing a mole of oxygen and finding that it weighs 15.9994 grams, not 16 as that misguided axiomatist claimed. Well, so it does. But I already noted that the results in this paper were only accurate to three significant digits. If three significant digits of accuracy are sufficient for practical applications, there is nothing wrong with using simple axioms that are known to be approximations. For instance, I developed an Android application for mortar fire control based on three axioms, one of which is that the Earth is flat; that is, gravity is everywhere pointed downwards. In point of fact, the Earth is not flat; it is a sphere. But knowing it to be a sphere is no recommendation to stand calmly on the receiving end of my mortar fire. Howitzers should not be fired at targets over the horizon based on such simplifications, but my app is more than sufficient to hit a Shilka lurking behind a building six blocks away.

So, having dispensed with the empiricist's inevitable criticism that our axioms are not perfect, let us return to our intrepid chemist, Mr. Benson. From our second axiom, we can deduce that the atomic mass of three ammonium nitrate molecules is 3(4 + 28 + 48) = 240 atomic mass units, AMU. Furthermore, we can deduce that the atomic mass of two nitromethane molecules is 2(3 + 12 + 14 + 32) = 122 AMU.

Therefore, using only deductive logic based on two reasonable axioms, we have found that we need 122 parts nitromethane to 240 parts ammonium nitrate, by weight, to achieve complete detonation. This result can be rounded off to one part nitromethane to two parts ammonium nitrate, by weight. And if it fails to detonate? It probably absorbed moisture from the air. Try again. Perfect equations and aesthetic axioms always supersede anecdotal evidence.

Clearly, the axiomatic method is vastly superior to the empirical method of conducting random experiments until something resembling a result appears. But empiricists never learn. Here, Mr. Benson (1992, pp. 139-140) describes a further line of research, conducted with his usual methodology of randomly mixing reactants together to see what happens.

> *The tip-off to a possible solution came while I was researching World War I's Messines Ridge sapper attack… Britain's World War I explosives manufacturers added finely ground aluminum powder to this explosive, called ammonal, to boost its brisance… Having made that discovery, I began to experiment with powdered aluminum. I added it to the ground ammonium nitrate before adding the nitromethane. At a level of about 5 percent (or about 20 grams) mixed thoroughly into 430 grams of NH_4NO_3, the effect was dramatic.*

I leave it as an exercise for the reader to derive equations (7), (8) and (9) employing only deductive logic – no experiments – based on our two axioms. (Note that aluminum has 13 protons and 14 neutrons.) The atomic masses are written below the reactants.

$$6NH_4NO_3 + 2CH_3NO_2 + 2Al \rightarrow 15H_2O + 2CO_2 + 7N_2 + Al_2O_3 \quad (7)$$
480 122 54

$$9NH_4NO_3 + 4CH_3NO_2 + 2Al \rightarrow 24H_2O + 4CO_2 + 11N_2 + Al_2O_3 \quad (8)$$
720 244 54

$$12NH_4NO_3 + 6CH_3NO_2 + 2Al \rightarrow 33H_2O + 6CO_2 + 15N_2 + Al_2O_3 \quad (9)$$
960 366 54

Conclusion

It is truly sad that Mr. Benson nearly drove his family into poverty, spending a lifetime and squandering a fortune attempting to accomplish with random experiments what an axiomatist could have achieved in thirty minutes at no cost. My heart breaks for his wife and children! If I were a Liberal, I might advocate a program of socialized explosives so that poor folks could pick up a brick of C4 at the food bank, just as they can now pick up a loaf of bread or a can of soup.

Lest others drive their families into poverty with their many costly experiments, let us be rid once and for all of the idea that doing science means randomly mixing reactants together. In spite of all the highfalutin talk about these random mixtures representing "hypothesis" that are to be tested empirically, accompanied by a barrage of statistical "data" if the experiment is conducted more than once, it is obvious that this is not science. Real scientists employ the axiomatic method.

REFERENCES

Benson, Ragnar. 1992. ***Big Book of Homemade Weapons***. Boulder, CO: Paladin Press