Well-Ordered Science: Evidence for Use

Nancy Cartwright†‡

This article agrees with Philip Kitcher that we should aim for a well-ordered science, one that answers the right questions in the right ways. Crucial to this is to address questions of use: Which scientific account is right for which system in which circumstances? This is a difficult question: evidence that may support a scientific claim in one context may not support it in another. Drawing on examples in physics and other sciences, this article argues that work on the warrant of theories in philosophy of science needs to change. Emphasis should move from the warrant of theories in the abstract to questions of evidence for use.

1. The Introduction. The issues raised here are at the core of a joint London School of Economics and Political Science–Columbia research project, and I should like to urge anyone who is sympathetic with our concerns to join us in that project.[1]

2. Well-Ordered Science. Nick Maxwell has long urged that for science we need not just knowledge but *wisdom*.[2] Recently Philip Kitcher (2001) has been expressing similar concerns in arguing that the most important demand we should make of science is not that it be accurate, progressive, or problem solving (or whatever are your favorites from the traditional

†To contact the author, please write to: Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England; e-mail: n.1.cartwright@lse.ac.uk; or Philosophy Department, 0119, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119.

‡I am very grateful to the Latsis foundation, the British Academy, the National Science Foundation, and a London School of Economics and Political Science–Columbia project grant for support on this research and to Philip Kitcher, Julian Reiss, and Damien Fennell for help with ideas and presentation. (The material is based upon work supported by the National Science Foundation under grant no. 0322579. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the view of the National Science Foundation.)

1. Contact d.j.fennell@lse.ac.uk.

2. See Maxwell 1984.

Philosophy of Science, 73 (December 2006) pp. 981–990. 0031-8248/2006/7305-0044$10.00
Copyright 2006 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

lists of scientific virtues). But rather that it be well-ordered and that it answer the *right* questions in the *right* ways, where value judgements and methodological issues are inextricably intertwined in determining what is *right*. Kitcher focuses on biomedical research. For instance, he objects that we spend too much effort and money trying to develop treatments that will make a small marginal difference to the life span and life satisfaction of first-world people (though, of course, perhaps a large difference to any one individual suffering from a given problem) at the cost of efforts to develop treatments and preventatives for third-world problems.

It is important to notice that Kitcher does not urge that ethics alone—or, more realistically, ethics mixed with a huge dose of self-interest—should dictate what questions get pursued. We also need to mix in from the start considerations of what are the *right methods*. For instance:

- What questions *can* be reasonably pursued at a given time? Kitcher does not, for instance, focus on certain third-world problems just because they affect a huge number of people—and more dramatically than even our awful cancers and heart diseases affect us—but also because he believes that these problems may be improved as a result of research that is neither very costly nor requires great imaginative breakthroughs. Developing variations on known treatments and vaccines so that they will not require refrigeration is one kind of case here.
- What are the effects of pursuing a given question or given line of research? This was the focus of Kitcher's well-known work on the human genome and the effects the results could have on society given what we know about both our power and our political will to guarantee safeguards.
- What methods can get us the kinds of results we are really looking for: exactly what can they deliver and at what cost?

It is these last methodologically oriented issues to which I want to direct your attention because (i) they are truly pressing and thinking about them in science is often confused (or nonexistent). (ii) Like most methodological issues in science, I am convinced, they will benefit from the kind of detailed careful attention that we philosophers are trained to provide. (iii) We are not providing that attention. My aim then is to urge us to direct our efforts away from the more abstract questions that usually entertain us—from highly general questions of warrant (like do we have reason to believe our theories are true rather than merely empirically adequate, is simplicity a symptom of truth, the 'principle principle', and the like) to much more specific questions about particular methods and their problems of implementation, their range of validity, their strengths and weaknesses, and their costs and benefits.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

