

What Is Kitcher's Real Realist Really a Realist about?

Matthias Egg

Summary: I review Philip Kitcher's defence of scientific realism against the so-called pessimistic induction. While supporting his overall strategy, I claim that there is a lacuna in Kitcher's argument, which needs to be filled by spelling out more precisely what parts of scientific theories we should be realists about. I attempt to do this and to show that scientific realism can thereby not only be protected against the pessimistic induction but also against a recent argument by Kyle Stanford, known as the "new induction".

1 Introduction

Some time ago, I attended a talk entitled *Realisms for Sale: entity, experimental, structural (epistemic and ontic), partial, 'full-on' ... But will the real realism stand up please?*¹ The speaker did not refer to *real realism* in the sense of Kitcher (2001), but his title neatly captures what I see as one of Kitcher's central aims: to provide a substantive argument for scientific realism without being drawn too far into the rivalries among realism's different strands. This new strategy seeks to improve on previous realism debates, which "are often perceived as boring and unproductive".² Thus, instead of entering into the technical intricacies that sometimes preoccupy scientific realists (as well as their opponents), Kitcher argues for a strategy that starts from a largely uncontroversial assumption (called the "Natural Epistemological

¹ The talk was delivered by John Worrall at the Workshop *Scientific Realism Revisited* at LSE in April 2009.

² Kitcher (2001: 153)

Attitude” or NEA) and arrives at real realism (RR) via the twin extrapolations, (Ex1) and (Ex2):

- (NEA) If we observe the relations between people and their environments, we assume that people sometimes form representations of things that are independent of them. While some of these representations may be inaccurate, a person’s success in responding to and shaping her environment would be inexplicable unless some of her representations were accurate.
 - (Ex1) The correlation between accurate representation and practical success is not confined to everyday interaction with observable objects but carries over to the unobservable entities of advanced science.
 - (Ex2) The accuracy of representations that people form of things independent of them does not depend on their being observed by us, or by any other observer.
-
- (RR) Some of the claims of advanced science are accurate representations of observer-independent entities in the world.

In the formulation given here, real realism may look so modest as to become uninteresting. In order to give substance to the assertion that *some* scientific claims accurately represent the world, an answer to the inevitable question “which ones?” is needed. The first goal of this paper is to argue (in sections 2 and 3) that Kitcher needs to say more than he does in answer to this question. In a second step (sections 4 and 5), I will suggest such an answer, claiming it to be entirely compatible with Kitcher’s account.

2 Success, truth and the pessimistic induction

In his defence of (Ex1), Kitcher attempts to refute an antirealist argument known as the *pessimistic induction*, which claims that history gives us little reason to suppose that successful scientific theories are even approximately true. A famous example of

a successful but fundamentally mistaken theory which Kitcher discusses at some length is Fresnel's wave theory of light. It made astonishingly successful predictions, but it was based on an assumption which we now consider untenable, namely the existence of an all-pervading ether, in which light waves are propagated. Kitcher now claims that the ether hypothesis, far from discrediting Fresnel's complete theory, can be decoupled from the rest of the theory, because "although Fresnel *thought* he needed to assume a medium in which light waves were propagated, he could actually have delivered his explanations and predictions without that assumption."³ The ether is therefore a prime example of what Kitcher calls an "idle wheel", as opposed to a "working posit". The following quote shows how he uses this distinction in order to counter the pessimistic induction:

"Instead of thinking about the virtues and vices of whole theories, we should distinguish the hypotheses that are genuinely put to work, claiming that the success of a theory provides grounds for thinking that those hypotheses — the hypotheses that characterize 'working posits' — are approximately true." (Kitcher 2001: 170)

By distinguishing between *working posits* and *idle wheels*, Kitcher seeks to retain the success-truth link on the level of *hypotheses* in spite of its failure on the level of *theories*. Several authors have already criticised this endeavour (see Chang (2003), Elsamahi (2005) and Stanford (2006: Ch. 7))⁴, their main argument being that the distinction in question can only be drawn with the benefit of hindsight, if it can be drawn at all.

