Why good science is good business

Ideas developed by a twentieth-century philosopher, when applied by statisticians, can help companies improve decision-making, says **Simon Raper**

Main image: ismagilov/Bigstock.com

"Science", wrote the philosopher Karl Popper, "must begin with myths and with the criticism of myths".[1]

In the business world, where I work, vast amounts of money are invested each year on the strength of firmly held beliefs that started life as little more than water-cooler anecdotes or out-of-context quotes from obsolete research papers. These are the boardroom myths on which businesses are built.

Our job, as scientists and statisticians, says Popper, is to subject the prevailing wisdom, and indeed our own theories, to unrelenting criticism. Since business decision-makers act on their beliefs, this focus guarantees that our work will be relevant. The alternative, to sit in our mathematical ivory towers attempting to build the world from the bottom up, is not only misguided but produces work that will almost certainly be ignored.

I think Popper's analysis of what science is, and how it works, shines a light on the practical use of statistics in the workplace. It shows us what we ought to be doing and why, as well as the traps that we will almost inevitably fall into. In what follows I want to explore three of his ideas and examine what they mean for us in practice.

Simon Raper is a statistician and the founder of Coppelia, a London based company that uses statistical analysis, simulation and machine learning to solve business problems

When science is no longer scientific

The first idea is his answer to a very specific problem: How can we tell a science from a pseudoscience? Why is biology a science and astrology not? After all, as Popper points out, both make inferences from empirical data – and surely that is what science is supposed to be about?

It is worth pointing out that the problem of demarcation, as it is known, was a big deal in the first half of the twentieth century. We may have to contend with homeopathy, crystals and the cleansing of auras, but back then Popper was facing down not only the worst excesses of Marxist and psychoanalytic thinking but also the pseudoscience of racial types that props up fascism. Popper felt strongly (and with good personal reasons) that he would be doing the world a service by demonstrating that these things were not sciences.

His very famous answer is that what sets apart science from pseudoscience is the possibility of falsification. Good science sets out well-defined theories and, in doing so, makes clear what might prove these theories wrong.

My theory that all worms have backbones, though a bit ridiculous, is nevertheless a good scientific theory. It predicts that if you slice open a worm you will find a spine – something that is easily falsified. Pseudoscience does not set out its wares in this way. In fact, one mark of a pseudoscience is that there is always an answer to accommodate what might at first look like a refutation.

As Popper points out,[1] there is no conceivable form of human behaviour that can contradict the theories of psychoanalysis. If, as a patient, I vehemently deny that my behaviour is

motivated by repressed sexual urges then this only counts as proof that the repression mechanism is working as expected.

Another strategy of pseudoscience is to make claims that are so general and vague that prediction rarely fails. We see this most obviously in astrology when it predicts a stroke of luck or an unexpected encounter.

But falsification, according to Popper, is not just a concept we can use to demarcate science from pseudoscience, it is the engine of science itself, it is how it moves forward. This takes us to the second idea.

Falsification as the engine of science

Prior to Popper it was generally thought that science moved forward by a process called inductive inference. It goes like this:

I kick the table. The book falls off.
I kick the table. The book falls off.
I kick the table. The book falls off.

Then, with sufficient examples, I make a jump, a leap of inference, to the conclusion if A then B:

If I kick the table then the book will fall off.

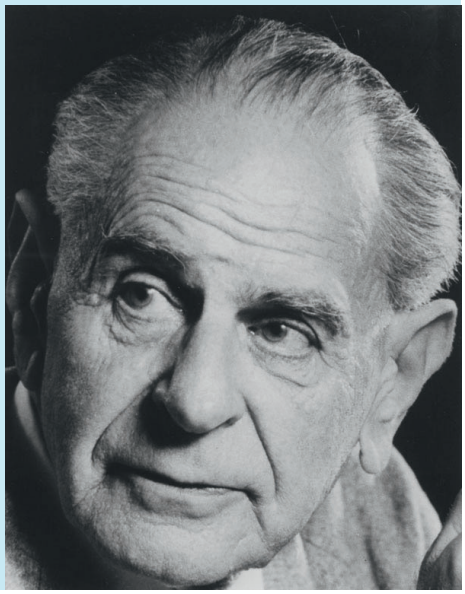
Popper points out that this is not how science works. We will not find our laboratories full of scientists endlessly testing the positive ramifications of a theory. Yes, they do repeat experiments, but only very targeted ones – those that could refute their theories.

