**Follow the Science?**

**David Papineau**

**Rivkah Hatchwell**

**Institute of Philosophy**

**February 28 2023**

An influential body of contemporary philosophers of science holds that evaluative considerations ought to shape the way that scientists report their findings. I disagree, not least because this is likely to bring science into disrepute.

**Plan**

1. Initial Meta-Ethical Observations

2. Evidence not Convenience

3. Rudner and “Inductive Risk”

4. Whose Values?

5. Credences versus Significance Tests

6. Higher-Order Judgements

7. The Problem of the Priors

8. Pragmatic Encroachment in Epistemology

9. Racing to the Bottom

*1. Initial Meta-Ethical Observations*

We are often told that science should avoid value judgements. But why exactly? All too often the rationale seems to rest on nothing but the positivist idea that evaluative claims are an optional boo-hurrah add-on to the descriptive facts. Science is supposed to describe the facts. It has no authority to tell us how to respond emotionally.

By the standards of philosophical metaethics, however, this line is half a century out of date. Is it really optional how to react to the news, say, that children are dying as a result of the Turkish earthquake? Non-naturalist realism, naturalist realism, quasi-realism, projectivism, . . . all hold in their different ways that rational humans have no option but to disvalue this, and accordingly see nothing wrong with scientific findings having built-in evaluative implications.

*2. Evidence not Convenience*

This is not to say it’s acceptable to adopt some factual claim *because* of your values. Still, why would anyone want to do that? Whatever values are in play, surely it is in everybody’s interests to believe truly if they can, and therefore to tailor their beliefs to their evidence not their values.

Of course, if I own a coal mine/am rich/grow crops, then I might want to *persuade* *others* who care about global warming/public revenue/bees that the evidence doesn’t show that stopping mining/taxing rich/banning neonicotinoids will further their aims. But it would be silly for me to believe this myself, if the evidence doesn’t support it. (After all, if I care about global warming/public revenue/bees at all, I don’t want to delude myself about the impact of my activities.)

As a matter of fact, of course, people do often delude themselves in this way. It is an interesting question what mechanisms produces this result. Here we are in the realm of cognitive dissonance and self-deception.

Still, I take it to be uncontroversial that it’s bad to induce others to form unevidenced beliefs because it will further your sectional interests to do so, and it’s not made any better if you end up duping yourself into those beliefs too.

Surprisingly, a majority strain in contemporary philosophy of science comes very close to arguing that it is the duty of scientists to do just that.

*3. Rudner and “Inductive Risk”*

The line of thought goes back to the 1950s, when Richard Rudner pointed out that the evidence will never conclusively prove or falsify any scientific claim P. So there is always some “inductive risk” involved in definite commitment to P (or not-P). Rudner argues that in deciding whether to accept this risk, scientists can and should consider the costs of error.

Suppose P is *neonicotinoids reduce bee populations by more that 10%* and that neonicotinoids will be banned if and only if this is accepted. Then Rudner’s thought was that our evidence thresholds should be set in a way that reflects the costs of a false negative/positive. Since it will be so much worse to destroy the bees than reduce the crop yield, we ought to be relatively ready to endorse P, not demanding a high standard of evidence, and correspondingly be slow to reject it. The entomologists should announce P as established on far weaker evidence than they’d require to reject it.

*4. Whose Values?*

The obvious problem here is: “Whose evaluations of the costs and benefits should go into the calculation?”

In effect, the scientists are using their values to decide that the evidence warrants banning neonicotinoids. But the farmers, or indeed the general public, might well draw a different conclusion about the right policy even given the same evidence. It looks as if the scientists are short-circuiting something that should be subject to democratic debate.

Heather Douglas has argued that the scientists shouldn’t use their own values in deciding what to commit to, but rather those of the policy makers they are advising. That strikes me as better, but not a lot better. Not everybody shares the values of the policy makers in power, and they might reasonably take exception to experts who are ready to assert P on weak evidence, just because that’s the view that backs the choices their political masters favour.

But what is the alternative? How else are we going decide when to bridge the “inductive risk gap” except by considering the stakes by some standards?

This debate has had a new lease of life over the past couple of decades. There are views on all sides, but a surprising number of philosophers of science side with Douglas in holding we should look to values to guide us about which scientific claims to endorse, including Philip Kitcher, Quill Kukla, Katie Steele, Eric Winsberg, and Torsten Wilholt.

Much of the debate involves relatively practical issues. In what follows I shall mostly by-pass these, focusing on more foundational matters instead, though I shall return to some practical considerations at the end.

*5. Credences versus Significance Tests*

Note how the whole debate is premised on the assumption that science’s job is to arrive at categorical judgements to the effect that P (or not-P).

Rudner’s original paper in 1953 soon prompted a response from the Bayesian Richard Jeffrey, who pointed out that values wouldn’t come into it if scientists simply reported the credence/degree of belief/subjective probability for P warranted by the evidence. If you don’t try to bridge the “inductive risk gap”, you won’t need to evaluate the costs of doing so wrongly.