³ Kitcher's comment on my proposal

⁴ It could be replied that Chang's and Elsamahi's critiques are almost exclusively directed against Psillos's (1999: 108-114) account of the distinction, which differs from Kitcher's. This is correct, but then it has to be noted that Psillos's version is an improvement on earlier efforts by Kitcher (1993: 140f.). Kitcher seems to acknowledge this in his later work (2001), since he there abandons his earlier terminology, replacing "presuppositional posits" by "idle wheels", which comes close to Psillos's "idle constituents".

This does not bother Kitcher. Although he admits that there may be cases in present science where we fail to realise that some of the posits of our theories are in fact dispensable in just the way the ether turned out to be, he nevertheless suggests that we are entitled to make the judgments we do, while being conscious of our own fallibility. The situation is, according to Kitcher, reminiscent of the so-called “preface paradox”: “Our predicament is like that of the author who confesses in her preface that she is individually confident about each main thesis contained in her book but equally sure that there’s a mistake somewhere.”⁵ This strikes me as an understatement of the antirealist’s charge, the issue of which is not just the ordinary human hazard of being wrong from time to time, but a specific epistemic problem that affects science in a way in which it does not affect other human endeavours. To be sure, antirealists differ in their characterisation of that problem: while Bas van Fraassen (1980) takes issue with scientific claims about unobservable entities, Kyle Stanford (2006) claims that we may not yet have considered all the possible alternatives to our present scientific theories. But the basic charge against scientific realism is the same in both cases: science involves an epistemic risk that differs in kind from the epistemic risk involved in other fields of knowledge. Therefore, simply admitting that scientists are as fallible as anyone else will not suffice to silence the antirealist.

3 Stanford’s “new induction”: a threat to real realism?

Mentioning van Fraassen and Stanford shifts the focus of the discussion, since they both base their positions on arguments that go beyond the pessimistic induction. More precisely, in van Fraassen’s “constructive empiricism”, the pessimistic induction plays no significant role. I will not investigate van Fraassen’s arguments

⁵ Kitcher (2001: 171). The preface paradox was first introduced by Makinson (1965).

here, because they are, in my view, successfully addressed by Kitcher's *Galilean strategy*. On the other hand, Stanford's position, being more recent than Kitcher's work on realism, is not directly addressed by it. This section briefly introduces Stanford's argument, in the next section I will sketch a possible counter-strategy. As I said above, Stanford argues that the problem indicated by the pessimistic induction is substantially more serious for scientific claims than it is for other claims. The reason for this is what he calls *the problem of unconceived alternatives*. If we reach the conclusion that a certain hypothesis is true because we have managed to exclude all alternative explanations for a given phenomenon, our warrant for this conclusion depends on whether we have considered all the relevant possibilities in the first place. In everyday situations, we often have good reasons to assume that we have in fact exhausted the space of plausible possibilities. Suppose I want to know who took the last bottle of beer from the fridge. Suppose further that I know *I* did not take it and that my friend Bob does not like beer. I can then quite safely conclude that it was in fact my friend Bill who took the beer, the plausible background assumption being that no one apart from me and my two friends has access to that fridge.

By contrast, such eliminative inference has often proved fallacious in the context of scientific theorising, as is illustrated by the following quote: "Light may be a swarm of projectiles, or it may be a vibratory motion whose waves are propagated in a medium; is it forbidden to be anything else at all?"⁶ Today, of course, we regard light as neither one nor the other. The concept of light we now consider as accurate was just not among the alternatives that were taken into consideration at the beginning of the 20th century. An aspect of this was already highlighted by Kitcher's remark on Fresnel's theory of light: the only way in which Fresnel could make sense of the observed wave-like phenomena was by assuming a medium for the waves to be propagated in. The alternative idea with which we are now so familiar, namely that

⁶ Pierre Duhem, quoted in Stanford (2006: 28).

electromagnetic waves do not need a medium, was not one that Fresnel considered.