In fact, looking for endless confirmations of their own theories is exactly what the practitioners of pseudoscience do. Ah, the Freudian psychoanalyst says, yet another example of a repressed sexual urge confirming the theories of psychoanalysis.

For Popper, science works like this:

1. I form a theory (this is a creative, imaginative process) that in some way betters existing theories in that it solves problems that they do not.
2. I then throw everything I have at this theory. I try my hardest to refute it. If it withstands my best efforts, and the efforts of others, then it becomes the current theory. Current until it is eventually knocked from its pedestal by a better theory. For Popper there is never (and, for good logical reasons, never can be) a final theory. Popper described this as a process of conjecture and refutation (also the title of his most accessible book).

Our job, says Popper, is to subject prevailing wisdom and theories to unrelenting criticism. Since business decision-makers act on their beliefs, this focus guarantees that our work will be relevant



LSE Library/flickr.com

Now this process of conjecture and refutation is very much a social, collaborative process, embedded in the practical world, and this brings us to the third idea, the idea with which we began.

Science must be a criticism of existing theories

This is the idea that in any enterprise involving knowledge we always begin where others left off. The pioneers of chemistry started from the theories of alchemists. The science of thermodynamics began as a challenge to Aristotle's belief that "nature abhors a vacuum". The history of science is full of such examples.

This led Popper to the conclusion that science ought to proceed in a piecemeal fashion, working with what we currently have rather than sweeping the table clear each time. (Here he was attacking the utopian thinking of, among others, Marxists and socialists who sought a complete re-engineering of society with what he regarded as disastrous consequences.)

It might seem blindingly obvious that current theories should be the focus of science and statistics, but the truth is that applied statistics in the business world (and I suspect outside it too) is mostly a question of navel-gazing – a constant bettering of the tools and techniques for their own sake, while powerful beliefs and prejudices about business, about economics and about human nature lumber about the boardrooms unchecked.

So that is it. Three ideas:

1. Falsification as a way of demarcating proper science.
2. Science proceeds by a process of conjecture and refutation.
3. Science must be a criticism of existing theories.

How does this apply to us statisticians in our day-to-day work, facing a tough deadline with just our statistical knowledge and a handful of .csv files?

Bad Bob

Meet Bob. Bob is a parody, a caricature, a completely fictitious person who has been stitched together from the worst examples of statistical malpractice I have witnessed.

Bob has been instructed by his superior to evaluate the success of a TV marketing campaign for a client. Bob has no theories about the effectiveness of TV marketing when applied to the client's product. In fact, he believes that having such ideas is way above his pay grade. For him the objective is to churn out a number – maybe return on investment, or something a little more pragmatic, like TV's recommended share of budget for the next ad campaign. His company has a method of doing this kind of analysis. It could be based on analysing survey

data, it could be econometric modelling, it could involve some kind of direct tracking of customers. The method does not matter. What matters is that Bob blindly and unquestioningly accepts the method.

Bob runs some code and out pops a number, only it is very low and this worries him. The client, he believes, will not be happy. But he also knows that there are some things that he can try which may yet produce a high number. If he is running an econometric model he can add or take away explanatory variables, or change the functional form. If he is looking at a survey he can choose a subset of the data or look at another variable. There are usually hundreds, even thousands, of possibilities to try, so it is not hard to find one that yields, probably now by chance alone, a much more palatable number. A lot of this is happening at a subconscious level, so Bob sees it all as yet another confirmation of the idea that marketing always works.

The problem here is that in setting things up so that the rules of the game are not defined in advance, Bob has made sure that the implicit theory, "marketing always works", is as good as unfalsifiable. In statistics, we have various names for this crime – overfitting, data dredging, *p*-hacking – but the essential failing is the one noted by Popper: by not specifying, prior to testing, a testable theory, statistics has been turned into pseudoscience.

