

Testing, testing

The social sciences have often been charged with ‘physics envy’ — accused of having an unwholesome desire to appear more scientific by copying the methods and style of physics. Theoretical economics is a prime and justifiable target with its peculiar fascination with mathematical proof and an axiomatic approach to human behaviour. Envyng the authority of physics, the charge goes, economists have used mathematics more as fashion than as a real tool of understanding.

With economics, I tend to think ‘guilty as charged’. Just look at the general equilibrium models currently used by the European Central Bank, which predict economic outcomes by having a continuum of households and firms optimize their ‘intertemporal utility’ by solving problems in the calculus of variations.

But the attack on ‘physics envy’ sometimes reaches further, and questions even the value in social science of testing ideas and hypotheses with empirical data. For example, in a recent essay in the *New York Times* (<http://go.nature.com/aZlBqh>), and in a forthcoming book, political scientists Kevin Clarke and David Primo argue that contemporary social science suffers from an over-commitment to evidence, styled on practice in the physical sciences. “Theoretical models can be of great value even if they are never supported by empirical testing,” they suggest, and the demand for such tests comes with costs: “the belief that every theory must have its empirical support constrains the kinds of social science projects that are undertaken.”

No doubt my training as a physicist biases my view, but I’m sceptical. For one thing, physicists often develop and use theories that have not yet been tested, so if social science suffers from an obsession with empirical testing, this can’t really be the result of ‘physics envy’. Every branch of physics has currently untested theories, and this is natural. Science cannot exist without creative, speculative work; it’s just that these theories are considered to be ‘untested theories’ and interpreted that way — as possibilities that may or may not turn out to have value.

It’s also true that ideas need a little room to breathe and develop before being put to the test. When introducing the general theory of relativity, Einstein didn’t report empirical evidence — other than logical consistency with other ideas of physics — and empirical tests didn’t arrive for decades.

But do ideas in social science really get brutally crushed under immediate empirical tests? That picture certainly doesn’t



If social science suffers from an obsession with empirical testing, this can’t really be the result of ‘physics envy’.

fit economics, where ideas often linger even when empirical tests show them to be unfounded. For example, the rational-choice view of human behaviour remains the workhorse of economics, even though, as one economist has noted, its empirical disconfirmation “appears to be one of the few really robust results achieved by the human sciences.”

Another idea to which Clarke and Primo object is that a theory should make predictions. In the 1950s, they note, an economist named Anthony Downs proposed an explanation for why opposing political parties during elections often try to adopt very similar platforms, despite their apparently different views. If one party stakes out a central position likely to appeal to a wide swathe of voters, Downs reasoned, there’s not a lot the other can do. If one position is particularly advantageous, then there’s a natural tendency for competitors to mimic each other.

“This framework,” Clarke and Primo write, “has proven useful to generations of political scientists even though Mr Downs did not empirically test it and despite the fact that its main prediction, that candidates will take identical positions in elections, is clearly false. The model offered insight into why candidates move toward the center in competitive elections, and it proved easily adaptable to studying other aspects of candidate strategies. But Mr Downs would have had a hard time publishing this model today.”

This may be true, but if so, it also seems understandable. Downs’ idea is intuitively plausible. It may be useful for telling stories that seem to make sense about lots of elections. But is it really true? It shouldn’t be impossible to test the idea, at least roughly, given the masses of polling data from elections around the globe, and recorded statements of politicians. Indeed, if the claim is no more than that parties ‘might adopt identical platforms’, this seems pretty easy to support — just find a few examples. A stronger claim — that this is typically the case, for example — would be both more

interesting and require more evidence to establish.

Finally, can a theory be useful even if it makes no testable predictions? Clarke and Primo say yes. “Theories,” as they put it, “are like maps: the test of a map lies not in arbitrarily checking random points but in whether people find it useful to get somewhere.”

I think most scientists would agree with that analogy. But the analogy actually suggests the opposite of what Clark and Primo seem to think. After all, the usefulness of a map ultimately comes from its actual correspondence with an underlying territory; its ability to capture something true about the world. If we believe a map is useful, we generally have reasons for believing so — it’s worked in lots of situations, got us home when lost, led us straight to a pub when we wanted one.

In other words, true practical usefulness suggests that a theory must make some testable predictions and pretty good ones. The logic seems hard to avoid. If there is absolutely no evidence to judge the legitimacy of an idea — and it’s not a mathematical theorem, for which evidence is irrelevant — then we’re outside of science. That’s okay, of course, not everything is science. If there is evidence — even if it is scant, messy or hard to come by — then theories can be tested, and useful theories do make predictions, otherwise they wouldn’t be useful.

A good theory, as Richard Feynman once said, “sticks its neck out” by making claims that can, at least in principle, be found to be wrong. This isn’t an arbitrary, fussy demand, but part of the engine of science and learning. We learn from mistakes, and we learn best from making explicit mistakes in public so others learn too. The worst thing of all is ideas that cannot even in principle be tested.

Ultimately, Clarke and Primo’s argument seems to be that social science is so hard that progress is limited, if measured by the prevailing scientific standards (ill defined as they of course are). Hence, we should lower the bar on what counts as knowledge, so that even baseless speculation can count. Then we will make wonderful progress — and can go on saying we’re doing science at the same time.

As far as I can see, however, there’s plenty of good social science being done, and the best social science benefits from empirical testing, rather than being harmed by it. □

MARK BUCHANAN