3. Evidence for Use. My own particular concern in this regard right now is with *evidence for use*. We philosophers tend to buy into the Positivist/Popperian picture of *exact science*, in particular into the view that science can and does establish *stable unambiguous* results, what I think of as "off-the-shelf" results, results that are warranted and once warranted can be put on the shelf to make them generally accessible, from whence they can then be taken down and put to various uses in various different circumstances. For large chunks of the sciences I know about this is a very mistaken picture of warrant; it is a picture we have, I believe, because as philosophers we pay a lot of attention to how scientific claims get tested but pay very little to how they get used. I argue that there is a sense in which our scientific claims are not unambiguous: what a claim means in the context in which it is first justified may be very different from what it means in the different contexts in which it will be put to use. If I am right about this, it follows that:

> What justifies a claim depends on what we are going to do with that claim, and evidence for one use may provide no support for others.

4. Physics Is Not Immune. My own recent concerns about this problem are in the human sciences—economics, other social sciences, and medicine. But they originated in my work on quantum mechanics, and I want to summarize what I noticed there lest we think that the problems are peculiar to the sloppy and unregimented studies of society.[3]

I was looking at cases where quantum theory was uncontroversially central to use, in particular at the role the theory plays in the treatment of lasers, squids, and other superconducting devices. My experience was that the quantum mechanics of the laser engineers was a different animal altogether from the quantum mechanics of quantum theory. Central ideas and language were shared, such as modeling techniques and equation forms. But in engineering lasers this was so intermixed with specifics that depend on materials or that use their own peculiar approximations or import assumptions from other theories that even equations that look very much the same in the two cases were really instead more of a pun. Indeed, we do not need to go all the way to engineering to see this. The work of Sang Wook Yi (2001) shows that it is already the case in condensed matter physics, which on standard philosophical accounts should just fall under quantum theory.

Let me remind you as well of the work of Peter Galison, who shows for specific cases in contemporary physics that experimenters and theo-

3. See Cartwright 1983, 1999.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

reticians have very different understandings of what looks on the face of it to be the same claim.[4] Each implicate the claim in a radically different network of inference and assumption, so different that the claim must be assigned a different sense for the two groups (which, moreover, are obviously not homogeneous within themselves).

If we combine my observations with Galison's we have a real problem for warrant in the use of physics results. First, it is difficult to see how experiment can warrant a theoretical claim, since the theoretical claim both supports and presupposes a very different set of inferences than does the experimental claim. (There is a vast amount of mathematics in the theory that gets no experimental warrant at all.) Then it is equally hard to see how the theory can warrant the use. How then can warrant travel from experiment to use? Or does it? And if not, then what? What philosophical account can we offer of the evidence we need for the assumptions of quantum mechanics *as used* and as understood in those uses?

5. Some Examples. Moving away from physics, let me cite some other examples.

- First from philosopher and sociologist of science Jerry Ravetz, who specializes in questions of use: we may have excellent evidence, from randomized controlled trials even, that a particular fertilizer is both safe and effective. Then we send the fertilizer in bags with English-language instructions to a distant country with dramatically different geology—say very steep slopes with vast runoff—and no culture of fertilizer use. There it is applied just before the huge rains come at 10 or 12 times the tested doses. The river is poisoned, people grow sick, animals die, and no good is done for the crops.

This raises a typical problem. Natural science results—like fertilizer effectiveness and safety—are warranted by natural science methods. But the implementation of those 'same' results is seldom a pure natural science process. It involves social processes as well, and those need to be understood up front. The tests cannot provide warrant for an off-the-shelf result. The result that is warranted by the test is not the one we need to know about for use. That result—the one that we need to know about the safety and effectiveness of the fertilizer in situ—will be highly context dependent, and even knowing what result it is we need to know will require a great deal of social science input.

The problem is that we do not know how to do this. For one, we do not know how to include evidence about social processes into decisions

4. See Galison 1987, 1997.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

that depend heavily on natural science. Consider one anecdotal example to make the point.