Looking for endless confirmations of their own theories is exactly what practitioners of pseudoscience do

Furthermore, if Bob is questioned on the validity of his findings he reaches instinctively for some apparently incomprehensible maths (incomprehensible both to him and the questioner) in much the same way that, in Popper's day, you might have heard a psychoanalyst reach for the psychobabble. For Bob, it is the fact that the maths is hard that makes it credible. He sees no need for his model to address the current thoughts and beliefs of the industry, which he views as mere opinion. These are swept aside and replaced by "hard maths stuff".

Beyond Popper

In a way that is somehow gratifying, given the ideas we have discussed, falsification turned out to be a far from perfect theory. Popper's ideas about science were refined and attacked by later philosophers. Most of the criticisms were centred around the problem of deciding which beliefs to give up should a theory be falsified. Should we reject the theory itself or one of its many background assumptions? Newton's theory of gravity and Bohr's theory of the atom were all falsified against the background theories of their day.[2] Thomas Kuhn and then Imre Lakatos tried to solve this problem by constructing alternative accounts which saw science as advancing against a background of shifting thought paradigms and research programmes, respectively. More recently, Mary Midgley has used a critique of science's background assumptions to attack the very idea that science should be separated from other disciplines.[3] We should note, however, that nearly all the subsequent thinkers place an even greater emphasis than Popper on the idea that *science is embedded in myths*. We are never starting from a blank slate. We are always unknowingly in the shadow of the latest scientific paradigm.

In statistics, the rise in popularity of Bayesian methods has brought the spirit of Popper's approach closer to home. Indeed, Bayesian inference can be seen as a much more general theory in which data and model modify the probability of beliefs rather than simply ruling them out. Popper's criterion of falsifiability is then just a special case where the rejection of the belief is entailed by the evidence with absolute certainty (thanks to Eliezer S. Yudkowsky of the Machine Intelligence Research Institute for pointing this out).

The current controversy over the use of *p*-values in published research revives several of Popper's themes: the way that a lack of transparency over which hypotheses are being tested can lead statisticians into pseudoscience, and the need to present the results of statistical tests not in isolation but in the context of all other evidence.

Before you object that Bob is obviously unqualified and should be quickly shown the door, I have to point out that I've met many Bobs with impressive qualifications. But enough of Bob.

Best behaviour

Now meet Jo. Again Jo is somewhat idealised, behaving in a way that is more perfect than we can usually manage (and that includes me). What does Jo do differently?

First, Jo challenges the brief from the client. She wants more information. What is the reason for evaluating the marketing campaign?

She is told that the new marketing director believes that the client is overspending on TV. They are thinking of looking at more adventurous options instead, perhaps a multichannel approach involving print, web and social media.

Jo then asks the million-dollar question: what do you think the client would do if they did not have our help? This tells her the nature of what we might call, following Popper, the current conjecture, the one that is being taken most seriously. After some more questioning, Jo establishes that in the absence of any contrary evidence the client will cut the TV budget to £120 000.

Now this certainly is not pseudoscience since we have a claim that is falsifiable – any spend over £120 000 delivers no (or negligible) additional return. As in all science, the hard part is now to design an experiment to test this claim – or what we might call the "null hypothesis". Note that this is not how Bob saw it. It never crossed his mind that he was designing an experiment. He worked by rote, cranking out a standard model from a template with the usual suspects as variables.

But Jo thinks very carefully about this particular client, product and situation. To isolate the effect of advertising on sales she will need to control for other factors such as the weather, competitor activity and price changes. In other words, she will need to a build a model. Unlike Bob, this is not just a matter of working zombie-like through some highly standardised modelling process designed to fit all situations and therefore none. Her model is carefully crafted to reflect aspects of the world that she and others consider important in this particular case. This means building in the background assumptions and current theories of the advertising industry – theories which she has taken the time to research. Some of these assumptions will be tested in the process of building the model; others – for example, study findings about the psychological effects of advertising – will have to be accepted for the time being. These are the "myths" that she is not attacking at this point – that is fine, so long as she tracks them as assumptions.