Cases like this form the basis of Stanford's *new induction over the history of science*:

“[W]e have, throughout the history of scientific inquiry and in virtually every scientific field, repeatedly occupied an epistemic position in which we could conceive of only one or a few theories that were well confirmed by the available evidence, while subsequent inquiry would routinely (if not invariably) reveal further, radically distinct alternatives as well confirmed by the previously available evidence as those we were inclined to accept on the strength of that evidence.” (Stanford 2006: 19)

It is easy to see how this *new induction* poses a threat to Kitcher's distinction between working posits and idle wheels. Today we recognise that the ether was an idle wheel, because we have a theory that incorporates all the successes of Fresnel's theory without depending on the ether hypothesis. Why think that the entities we now consider as working posits will not some day suffer the same fate?

4 The causal strategy against the new induction (with an example from particle physics)

My suggestion is that in order to overcome this stumbling block, we need to be more explicit about what we mean by a working posit. What kind of work is implied here? Kitcher speaks of hypotheses that are “put to work” in order to achieve a certain success. Successes, in turn, “are matters of prediction and intervention”⁷.

⁷ Kitcher (2001: 167). Surprisingly, Kitcher does not mention *explanation*. Perhaps he sees this as included in a wide sense of “prediction”, as when physicists say: “the Schrödinger equation allows one to predict the hydrogen spectrum”, even though the latter was known long before the former.

Now, if a working posit is involved in achieving a successful *intervention*, "work" undoubtedly carries a causal connotation: the intervention is successful because a posited entity is manipulated in the right way to bring about the desired outcome. This suggests that the entity in question really does exist. (As Hacking (1983: 23) famously put it: "if you can spray them then they are real.")

Things are less clear, however, in the case of *predictive* success. Many predictions are successful without positing any causal mechanism, for example by exploiting some symmetry or other kind of regularity. Thus, Kitcher's working posits include causally efficacious entities as well as some more abstract entities that do not cause anything. In other words, the class of working posits may be subdivided into a *causal* and a *theoretical* kind.

I claim that only the first kind of working posits, the ones to which we can ascribe concrete causal roles, can be saved from Stanford's attack. By contrast, I think that Stanford is right about those working posits that appear only in theoretical hypotheses (successful though these may be).

Let me illustrate this idea with an example from particle physics. In 1930, Wolfgang Pauli postulated the existence of a previously undetected particle which came to be known as the *neutrino*. The main reason for its introduction was an observed inconsistency in the energy balance of nuclear beta-decay. Some energy seemed to disappear in this reaction, and Pauli's idea was that a hitherto unknown particle might carry away some energy without itself being detected. He called his assumption "a desperate way out", and it was generally regarded as an *ad hoc* hypothesis, since there was no independent evidence for such a particle. In the following years, however, the neutrino hypothesis proved to be a powerful theoretical concept. This was mainly due to Enrico Fermi's theory of beta-decay, which rested on the assumption that each such decay involved the creation of a neutrino. The success of Fermi's theory led to widespread acceptance of the neutrino hypothesis, as is illustrated by the following quote from 1948: "Not everyone would be willing to say that he believes in the existence of the

neutrino, but it is safe to say there is hardly one of us who is not served by the neutrino hypothesis as an aid in thinking about the beta-decay process.”⁸

It seems to me that the status of the neutrino around 1950 is similar to the status of the ether at the end of the 19th century: although direct evidence was missing, the neutrino’s existence was generally accepted, because it was taken to be an essential part of Fermi’s theory, just as the ether was once taken to be an essential part of electrodynamics.

But the story of the neutrino did not end there. In the mid-1950’s Frederick Reines and Clyde Cowan performed a series of experiments by which they achieved the so-called “direct detection” of the neutrino. It may, of course, not be obvious to the non-physicist what “direct” means in this context. Neutrinos are neither visible themselves nor do they produce visible tracks in particle detectors. All that can be detected are the products of so-called *inverse beta-decay*, a process that the neutrino can bring about when it hits a proton. These products (a positron and a neutron) initiate further processes, which then produce a characteristic double-peak signal in the particle detector. By 1956, Reines and Cowan concluded that at least some of these events were in fact produced by neutrinos.