- The late John Maynard Smith was a brilliant biologist, himself cautious about the great boon to our health that is often promised on behalf of the human genome project. Asked about designer babies, to the extent that they will be possible, and what policies and safeguards should be put in place, Maynard Smith replied: let the mother decide. She is the person who has naturally evolved to have the most concern for the welfare of the baby.

Maynard Smith's answer was based on his understanding of natural science. It did not occur to him that for sensible policy we need some understanding of the social and political processes: What pressures will mothers be under (e.g., if we let the mother decide, will that be tantamount to letting the father decide)? What do mothers know? And so on.

Worse, Maynard Smith was dismissive about the study of society. In response to a different question after the very same talk, he urged, "The very worst thing would be to let the social scientists get involved." He was cheered for this by a number of biologists in the audience—and this despite the fact that the talk was hosted by the London School of Economics and Political Science.

Unfortunately, I am afraid that our attitude here in the Philosophy of Science Association is much like Maynard Smith's and his biological audience. Science faces pressing epistemological questions. Not the ones we usually ask, "What warrants a theory," but rather, "What warrants the conclusions we draw on the basis of that science in putting it to use?" This is an incredibly hard question (probably a great number of questions bundled into one), and it is one which we do not have a strong starting position to build from. That is clearly part of the reason that so few of us work on the problem—I know I find it very daunting.

But there are other reasons, and one, I believe, is Maynard Smith's: when it comes to results that require the input of both natural and social science, we look the other way. The social sciences are the poor sister to philosophy of natural science—philosophy of physics, philosophy of biology, and the logic of statistical inference. When we do turn our attention to the social sciences it is economics that gets center place, and even there it is not labor economics, the design of measures of poverty, or the kinds of questions Joseph Stiglitz (2002) raises in criticizing the International Monetary Fund about the separation of economic science and self-interest or the fit of universal economic models to highly various local situations. Rather it is the upper reaches of game theory and decision theory that take up the bulk of our attention.

Besides the problems of integrating—or even obtaining—social science

https://doi.org/10.1086/518803 Published online by Cambridge University Press

evidence from the start, the Ravetz story about the fertilizer should also remind us that we have very little to say about combining evidence at all.[5] Let me illustrate with an example I am now studying.

- British epidemiologist Michael Marmot (2004) urges that low status is bad for your health and that this is true not just at the bottom end but holds all the way up the social gradient. For instance, if you board the tube in central London and go six stops east, you lose one year's life expectancy with each stop. I want to focus on two interconnected issues: (1) How far do/should Marmot's conclusions stretch? For what populations and under what circumstances can we expect his conclusions to obtain? (2) What evidence is relevant to support these conclusions?

Marmot himself suggests that the conclusions hold across all situations where low socioeconomic status leads to increased social isolation and to a particular kind of stress (stress resulting from a combination of low control and high demand). It is interesting how he supports this. In his own work he has carried out detailed longitudinal studies across 20 years and more on Whitehall Civil servants, with startling results; for instance, the highest paid Whitehall civil servant has twice the chance of living to age 60 as the lowest paid.[6] But Marmot also has results from interviews and questionnaires on job control and job demand; about the association between laboratory-induced stress and various physiological reactions that are thought to increase the chance of stress-related illness; and on Whitehall status and lifestyle factors connected with illness, such as smoking, obesity, exposure to pollution and exercise, and more.

I think we can say (if only we knew how to amalgamate this evidence!) that Marmot's results have a high degree of *internal validity*: they are very well designed and well controlled to establish just the results he claims. To achieve this high standard of internal validity it helps to have a set of cooperating captive subjects with known characteristics like Whitehall civil servants. But what about *external validity*: for what other populations can we expect these same conclusions to hold? Or, to the point for us: what can we offer on external validity? Little, I think, beyond the truisms that there is generally a trade-off between internal and external validity and that the chance of external validity is enhanced if the subjects are representative of the target population.