TABLE 1 Karl, Bob and Jo: comparing two different approaches to statistical analysis, and how both measure up against Popper's scientific ideals

Concept	Bad practice (Bob)	Good practice (Jo)
Falsification demarcating science from pseudoscience	Unthinkingly modifies the model so it can accommodate any theory, including the one he thinks will most please his client. Sees the result as yet another confirmation of the idea that marketing always works. Thus practises pseudoscience.	In advance, formulates a theory that can be falsified by the data (the theory that spending over £120 000 delivers a negligible additional return).
Science proceeds by conjecture and refutation	Does not enquire into the current theory. Does not propose a new one.	Designs an experiment that will test the default beliefs of the client, beliefs which they will act on in the absence of any evidence to the contrary. If her experiment knocks down the current theory, she will attack the new theory with just as much vigour to check that it stands up to scrutiny (for now!).
Science must be a criticism of existing theories	Does not understand how his analysis or model relates to the current theory (marketing spend is too high) and so cannot relate it back to the decision. Hence falls back on maths technobabble when questioned.	Takes great care to understand how her analysis relates to the beliefs held by her client and the industry. Her model incorporates some of these beliefs and challenges others. This makes it easy for her to explain how her findings relate to decisions.

True to the Popperian view of things, Jo has been very diligent about attacking her own position. She has sensitivity-tested the above-mentioned assumptions (are there radical changes to the conclusions if she changes them?). She has also gone to great lengths to understand the accuracy of her conclusions. She is aware, for example, that the more models she tries and the more hypotheses she tests, the more likely it is that an apparently significant result will be due to chance alone. She also notes the statistical hazards involved in peeking (using the data to generate hypotheses and then carrying out the experiment on the same data). She tracks any obvious violations of modelling assumptions and assesses their likely impact. She is wary of problems that might exist with the accuracy of the data and how they might bear on her conclusions. And, most importantly of all, since she is not, like Bob, operating in a vacuum: she weighs the results of her analysis against what else is out there – the common sense of the business professionals, other comparable studies, new data that has arrived since the work began – before making her recommendations.

In contrast to Bob (who is terrified of delivering a low value), Jo cannot lose. If the null hypothesis (spending over £120 000 delivers negligible return) holds, and she explains just how much pressure she has put on it, then the client is happy. They feel that they have done the right thing. Everything possible has been done to question their gut feeling. Even Jo's caveats, couched as many of them are in the language of the client's business and of common sense, only reinforce the impression that she has thought it all through, leading the client to trust her expertise when it comes to those statistical concerns that cannot be so readily expressed in layman's terms. If, on the other hand, evidence stacks up against the null hypothesis then she has put a big question mark against a strategy that could end up costing them thousands (and perhaps the marketing director his job).

Philosophy's investigation into the methods of science did not end with Popper, but he set the terms for all future debate

The appraisal

Jo and Bob are clearly very different in their approach, especially as it relates to the three ideas we took from Popper. Table 1 outlines these differences.

Looking first at the idea of falsification, we see Bob as a practitioner of pseudoscience, building models that can accommodate any desired belief. In contrast, Jo advances a clear and falsifiable theory.

Second, considering science as a process of conjecture and refutation, we find Bob barely conscious of his client's (or even his own) theories about marketing, let alone in a position to refute them. Jo, on the other hand, considers it her job to rigorously test the client's theory and then repeat the process on any theories of her own, paying special attention to any weaknesses in statistical method that might allow a bias towards her own conclusions to creep in.

Finally, when it comes to the idea that science must concern itself with current beliefs, we note that Bob is uninterested in the real problems faced by his client, wholly caught up as he is in modelling abstractions and technical concerns. Jo on the other hand, bases her model on the current picture of how the business, the market and human psychology all work. She understands that her approach must be piecemeal, challenging one belief at a time, while acknowledging the others as assumptions. There is no complete bottom-up solution.

Philosophy's investigation into the methods of science did not end with Popper (see "Beyond Popper"), but he set the terms for all future debate. He did this by showing that an understanding of the difference between good and bad science is reached by grasping how it really works in practice.

Popper's demands on scientific practice are difficult – perhaps impossible – to live up to, but at a time when the enthusiasm within business for the statistical sciences is soaring, when expectations are overblown and the resulting pressure sidelines self-criticism, it is more than helpful to have a set of principles that remind us what real science should look like, whether in the laboratory or the boardroom. ■

References

1. Popper, K. (2002) *Conjectures and Refutations*. London: Routledge.

2. Chalmers, A. F. (1999) *What Is This Thing Called Science?* Buckingham: Open University Press.

3. Midgley, M. (2006) *Science and Poetry*. London: Routledge.