On a first glance, this claim seems equally vulnerable to Stanford’s charge as the hypotheses concerning the nature of light discussed above. If one looks at the relevant articles⁹, one sees that Reines and Cowan considered several alternative explanations of their measurement results and performed various tests to exclude them until only the neutrino hypothesis remained. But how do we know, Stanford might ask, that they exhausted the space of plausible alternatives? The *new induction* seems to invalidate the assumption that they managed to do so.

This is too superficial, however. The alternatives among which Reines and Cowan had to decide were not some general theories about the structure of matter, but concrete accounts of what

⁸ Crane (1948: 278)

⁹ e.g., Cowan et al. (1956) or Reines et al. (1960).

could cause the signals that their detectors recorded. For example, they investigated if anything else apart from a positron could be responsible for the first of the two peaks or anything else other than a neutron for the second one. These considerations, I claim, are much closer to the beer-and-fridge example I gave in section 3 than to the theoretical inferences of which Stanford's new induction speaks. I am not denying that the identification of causes in particle physics experiments is sometimes problematic, but these problems are of a technical and computational kind that has nothing to do with the supposedly fundamental *problem of unconceived alternatives*.

As an argument for realism about neutrinos, this story is incomplete, because it depends on realism about neutrons, positrons and the like; entities that Stanford would not allow into his ontology. Here is a sketchy suggestion about how it might be completed: What the neutrino story does show us is how the step from realism about fairly well-known entities (neutrons, positrons) to realism about a more hypothetical kind (neutrinos) can be performed. Iterating this process backwards, i.e., from neutrons and positrons to even more familiar entities, will ultimately show that our knowledge of neutrinos can be based on our knowledge of entities that even Stanford, being not a radical sceptic, has to accept. It will be noticed that this argument is similar to Kitcher's Galilean strategy, and if successful, it will undermine Stanford's antirealism in the same way that the Galilean strategy undermines constructive empiricism.

5 Compatibility with Kitcher's view on causation and possible objections

For this causal strategy to succeed, a robust notion of causality is called for. One may wonder if such a move would be acceptable to Kitcher, who, in his earlier work (1989: 497), showed little support for "the idea that there are causal truths that are independent of our search for order in the phenomena". But this is a topic on which Kitcher has changed his mind since 1989. As

he declares in his comment on my proposal, he no longer thinks that it is possible to avoid a robust commitment to causation by grounding our causal claims in a prior notion of explanation. One aspect of this change of mind becomes visible in Kitcher's defence of (Ex2), where he relies on a causal theory of reference which includes the notion that "the obtaining of a particular set of causal relationships is independent of human beings and their interests".¹⁰

The compatibility of my causal strategy with Kitcher's Realism reveals itself even more explicitly in Kitcher's forthcoming paper on "Mathematical *Truth*". He there contrasts examples of mathematical discoveries with examples of discoveries in archaeology, chemistry and biology, claiming that only the latter examples are genuine discoveries of pre-existing (but previously unknown) entities. The key difference between the two classes of discoveries seems to rest on the following assumption: "Satisfactory explanations of the discovery of a new entity X must offer an account of X and the ways in which it became accessible."¹¹ Such explanations cannot be given in the case of mathematical entities, while they can be given in the other cases, thanks to the causal properties of the discovered entities. Thus Kitcher seems to agree that we should accept as real those and only those working posits to which we can ascribe concrete causal roles.

I conclude this paper by mentioning two possible objections that might be raised against my strategy, and, if what I have just said is right, against Kitcher's account as well. The more obvious one comes, again, from Stanford. He will not be impressed with the causal role considerations I invoke in order to avoid the problem of unconceived alternatives. Anticipating something like my causal strategy, he writes:

"Although perfectly natural, this suggestion seems to run afoul of any number of discarded theoretical posits that *were* ascribed direct causal roles in the production of phenomena by the successful explanatory practices of their

¹⁰ Kitcher (2001: 186).