We are pretty good at many questions of internal validity: we argue—and rightly so I think—about the real benefits of randomization in clinical

5. See Ravetz 1999, 2000, 2002.

6. For more on the Whitehall studies, see http://www.workhealth.org/projects/pwhitew.html and http://www.workhealth.org/projects/pwhitepub.html.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

trials, about an approach to statistical inference based on Fisher's ideas versus those of Nehman/Pearson, about the causal Markov condition, and about whether Holland and Rubin are right to justify standard randomized-control-trial techniques on the basis of singular counterfactuals.[7] But we have little to say about external validity—and that is what matters for use.

The related question is about combining evidence. How does Marmot himself support the move from Whitehall civil servants to a far broader population? He does so by marshaling a great deal of evidence of different kinds: for instance, experiments on monkeys that put together the top monkey from a number of different troupes. The monkeys again form a hierarchy, and the ones at the top are by far the healthiest. And he does so by looking at health data across Canadian provinces and at what happened to health in Russia—especially among Russian men—after the change from socialism. And so forth.

Altogether, informally, it is an impressive package. Where did he publish it? That helps to make my point—in one of those high-caliber 'semipopular' books. For this is not the kind of thing that goes into a serious journal, and in a sense rightly so. Even review articles in journals tend to cite studies that have a great deal of commonality of language and method—that way they can be adequately policed by the experts in the field. That is just the problem. We have no experts on combining disparate kinds of evidence (apart from some neat metastatistical techniques, which do not stretch very far). But doing so is at the heart of scientific epistemology when that epistemology is directed at establishing results we can use. So we here in this association should be tackling it.

- We spend a lot of energy and imagination on questions of when we are entitled to count a scientific conclusion as true. But we spend little effort in thinking about what truth buys us. Think about causal modeling in political economy. As John Stuart Mill (1836) stressed, the causes operating in the economy change frequently and usually unpredictably. So, as econometricians Michael Clements and David Hendry (2002) argue in recent work on forecasting, even a very accurate causal model cannot be relied on to forecast correctly. The best evidence for the truth of the model is not good evidence for its forecasts.

This is the same kind of conclusion that social-psychologist Gerd Gigerenzer urges when he talks about "cheap heuristics that make us rich." Gigerenzer illustrates with the heuristic by which we catch a ball in the

7. See Holland 1986; Rubin 1974.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

air.[8] We run after it, always keeping the angle between our line of sight and the ball constant. We thus achieve pretty much the same result as if we had done the impossible—rapidly collected an indefinite amount of data on everything affecting the ball's flight and calculated its trajectory from Newton's laws.

The point about cheap heuristics is that they are not anything like the 'true' account. They are not approximations to it nor idealizations from it; they do not, as many antirealists (e.g., constructive empiricists, NOA-ers, etc.) demand of 'good' theory, have all the virtues of truth just failing truth (or good grounds for it); they do not improve by adding more realistic assumptions (to the contrary, this usually undermines the 'trick' by which they work in the first place); and so forth. This puts them entirely outside our usual debate. But cheap heuristics are crucial for practice. What evidence is necessary to justify the use of a conclusion derived from a cheap heuristic? Must we first have the 'true' model and then show that the results converge often enough? Or, ... Again, these are key questions in scientific epistemology as soon as we stop focusing on theory and turn to use. We should be working on them.

- There is one area of use in which we philosophers of science are doing good detailed work at the moment—methods of causal inference. But I would like to close by suggesting ways in which we should be stretching this work. We have on offer right now a lot of alternative accounts of what causality consists in: probabilistic theories of causality, invariance accounts, manipulation theories, causal process theories, and so on. Each, it turns out, is closely associated with one or another well-known method for establishing causal conclusions: tests for Granger causality, stability tests, controlled experiments, identifying causal mechanisms, and so forth.