Of course, this ran counter to the official methodology of science, which sought to deal with inconclusive evidence via the apparatus of *significance tests* with rejection regions and p-values and so on. This was designed as a methodology that would allow you to definitely assert P (reject the null) if you got a significant result, and reject P (accept the null) if you didn’t get a significant result (or perhaps more cautiously in that case, depending on the power of the test, suspend judgement).

Indeed Rudner and a number of others in the ongoing debate explicitly refer to significance levels as what they mean when they say thresholds for acceptance should be sensitive to stakes.

Still, the consonance with significance testing lends no support to Rudner et al’s way of setting things up. I hope we can all agree that the logic of significance testing is deeply flawed and that the recently recognised replication crisis is nothing but the chickens of significance testing coming home to roost.

(If your P was unlikely to start with, then data that are significant at the 5%/1%/whatever level are no kind of reason to outright accept it—after all its probability might still be very low. No wonder many “significant results” aren’t repeatable.)

The old orthodoxy didn’t like the Bayesian alternative to significance testing because of the apparent arbitrariness of the priors. (Do we want the opinions of old male doctors to influence our medical practice?)

I’ll come back to this arbitrariness. But the basic point is that no sensible methodology, even if it aims at categorical judgements, can do without prior probabilities. That’s the lesson of the replication crisis.

Indeed, come to think about it, the Rudnerian idea that we should set thresholds by considering the costs of false positives/negatives can’t do without them either. Rudner and others skirt round this by skipping the sums and just alluding to the setting of significance levels, but just consider this table:

|  |  |  |
| --- | --- | --- |
|  | P | Not-P |
| Accept P (ban spraying) | 0 | 0 |
| Reject P (allow spraying) | -100 | 20 |

You can’t use this to tell you whether the evidence justifies Accepting P unless you have some idea *how likely the evidence makes P*. The idea of some non-credence-based calculation that will rationally set a threshold for acceptance given the stakes is bogus, along with the supposed logic of significance tests.

*6. Higher-Order Judgements*

A standard objection to Jeffrey’s alternative is that it’s going to require a higher-order categorical judgement that *P is probable to the claimed degree*, and the threshold for that judgement should again be sensitive to the stakes.

This strikes me as muddling up credences and objective probabilities.

Objective probabilities can be part of the subject matter of scientific claims P. For example:

*Neonicotinoid spraying has a 90% probability of reducing local bee populations by more that 10%*

Credences express attitudes to scientific claims P. For example we might attach 80% credence to some P, perhaps about the effects of neonicotinoids.

Sometimes credences and objective probabilities are related. When the proposition at issue is an outcome of a mechanism producing results with an objective probability, like a coin toss, then the “Principal Principle” tells us, roughly, to set our credence to the objective probability of the outcome. But that’s not the kind of case at issue here. Nobody thinks that scientific claims P are produced with objective probabilities by mechanisms. The credences at issue in this debate might be in claims *about* objective probabilities, but they aren’t *reflections* of objective probabilities. They are more like expressions of informed opinion.

So in particular, the credence c that Jeffrey says the scientist should express in P won’t answer to some categorical judgement that the objective probability of P is c. Rather it will simply be the expression of the scientist’s expectation that P is true, grounded in the scientist’s initial judgement of P’s credibility and modulated by subsequent evidence. So there’s no need for any higher-order judgement *that P is c% probable*.

*7. The Problem of the Priors*

Of course, this response to the higher-order charge highlights the way in which the Bayesian alternative does rely on prior probabilities. And that itself might seem to reduce the attractions of Jeffrey’s Bayesian line. Do we want scientists to be saying “Given my evidentially ungrounded initial assessment of P’s credibility, plus the evidence that has now come in, I’d say that the odds in favour of P are 75%”? Apart from anything else, other scientists with different prior probabilities will have different odds.

Still, what’s the alternative? Is it really better in such a case for scientists to speak with one categorical voice, just because that supports the action they or their political masters would choose, even when the evidence is in truth equivocal?

How bad is the indecision that comes with the dependence of Bayesian judgements on prior probabilities? One initial point is that the priors of scientists won’t just be random guesses, but informed by their background knowledge and experience, even if not by formal empirical studies. (Don’t take the arbitrariness of traditional medical practice as the model for scientific “judgement”.) And a second point is the way that, once we do have empirical evidence, that will tend to wash out differences in priors, and push credences in the same direction. In truth, there’s no reason why the Bayesian judgements of informed scientists shouldn’t display reassuring conformity even in cases where the evidence doesn’t warrant any categorical conclusion.

(Note also that there will be plenty of cases where unanimous categorical judgements are just fine. We’ve moved beyond Rudner’s 1950s idea that evidence will never conclusively prove or falsify any scientific claim P. The evidence mightn’t ever *logically prove* any P, but there’s no need for any real doubt that COVID-19 is caused by a virus, that smoking causes cancer, that the sun is powered by hydrogen fusion, . . .)