¹¹ Kitcher, "Mathematical *Truth*?" (MS), section III.

respective theories. To such familiar examples as phlogiston and caloric fluid we've seen that we might fairly add Darwin's gemmules, Galton's stirp, and Weismann's biophors." (Stanford 2006: 172)

Though I have not yet analysed these cases in any detail, it seems to me that this objection may be answered by being more specific about what kind of causal roles are necessary for a realistic interpretation. As we have seen, the neutrino was ascribed a causal role from the beginning, namely the role of carrying away some energy in beta-decay. This was not yet enough for realism about neutrinos. Later, however, the neutrino was ascribed a causal role in the much stronger sense of bringing about an event that would not otherwise have occurred. I doubt that there is "any number of discarded theoretical posits" that were once thought to be causally effective in this strong sense.

The second objection is probably more interesting, because it comes from the realist side. Kitcher repeatedly (and correctly, I think) stresses that realism is a piecemeal affair. We should only be realists about posits that become detectable by means of the causal work they perform. The worry here is that the properties we ascribe to these entities on the basis of their causal roles do not sufficiently determine *what these entities really are*. The resulting realism may seem to be a very poor one, saying merely: "there is something which does this and that, but I don't know what it is." Indeed, when I once presented my neutrino case study at a philosophy of physics workshop, one of the commentators said that this was all well and good, except that it wasn't realism anymore. More specifically, a comprehensive characterisation of the neutrino would have to rely on quantum field theory, and insofar as it is unclear how quantum fields can be causally efficacious,¹² they are precisely the kind of posits that the real realist would not want to call real. Hence, he would not be able to say what a neutrino is, only that it is something which is produced in beta-decay and can in turn produce an inverse beta-decay. If this is

¹² See Teller (2002: section 8.3).

sufficient to count as a realist, good. If not, we have to conclude that Kitcher's real realist is not really a realist after all.

References

- Chang, H.: Preservative Realism and Its Discontents: Revisiting Caloric. In: *Philosophy of Science*, 70, p. 902–912, 2003.
- Cowan, C. L./ Reines, F./ Harrison, F. B./ Kruse, H. W./ McGuire, A. D.: Detection of the Free Neutrino: A Confirmation. In: *Science*, 124, p. 103–104, 1956.
- Crane, H. R.: The Energy and Momentum Relations in the Beta-Decay, and the Search for the Neutrino. In: *Reviews of Modern Physics*, 20 1, p. 278–295, 1948.
- Elsamahi, M.: A Critique of Localized Realism. In: *Philosophy of Science*, 72, p. 1350–1360, 2005.
- Hacking, I.: *Representing and Intervening*. Cambridge: Cambridge University Press, 1983.
- Kitcher, P.: Explanatory Unification and the Causal Structure of the World. In: Kitcher, P./ Salmon, W. (eds), *Scientific Explanation*, vol. XIII of *Minnesota Studies in the Philosophy of Science*, p. 410–505, Minneapolis: University of Minnesota Press, 1989.
- *The Advancement of Science*. New York: Oxford University Press, 1993.
- Real Realism: The Galilean Strategy. In: *The Philosophical Review*, 110, p. 151–197, 2001.
- Makinson, D. C.: The Paradox of the Preface. In: *Analysis*, 25, p. 205–207, 1965.
- Psillos, S.: *Scientific Realism: How Science Tracks Truth*. London: Routledge, 1999.

- Reines, F./ Cowan, C. L./ Harrison, F. B./ McGuire, A. D./ Kruse, H. W.: Detection of the Free Antineutrino. In: *Physical Review*, 117, p. 159–174, 1960.
- Stanford, P. K.: *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press, 2006.
- Teller, P.: So What *Is* The Quantum Field? In: Kuhlmann, M./ Lyre, H./ Wayne, A. (eds), *Ontological Aspects of Quantum Field Theory*, Chapter 8, p. 145–164, London: World Scientific Publishing, 2002.
- van Fraassen, B. C.: *The Scientific Image*. Oxford: Clarendon Press, 1980.

Department of Philosophy
University of Lausanne
UNIL-Dorigny
Bâtiment Anthropole
1015 Lausanne, Schweiz
matthias.egg@unil.ch