We put a lot of energy into trying to figure out which of these accounts of causality is correct. I would like to see us divert some of that energy to a more refined question:

> Which account—with its concomitant method—is right for which kind of system in which kinds of circumstances?

When we can answer that we will know about the proper use of the different associated methods.

The currently fashionable Bayes-nets methods are doing better than most in this regard. For they lay down three assumptions about causality, then show that any time causes meet these three conditions, their methods

8. See Raab and Gigerenzer 2005.

https://doi.org/10.1086/518803 Published online by Cambridge University Press

will not give erroneous results (though they may often yield no results at all) if the input information on the probabilities is correct.

This is a good start. But it does not go far enough. What are these three assumptions? They are so-called faithfulness, the causal Markov condition, and minimality. And what does all that mean? I can write them out for you (many of you know them already), and you will understand them—in a sense. But what I write will not help a practicing scientist. What do these conditions amount to in the real world? Are there any even rough identifying features a system may have that will give us a clue that it is faithful or satisfies causal Markov or minimality?

Bayes nets experts are very good at proving theorems. They are also, many of them, getting good at what turns out to be the terribly complicated and subtle matter of applying the methods in real cases. But little is done on criteria in more concrete terms of *when* to apply these methods. And our other accounts of causality lag far behind Bayes-nets in this regard. They shouldn't.

6. Conclusion. We ought to aim for a well-ordered science. That involves a number of different issues to which philosophy of science can—and should—contribute. The ones I have focused on involve questions of warrant and evidence. Most of our work on warrant in the Philosophy of Science Association is still fixated on theory. If we want to contribute to a well-ordered science that answers the right questions in the right way, we need to shift our emphasis and work instead to questions of *evidence for use*.

REFERENCES

Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Oxford: Oxford University Press.
——— (1999), *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
Clements, Michael, and David Hendry (2002), "An Overview of Economic Forecasting," in Michael Clements and David Hendry (eds.), *A Companion to Economic Forecasting*. Oxford: Blackwell, 1–18.
Galison, Peter (1987), *How Experiments End*. Chicago: University of Chicago Press.
——— (1997), *Image and Logic*. Chicago: University of Chicago Press.
Holland, Paul (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association* 81: 945–970.
Kitcher, Philip (2001), *Science, Truth and Democracy*. Oxford: Oxford University Press.
Marmot, Michael (2004), *Status Syndrome: How Your Social Standing Directly Affects Your Health and Life Expectancy*. London: Bloomsbury.
Maxwell, Nicholas (1984), *From Knowledge to Wisdom: A Revolution in the Aims and Methods of Science*. Oxford: Blackwell.
Mill, John Stuart (1836), "On the Definition of Political Economy and the Method of Philosophical Investigation Proper to It," in *Collected Works of John Stuart Mill*, vol. 4, J. M. Robson (ed.), *Essays on Economics and Society*. Toronto: Toronto University Press.
Raab, Markus, and Gerd Gigerenzer (2005), "Intelligence as Smart Heuristics," in Robert

https://doi.org/10.1086/518803 Published online by Cambridge University Press

Sternberg and Jean Pretz (eds.), *Cognition and Intelligence: Identifying the Mechanisms of the Mind*. Cambridge: Cambridge University Press.
Ravetz, Jerome (1999), "What Is Post-normal Science?" *Futures* 31: 647–653.
——— (2000), "The Future Politics of Science," *Futures* 32: 505–507.
——— (2002), "Food Safety, Quality and Ethics—a Post-normal Perspective," *Journal of Agricultural and Environmental Ethics* 15: 255–265.
Rubin, Donald (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology* 66: 688–701.
Stiglitz, Joseph (2002), *Globalization and Its Discontents*. New York: Norton.
Wook Yi, Sang (2001), *How to Model Macroscopic Worlds: Towards the Philosophy of Condensed Matter Physics*. PhD Dissertation. London: University of London.

https://doi.org/10.1086/518803 Published online by Cambridge University Press