*8. Pragmatic Encroachment in Epistemology*

These issues also come up in mainstream analytic epistemology. A number of writers, most prominently Jason Stanley, hold that knowledge is stakes-sensitive or interest-relative. (You might ordinarily know that the bank is open on Saturday, but you won’t in a context where you’ll lose $100,000 unless you can cash a cheque then.)

Others object to this kind of “*pragmatic encroachment*”. They say that the formation of beliefs and other doxastic attitudes should be influenced by evidence alone, not pragmatic considerations.

My own view is that we need to start the debate earlier, and ask why we are dealing in categorical states of mind at all. Why not just form a rational credence in the bank being open, and decide what to do in the normal utility-maximizing way? All the problem arise only because we’re fussing about whether to form an unneeded *flat-out belief*.

There is amazingly little discussion of the virtue of flat-out beliefs in the literature. One idea is that it is good to *know*, and knowledge entitles you to flat-out belief. (But why is it good to know?) Another is that it’s too complicated to be thinking in credences all the time. (But shouldn’t we engage with the complications, when $100,000 or the fate of the bees are at stake?) For the most part, however, it’s simply assumed that there *must* be some rationale for trading in flat-out beliefs, because after all humans do it all the time. I have my doubts.

(The best rationale comes from my colleague David Owens, who observes that, given the way humans are constituted, we can’t engage emotionally with the world without forming flat-out beliefs. You can’t *blame* somebody for their bad behaviour, or be *angry* with them, or *pleased* at their success, or *happy* because Spurs won . . . if you only think it’s *likely* these things happened. Maybe so. But this is no argument for using flat-out beliefs in making serious decision.)

My overall view about the issue of pragmatic encroachment is that it is no doubt reasonable, given that there are no doubt pragmatic reasons for forming flat-out beliefs. But I don’t feel I can really say more until I’m told what these pragmatic reasons for flat-out beliefs are.

*9. Racing to the Bottom*

In much of the inductive risk literature, the dangers of *manipulation* and wilful *misunderstanding* are standardly mentioned as pragmatic reasons for scientists forming, or at least expressing, categorical rather than probabilistic views.

Vested interests and anti-science factions will latch onto any expressions of uncertainty or doubt to undermine the scientific case for sound policies, arguing that the evidence is equivocal, that there is no scientific proof of global warming, vaccine efficacy, bee deaths, . .

The solution, according to Kevin Elliott, is for scientists to adopt a “no passing the buck principle”. It’s no good their passing on their credences to the politicians and policy-makers, leaving it to them to make rational decisions. The uncertainty will only be exploited by bad actors. Instead the scientists should grasp the nettle themselves, identify the right course of action, and promulgate as definitively established those claims that support it.

Maybe this is the pass we have come to. But if so I think we are in trouble.

We’ll do remember that the bad actors in this story are not necessarily evil. In the first instance, all that distinguishes them from the consensus of right-thinking scientists is that they have different priorities. They might understandably place more weight their household incomes than the future of the planet. Or perhaps they prioritise generation of profit over environmental impact. In themselves these divergences are not pathological. They are the kinds of differences in values and interests that we can expect among reasonable people, and I see no compelling reason why democratic processes should not be able to manage and resolve them.

It's not these diverging priorities that are being invoked to justify categorical scientific pronouncements, but the consequent sin of manipulating the facts, of seizing on uncertainty to make it seem as if there is no evidence that unwanted measures are needed. And I agree that this is a sin that should be condemned. Democratic processes will not be able to find the right way of resolving conflicting priorities if it is misinformed about the relevant facts.

So let’s get our manipulation in first!

As far as I can see, the cure is just as bad as the problem. The “no passing the buck principle” is simply urging scientists to manipulate the facts in the opposite direction, and present partially evidenced claims as conclusively proven, when that supports the policies they favour.

And the real danger is that this policy is going to set us off on a race to the bottom. Perhaps it will have commendable upshots in particular cases, ensuring that we arrive at the policy that would be democratically favoured in the absence of propaganda from powerful bad actors.

But the trouble is that, once the practice becomes standard, it will undermine trust in science. Those who are unsure of the scientists’ preferred policies will quite rightly come to suspect that they are guilty of overegging the evidence. And this will only encourage the attitude that “we’ve had enough of experts”.

What we want is that democratic decisions should be informed by accurate scientific information, including accuracy about partial credences. These decisions can take into account differences in values and interests and use the science to see how best to resolve these. But our institutions won’t be able to do this if all sides are wielding propaganda to support their preferred policies. In the end, allowing evaluations to shape scientific decisions will only devalue the coin of science.

“It would be romantically nostalgic and practically impossible to disentangle science from commerce and government and return it to an Ivory Utopia. . . [Still, an] English proverb has it that he who sups with the devil needs a long spoon: civic institutions are not the devil, but their needs, after all, do not match with dedication to Truth.” Steven Shapin in the *LA Review of Books* “Is There a Crisis of Truth?” December 2019: