Metascience
DOI 10.1007/s11016-009-9299-1

REVIEW SYMPOSIUM

GRASPING AT REALIST STRAWS

Kyle Stanford, *Exceeding Our Grasp. Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press, 2006. Pp. xiv + 234. US $65.00 HB.

By Juha Saatsi

The history of science really matters for the scientific realism debate. Studying scientific theorizing of the past can vindicate appropriate epistemic attitudes towards today's theorizing. But historical facts wholly unadorned by philosophical embellishments are not enough by themselves – history requires an interpretation. And that interpretation needs to feed into a philosophical argument that rules out alternative interpretations. Stanford's monograph gathers historical facts aplenty, offers interesting interpretations, and provides an important book-length argument for selective anti-realism. In what follows I will focus exclusively on some interpretational issues and in particular on Stanford's main argument against realism. This argument can be resisted, I believe, by de-emphasizing the role of explanatory success in the realist's game plan.

Amidst several interesting and insightful philosophical ideas that Stanford puts forward there is one that gives the book its name. The 'problem of unconceived alternatives' combines elements from the two rather well known traditional challenges to realism: the underdetermination problem, and the pessimistic induction. The new problem for the realist, like the pessimistic induction, takes the form of an inductive argument over the history of science – hence also the name "New Induction". But what exactly is *new* about this New Induction (NI)? And does its novelty transform or redirect the prevailing debate that has so far revolved around the pessimistic induction (PI)?

To get started, here's a rough summary of Stanford's argument (cf. pp. 29–30; all page references are to Stanford 2006).

Published online: 16 September 2009

(P1) *A historical fact:* scientists repeatedly fail to consider all plausible explanations for some phenomenon.
(P2) *A fact about the scientific method:* science often proceeds by eliminating all but one of the theories that scientists do consider, and we place our trust on that theory.
(C) *Conclusion:* this eliminative method is unreliable, since it doesn't take into account good theoretical explanations that scientists would not have eliminated had they considered them.

Let's grant the premises for now for the sake of analysis (I'll return to them below). Clearly the conclusion isn't entailed by the premises above, as it is compatible with them that the true theory is somehow always within the set of theories that scientists do in fact consider. But there's more to Stanford's argument than is presented above: it is part and parcel of NI's historical evidence for (P1) that scientists often fail to consider theories that are later *de facto* accepted, and these later theories are radically at odds with the ones that we earlier placed our trust in. This looks very much like the good old PI, so the "old induction" seems to play a major role in NI as well.

Realists have developed a variety of responses to PI, of course. A common element to many of these responses is the idea that the truth-content responsible for successful predictions can come apart from the explanatory resources the theory offers to the scientific community adhering to the theory. Hence the structural realist, for example, claims only that "getting the structure right" is responsible for Fresnel's success in deriving novel predictions, say, and *not* that Fresnel could have understood light phenomena in purely structural terms. The same holds for other (sensible) realists.

Acknowledging this much seems to allow the realist to respond to Stanford's NI as well. If realism is not concerned with theories latching onto the correct *explanatory* posits, then our scientific method can be reliable in a realist sense *even if* the correct explanation is *never* within the set of theories actually considered. All that matters is that the best-confirmed theory latches onto reality in those respects that brought about its successful predictions. Emphasizing predictive over explanatory success is natural given the theory-realist's appeal to the *No Miracles* argument, which concerns first and foremost novel predictive success. Stanford clearly feels that realism that doesn't commit to the explanatory posits of our best current theories is a realism not worth having. I'll come back to this worry about watered-down realism towards the end. In the meantime I want to analyze further the nature of Stanford's New Induction.

To my mind there is something fundamentally unappealing about the very starting point of Stanford's argument. For any philosopher of science, realist or otherwise, it is (or should be) a basic fact about science that science progresses by refining and occasionally overturning its explanatory assumptions. Any realist position must allow for such a non-linear form of progression, the non-linearity itself being just what we would expect: it would be naïve in the extreme to hope that theoretical inferences, especially those far-removed from everyday reality, would habitually latch onto reality *in every explanatory respect*, each theoretical advance simply retaining whatever was taken to be explanatory before. Thus the realist shares with Stanford the sensible assumption that human beings are not "cognitive supercreatures who are adept at conceiving of all possible theoretical explanations for a given set of phenomena" (p. 45). So the realist wants to allow for a degree of *iteration* in the progress of science, and the basic realist claim really should be (and for most realists it currently is, I believe) that such iteration isn't incompatible with progress; that there is an identifiable cumulative growth of scientific knowledge underlying the refinements and occasional overturns in the explanatory posits. The anti-realist engages in a kind of "judo-epistemology", to borrow Peter Lipton's apt term, when she attempts to turn the *strength* of the iterative aspect of the scientific method against the method itself.

So what in the end is new about NI? To be more precise, Stanford's NI is an *induction over theorists in the predicament of transient underdetermination by unconceived alternatives*. Let's unpack this a bit. The underdetermination is *transient*, because the alternatives aren't fully empirically underdetermined: theories are merely underdetermined by the evidence available at the time. The alternatives are *unconceived*, because the alternatives are not actually known to any theorist at the time in question. Finally, the induction is over *theorists* (as opposed to theor*ies*), unlike the case for PI.

On the face of it there's no denying that NI is different from PI, with a different inductive base, and a different conclusion. Stanford motivates NI by claiming it to be stronger than the traditional anti-realist challenges. It is arguably stronger than the classic underdetermination problem, which threatens to collapse into mere Cartesian skepticism with dreamed-up alternatives far removed from the actual scientific practice. It is arguably stronger than PI, because certain moves the realist has made against the latter are allegedly not available against NI. Namely: the realist has countered PI by claiming that many instances of radically false past theories are not relevantly similar to the current, mature, bona fide successful theories. By focusing on theori*sts*, rather than theori*es*, Stanford aims to make the historical inductive base immune to such a move, and more pertinent to our current science. After all, we have no reason to think that scien*tists* qua cognitive agents have changed all that much over the relevant historical period.

I'm very sympathetic to Stanford's attempt to bring underdetermination considerations down to earth from the giddy heights of *total* underdetermination by all possible evidence, but I'm less convinced by the contrast drawn between PI and NI, partly for reasons already given. Also, Stanford completely ignores the fact that a central realist move against PI has been to tighten the notion of success to *novel predictive success* in a manner that chimes with the realist's No Miracles argument. This limits the inductive base of PI considerably, and it also powerfully works against NI: the instances of unconceived underdetermination cited by Stanford by and large do not involve any novel predictive success! My basic worry about the claimed novel impact of NI can then be summarized as follows. The inductive base of NI either serves as an inductive base for PI as well – in which case it is not clear what has been achieved – or it is so limited that it properly serves as an inductive base for neither.

This worry is independent of whether or not we accept the historical premises, but there are reasons to worry about these premises in their own right. According to (P1) there are plenty of historical instances of theorizing where two theories *would have been* equally well confirmed, had both been available at the time. Unlike the standard underdetermination problem, we now have a counterfactual claim, and questions immediately arise whether such counterfactual inferences from the historical data are justified. The data Stanford provides is simply this. We have instances of incompatible but explanatorily successful theories T and T', accepted at times *t* and *t*', respectively. When T' is accepted, the original evidence, initially used to confirm T, is taken to confirm T'. On this basis the counterfactual claim is then made that T' would have been equally well confirmed at *t* as T was. But it is simply not clear to me on what grounds we can claim to know that the Newtonians, for example, operating in their cultural and scientific context, would have accepted the relativistic framework as a plausible alternative to theirs, given the data they had (Magnus, 2006). It is also not clear to me how the instances of actual theorists occupying the alleged counterfactual predicament are to be counted. Often it seems that the focus is on a *single* theorist propounding a certain theory as the only way (s)he can understand some phenomenon. But this seems too restrictive: surely 'a theory' – as a unit philosophically interpreted in a way that is relevant for the realism debate – gets confirmed over a period of time when the relevant scientific community comes to an agreement over it.

Coming back to Stanford's vision of what realism requires, he very explicitly acknowledges that NI fully hangs on his rebuttal of the realists' project of spelling out a notion of partial/approximate truth that is fit to describe a level of continuity across prima facie radical theory changes. It is illustrative of Stanford's general perspective on approximate truth that he chooses Newtonian mechanics as his initial prime exemplar of a theory which *cannot* be approximately true in any substantial sense, and which is unequivocally mistaken about the fundamental description of the physical world. Allegedly it is exactly cases like this that should give the realist pause, for might it not be "that all our own scientific theories are both fundamentally mistaken and nonetheless empirically successful *in just the same way?*" (p. 9). Stanford's choice of his prime exemplar is a telling one. After all, Newtonian mechanics doesn't feature prominently in Laudan's famous list, for example, despite "gravitational *force*" presumably being a non-referring term. And most of the realist literature has focused on the various caloric and ether theories, despite the fact that the shift from classical to relativistic mechanics and gravitation is one of the most pronounced radical shifts in our world-view. Why exactly isn't Newton's theory *the* standard exemplar against the realist? And why does Stanford choose it to be *his* exemplar?

The critical difference between Newtonian mechanics and Fresnel's ether theory, say, is that only in the latter does it appear prima facie that novel theoretical predictions were successfully *derived* from radically false assumptions about the world. Such instances (if there are enough of them) are problematic for a realist who maintains that novel predictions are symptomatic of the underlying truth, for deriving such predictions (repeatedly) from falsehoods falsifies the realist thesis that "success without truth is miraculous".

Newton's considerable theoretical successes, on the other hand, cannot be similarly attributed to a set of hypotheses about the fundamental nature of gravity, space and time. We can fully explain Newton's success in terms of what he came to understand truly about the world: that a rather large pool of data about moving bodies could be captured with outstanding accuracy by his laws of motion, that certain heavenly and certain earthly phenomena are relevantly similar in nature, and that the trajectory of a cannonball would eventually become a celestial circle if fired with a great enough initial velocity, and so on.

It is undeniable that the *explanatory* framework furnished by the Newtonian theory is radically at odds with the relativistic world-view, with geometrical explanations replacing causal action-at-a-distance, and so on. Indeed, it seems that some Newtonian explanations have simply *failed* – if we (plausibly, but not undeniably) require that successful explanations have to be true – and it is exactly this *explanatory failure* that Stanford wants to capitalize on. This forms a critical point of departure from the realist focus on *predictive successes*, and as far as Newton's predictive successes are concerned it is natural to take Newton's theory to be partially/approximately true. No physicist would ever suggest that we do not completely understand, from our current perspective, why Newton managed to make successful predictions. Although in places Stanford does talk about successes in predictions and interventions, in general his emphasis on explanatory successes is a premise that underwrites all the novel case studies that form the historical heart of the book. The theoretical 'successes' of the false theories of Weismann, Galton and Darwin, are invariably of the explanatory kind.

Admittedly there is some ambiguity and a lack of precision in the realist literature regarding exactly what kind of success is meant to indicate underlying theoretical truth, and to what extent. But given that many authors (Leplin, Psillos, Worrall, to name a few) have flagged the predictive dimension of success quite insistently and precisely, it is perhaps a bit uncharitable for Stanford to operate so flexibly himself. It is a positive sign regarding the current state of debate that realists are still actively grappling with these core issues, fine-tuning the notion of success and the realist explanation of it to fit the positive arguments for realism, and comparing the package that results with the historical record. Stanford feels otherwise, and views the debates over pessimistic induction as having "reached something of a stalemate", with more sophisticated realist maneuvers "whiffing of ad-hoc-ery". But failing to justify such sentiments, one is left to wonder whether the debate can really be advanced by painting with so much broader a brush than that used by his opponents. Far from being a stalemate, the current state of the debate just displays unavoidable increase in the degree of sophistication: evaluating inductive inferences, meta- or otherwise, is always a rather subtle business.

Let's now finally consider Stanford's general argument to the conclusion that realists have been forced to water-down their concept of partial/approximate truth to the point of effectively giving up the game. Chapter Six focuses on the referential status of central theoretical terms – in my opinion a partial red herring and a relic of the more linguistic approaches to the philosophy of science – and I'm by and large sympathetic to Stanford's treatment of these issues. My disagreements lie more with Chapter Seven, focusing on "selective realism", viz. the idea that (i) as matter of descriptive fact those features of theories that are really responsible for their predictive successes get carried over across theory changes, and (ii) the realist is justified in selectively committing to the corresponding aspects of our current theories. Stanford's main complaint here is that the realist cannot *prospectively* state exactly which aspects of our current theories are success-fuelling and hence expected to get retained. And, on the other hand, any *retrospective* identification of such aspects is – it is alleged – almost trivially guaranteed to find those aspects as success-fuelling which do happen to get carried over.

Why is the retrospective identification trivial? Stanford argues that it is because one and the same successor theory is used as a standard for identifying *both* the respects in which the past theory is true *and* the aspects of the past theory that enabled it to be successful. But are there not independent criteria for identifying the latter aspects? Surely there are. It is probably impossible to give a general recipe for this, for much hangs on the details of how the prediction was actually derived. But as a rule of thumb we can consider, for example, a set of properties attributed to a system by an earlier theory, such that for any system instantiating these properties we can deduce a prediction by writing down the same derivation. Representing the system as having these properties is then the sole success-fuelling element of the theory, regardless of whatever else the theory says about the system. We can then check from the vantage point of the successor theory whether the earlier theory attributes these properties to the system correctly or not (cf. Saatsi, 2005). We can easily envisage a situation in which such a set of properties underwriting the derivation in the earlier theory does not get attributed to the system by the successor theory, and there are some troublesome historical examples of this as well (Saatsi and Vickers, forthcoming). Although I admit that realists have been occasionally a little sloppy with their case-studies, Stanford hasn't shown that the strategy of retrospective identification cannot be rigorously implemented so as to dissolve the risk of trivialization. The requirement of prospective identification is too much to ask from the realist, but not because it would be an impossible endeavor. Indeed, I agree with Psillos (1999) that scientists themselves evaluate the 'working'/'idle' status of their theoretical posits every day.

At the heart of Stanford's disapproval of selective realism seems to lie the intuition that realism should deliver definite answers to definite questions such as: Should I believe in the explanatory Higgs mechanism for the generation of mass to the elementary particles? Should I believe in the explanatory molecular mechanism for gene recombination? The realist should openly confess to not being able give a clear-cut answer to such singular questions. But this does not mean, as Stanford would have it, that there is no interesting, carefully qualified, and weaker epistemological attitudes towards our best theories that still properly qualify as being realist. Knowledge of the unobservable admits many a degree.

Despite my reservations about the details of Stanford's book-length argument, I consider *Exceeding Our Grasp* to be a significant addition to the contemporary literature on the topic, presenting timely challenges to many common realist presuppositions. The book succeeds in putting considerable pressure on the realist camp: one must be quite precise in stating what "success" amounts to, what "approximate truth" amounts to, and exactly what the realist claims to know of the unobservable world. The historical case-studies are careful and thorough, and I applaud Stanford for getting his philosophical hands dirty with serious history of science. Such an approach is exactly what is needed to advance the realism debate. I'm also sympathetic with various aspects of Stanford's overall philosophical outlook, his naturalism, and his general approach to the epistemology of science. This book is guaranteed to engage anyone interested in the realism debate.

School of Philosophy
University of Leeds
Leeds, UK

By Stathis Psillos

Stanford has advanced a sophisticated neo-instrumentalist alternative to scientific realism. In fact, he has gone as far as to suggest that it may be a mistake to try to identify 'a crucial difference' between instrumentalism and realism when it comes to the epistemic attitudes they recommend towards theories or theoretical assertions. There is, he says, at most 'a local difference in the specific theories each is willing to believe on the strength of the total evidence available' (2006, p. 205; all references are to Stanford 2006). Although I welcome this attempt at reconciliation, I will argue that Stanford's own way to achieve it, while keeping a distinguished instrumentalist outlook, is flawed.

NEW INDUCTION VERSUS PESSIMISTIC INDUCTION

Stanford starts with a bold claim, viz., that at any given stage of inquiry there have been hitherto unconceived but radically distinct alternatives to extant scientific theories. When, in the fullness of time, these alternatives came to be formulated, they were equally well-confirmed by the then available evidence; they came to be accepted by scientists in due course; and eventually they replaced their already existing rivals. This is a condition that he calls 'Recurrent Transient Underdetermination'. If theories are subject to this predicament, Stanford argues, belief in their truth is not warranted. Not all theories are indeed subject to this, but Stanford thinks that all fundamental scientific theories in a variety of domains of inquiry suffer from recurrent radical underdetermination by the evidence. Based on evidence coming from the history of science, he performs what he calls the *New Induction*: there are good inductive reasons to believe that for *any* fundamental theory scientists will come up with – and for *any* evidence that will be available – there will be hitherto unconceived theories that will be at least as well confirmed as the ones available. This kind of situation is supposed to be the springboard for breathing new life into instrumentalism.

It promises to show that (there are good reasons to believe that) fundamental theories are not accurate descriptions of the deep structure of the world but rather 'powerful conceptual tools for action and guides to further inquiry' (2006, pp. 24–25).

Suppose, for the sake of the argument, we grant all this. It should be immediately noted that realism about fundamental theories would be in jeopardy only if the pessimistic induction were sound. The New Induction (NI) can only work in tandem with the Pessimistic Induction (PI). Unless PI is correct, NI does not suffice to show that the new and hitherto unconceived theories will be radically dissimilar to the superseded ones. Hence, rehabilitating PI is an important step in Stanford's strategy.

RESISTING PI'S REHABILITATION

Recent realist responses to PI have aimed to show that there are ways to distinguish between the 'good' and the 'bad' *parts* of past abandoned theories and that the 'good' parts – those that enjoyed evidential support, were not idle components and the like – were retained in subsequent theories. This kind of response aims to show that there has been enough theoretical continuity in theory-change to warrant the realist claim that science is 'on the right track'. This kind of response damages (at least partly) Stanford's unconceived alternatives gambit. If there is convergence in our scientific image of the world, the hitherto unconceived theories that will replace the current ones won't be the radical rivals they are portrayed to be. Claiming convergence does not establish that current theories are true, or likely to be true. Convergence there may be and yet the start might have been false. But the convergence in our scientific image of the world puts before us a candidate for explanation. The generation of an evolving-but-convergent network of theoretical assertions is best explained by the assumption that this network consists of approximately true assertions.

Stanford's main objection to this way of blocking PI is that it is tailor-made to suit realism. He claims that it is the fact that the very same present theory is used *both* to identify which parts of past theories were empirically successful *and* which parts were (approximately) true that accounts for the realists' wrong impression that these parts coincide. He writes:

With this strategy of analysis, an impressive retrospective convergence between our judgements of the sources of a past theory's success and the things it 'got right' about the world is virtually guaranteed: it is the very fact that some features of a past theory survive in our present account of nature that leads the realist *both* to regard them as true *and* to believe that they were the sources of the rejected theory's success or effectiveness. So the apparent convergence of truth and the sources of success in past theories is easily explained by the simple fact that both kinds of retrospective judgements have a common source in our present beliefs about nature. (2006, p. 166)

I find this kind of objection misguided. The way I see it, the problem is like this. There are the theories scientists currently believe (or endorse – it does not matter) and there are the theories that were believed (endorsed) in the past. Some (but not all) of them were empirically successful (perhaps for long periods of time). They were empirically successful irrespective of the fact that, subsequently, they came to be replaced by others. This replacement was a contingent matter that had to do with the fact that the world did not fully co-operate with the extant theories: some of their predictions failed; or the theories became overly ad hoc or complicated in their attempt to accommodate anomalies, or what have you. The replacement of theories by others does not cancel out the fact that the replaced theories were empirically successful. Even if scientists had somehow failed to come up with new theories, the old theories would not have ceased to be successful. So success is one thing, replacement is another. Hence, it is one thing to inquire into what features of some past theories accounted for their success and quite another to ask whether *these* features were such that they were retained in subsequent theories of the same domain. These are two independent issues and they can be dealt with (both conceptually and historically) independently.

They can be mixed up, of course. A (somewhat) careless realist could start with current theories and then try to tell a story about the abandoned and replaced ones such that it *ensures* that some of the theoretical assumptions about the world that scientists currently endorse were present in the past theories *and* responsible for their empirical successes. But carelessness is not mandatory! One can start with some past theories and try on independent grounds – bracketing the question of their replacement – to identify the sources of their empirical success; that is, to identify those theoretical constituents of the theories that fuelled their successes. This task won't be easy, but there is no principled reason to think it cannot be done. Unless, of course, one thinks that when a prediction is made the whole of the theory is indiscriminately implicated in it – but this kind of blind holism is no more than a slogan, or a metaphor. When a past theory has been, as it were, anatomised, we can *then* ask the independent question of whether there is any sense in which the sources of success of a past theory that the anatomy has identified are present in our current theories. It's not, then, the case that the current theory is the common source for the identification of the successful parts of a past theory *and* of its (approximately) true parts. Current theories constitute the vantage point from which we examine old ones – could there be any other? – but the identification of the sources of success of past theories need not be performed from this vantage point.

What needs to be stressed is that the realist strategy proceeds in two steps. The first is to make the claim of convergence plausible, viz., to show that there is continuity in theory-change and that this is not merely empirical continuity; substantive theoretical claims that featured in past theories and played a key role in their successes (especially novel predictions) have been incorporated (perhaps somewhat re-interpreted) in subsequent theories and continue to play an important role in making them empirically successful. This first step, I take it, is common place – unless we face a conspiracy of the scientific community to make us believe that every time a new theory is advanced and endorsed scientists do not start from square one (though they actually do). As noted above, this first step does not establish that the convergence is to the truth. For this claim to be made plausible a second argument is needed, viz., that the emergence of this stable network of theoretical assertions is best explained by the assumption that it is, by and large, approximately true. The distinctness of these two steps shows that Stanford's criticism is misguided.[1]

[1]Stanford (2006, pp. 167–168) ponders a somewhat similar line of thought on behalf of the realist, takes it to be promising, but dismisses it on the grounds that it is unconvincing: it is merely one potential explanation among others, including Stanford's own, viz., that our judgements about the truth of past theories and our judgements about their successes have a *common source*. I fail to feel the weight of this point. Stanford's own potential explanation is external: it tells us something about the source of the scientists' (or of the realists') judgements, viz. that it this source is current theory. Even if true, this line is compatible with in an internal potential explanation of the emergence of a stable network of theoretical assertions along the realist lines, viz. along the lines that being part of this stable network is best explained by being truthlike.

LIBERAL INSTRUMENTALISM

Stanford's instrumentalism is sophisticated and liberal. Stanford accepts that predictions are theory-driven and that they involve theoretical descriptions of whatever is predicted – being observable or unobservable. He puts no special epistemic weight on the observable-unobservable distinction. He takes it that our understanding of the world is theoretical 'all the way down' (2006, p. 202); that theories are our best conceptual tools for thinking about nature (cf. 2006, p. 207). In fact, his point of view is not instrumentalism *tout court*. According to his core characterisation of neo-instrumentalism, theories are predictive and inferential tools, but the inferences they licence – relying indispensably on theoretical descriptions – are not from observables to observables but 'from states of affairs characterised in terms we can strictly and literally believe to other such states of affairs' (2006, pp. 203–204). In other words, Stanford's instrumentalism relies on a body of *strict and literal beliefs* that form the basis on which the reliability of the theories as instruments for inference and prediction is examined. This body should, emphatically, *not* be associated with more traditional instrumentalist commitments to observables or to sensations and the like. How is it, then, to be circumscribed?

Instrumentalists have always taken it to be the case that no matter how attractive and useful theories might be, their 'cash value' has to do with what they say about the macro-world of experience. This is, in essence, what Edmund Husserl called the 'life-world', which he took to be the 'pregiven' world we live in. The content of this world might be understood narrowly or broadly (as Husserl understood it). Be that as it may, the point is that the content of the life-world is supposed to be accessed independently of theories. This is very similar to what Stanford thinks. He takes it that there is a part of the world to which there is 'some independent route of epistemic access' (2006, p. 199). To be more precise, Stanford claims that some parts of the world can be understood in terms of a theory towards which there can be no instrumentalist stance; these parts of the world (so characterised) will be the benchmark against which the reliability of the instrumentally understood theories is checked.

There are, however, a number of problems that Stanford's liberal instrumentalism faces. An obvious one is that it is unfortunate that this view is based on the presence of a body of strict and literally true beliefs. I doubt there are any such beliefs. Even if there are, they are not very interesting. Most beliefs of common sense – those that are supposed to form the backbone of the independent route of epistemic access to the part of the world that the instrumentalist is interested in – are neither literally true, nor strict and precise. Is the surface of the pool table flat? Well, it depends. Is the height of John 1.73? Close enough. Is the earth round? For all practical purposes. Is the sea-water blue? Not quite. Is a whale a fish? Not really. Do unsupported bodies fall to the ground? Yes, but… Does aspirin cause headache relief? It's very likely. This is just a figurative way to make the point. And the point is that common sense is *not* a theory towards which we can have a stance of strict and literal belief. Many terms and predicates we use in our commonsensical description of the world are vague and imprecise. Gaining independent epistemic access to the very entities assumed by the common sense requires leaving behind (at least partly) the common-sense framework.

Let us suppose Stanford is right in what he says about the network of strict and literally true beliefs we have with regard to common bodies. Stanford's strategy is overly conservative. It favours rather elementary theories. The irony is that it is *known* that the favoured theories are elementary, for otherwise there would be no motivation (acknowledged by Stanford's instrumentalism too) to advance more sophisticated theories (like proper scientific theories) so as to improve our understanding of the very objects assumed by the elementary theories. There is an issue of *motivation*, then: why try to devise theories? If the common sense framework is already in place and consists of a body of (strictly and literally) true beliefs, why not stay there? The answer, of course, is that we know that this framework admits of corrections and that scientific theories correct it (as well as their own predecessors). This is now commonplace: science is not just piled upon common sense. It adds to it and it corrects it. What Husserl did not consider when he aimed to show the priority of the life-world over the scientific image, Stanford and others have now taken to heart (cf. 2006, p. 201). But then there is an issue of *explanation* – which I will phrase in very general terms: if newer theories correct the observations, predictions, commitments etc. of their predecessors, they cannot just be more reliable instruments than their predecessors – this would not explain *why* the corrections should be trusted and be used as a basis for further theoretical developments.

This last line of thought might be pressed in several ways, but the most promising one in this context seems to be the following. Assume there is a network of strictly and literally true beliefs on the basis of which we start our endeavours. As science grows, there are two options. The first is that this network does not change – no more strictly and literally true beliefs are added to it as a result of the scientific theorising, theory testing, etc. This would render all this talk about the indispensability of scientific theorizing as little more than bells and whistles. The whole spirit of Stanford's liberal instrumentalism would be violated. The second option (rightly favoured by Stanford himself) is that the network of strictly and literally true beliefs is enlarged – more stuff is added when new theories are advanced, accepted, tested and the like. Note that Stanford holds no brief for the observable/unobservable distinction and does not restrict his realist stance to simple empirical laws and generalisations. What happens then? Every new theory will enlarge the domain that is interpreted realistically (subject to literal and strict belief). So every successor theory will have more realistic (less instrumentally understood) content than its predecessor. As this process continues (or has continued in the actual history of science), one would expect that at least some parts of the theory that Stanford treats as instrumentally reliable will become so involved in the interpretation of the realistic parts of the theory that it won't be cogent to doubt them without also doubting the realistically interpreted parts of the theory. I take it that this is the case with the atomic hypothesis, nowadays. But the point is very general. And it is that this kind of neo-instrumentalist image of science might well undermine itself and allow realism to spread indefinitely.

Stanford might well claim that realism will never spread to high-level and fundamental theories. He does admit that there are cases in which the available evidence constrains the space of competing available explanations – he argues, for instance, that the possibility that amoebas do not exist is ruled out. But he also insists that there are contexts – mostly having to do with fundamental physics – in which the evidence will never rule out alternative competing explanations. Perhaps, what has already been said in relation to the unconceived alternatives predicament is enough to make it plausible that there is no principled difference between being committed, say, to the reality of amoebas and being committed to the reality of atoms. If scientists are poor conceivers, why should we think they can succeed with amoebas but not with atoms?

Ultimately, Stanford's instrumentalist – like many other instrumentalists – relies on double standards in confirmation. The typical strategy here is this: the content of a theory is split into two parts – let's call them the OK-assertions and the not-OK-assertions, respectively. This partition can be made along several dimensions, but typically it is made along the lines of empirical versus theoretical or observable versus unobservable. The OK-assertions are said to be confirm*able* and confirmed by the evidence. Then the further claim is made (or implied) that the principles of confirmation that concern the OK-assertions are *not* transferable to the not-OK-assertions. Stanford seems to be in a worse situation here because he allows (in fact he requires) that some theories (and some claims about unobservables) are strictly and literally believed. But then he has to show that the ways in which the beliefs he allows (the OK-assertions) are confirmed are radically different from the ways in which the non-OK assertions confront the relevant evidence. No such case has been made. In fact, the very motivation for double standards in the theory of confirmation is deeply problematic. There is plenty of reason to think that the very same principles and methods are implicated in the confirmation of both the OK-parts and the not-OK parts. In other words, the very distinction between the OK-parts and the not-OK parts of a scientific theory is suspect.

Department of Philosophy and History of Science
University of Athens
Athens, Greece

By Rasmus Grønfeldt Winther

A Dialogue

Characters (in order of appearance):
Realist
Instrumentalist
Scientist
Constructivist

Setting: Equatorial rainforest, perhaps Brazil. Small clearing in the forest. Sounds of birds and monkeys. Light, but no direct sunlight. Air damp. Earthy smell. Two men, Realist and Instrumentalist, stand talking. Two backpacks lie on the ground.

Real: Good morning Inst. Did you sleep well in your tent?

Inst: Yes, thank you Real. It was a bit hot and damp, but our sleeping accommodations worked well enough. And you?

Real: No complaints. Well, let's get on with today's task of evaluating Kyle Stanford's *Exceeding Our Grasp*. His philosophical book on the role that unconceived alternatives might play in the history of science left me completely unconvinced. Rather, let me be more precise: his analysis seemed both trivial and false to me. At any rate, it certainly makes a mountain out of a molehill, or rather a galaxy out of a speck of dust.

Inst: Oh, this will be fun! I couldn't disagree with you more. I thought Stanford's book was a brilliant rendition of a deep and pervasive problem in science. May I ask what you thought was trivial and false about his argument? And how can it be both at the same time?

Real: Look, his basic argument is that given any synchronic slice of the history of science there are many unconceived alternatives to the accepted scientific theory. This is especially true for fundamental scientific theories of the very small, the very far away, or the very distant biological or astrophysical past. Indeed, there are numerous kinds of unconceived alternatives, found within radically different conceptual spaces, that we have not even considered as viable theories. Let me find Stanford's book and quote to you [unzips backpack, rummages around, and takes out *Exceeding Our Grasp*]:

> …the problem of unconceived alternatives worries that there are theories that we should and/or would take seriously as competitors to our best accounts of nature if we knew about them, and that could or have been distinguished from them evidentially, but that are excluded from competition only because we have not conceived of or considered them at all. (Stanford 2006, p. 23)

Notice also that even in his title, he refers to the existence of unconceived alternatives as a "problem." I fail to see the problem. First, triviality. Describing scientific change in terms of the existence of unconceived alternatives is simply another way of saying that science clearly progresses through the replacement of old theories by new theories. Can anyone dispute the truism that new theories were unconceived prior to their birth? Stanford is providing a trivial restatement of the growth of theoretical science. Second, falsity. Even if we accept the there are always unconceived alternatives, as I also do, a problem or worry does not arise. Stanford correctly observes that we are most likely wrong about our best current theories, but future alternatives will retain the true, good part of our current theories. There has been, and will always be, partial continuity of (1) reference of theoretical terms and (2) mathematical formalisms between successor theories. There is nothing to worry about, epistemically. So by calling the admittedly real existence of unconceived alternatives a "problem", Stanford is sketching a false and twisted evaluation of the perfectly healthy development of scientific theory in its fallible discovery of truth. In short: he redescribes progress in the history of science in a trivial manner, and then glosses that redescription such that it falsely evaluates that very history.

Inst: Strong words! Fighting words! Maybe we should take a step back. I don't think that you are characterizing the problem of unconceived alternatives fairly.

Real: Really? Why not?

Inst: Stanford is adding something genuinely new and cogent to the philosophical discussion. Pierre Duhem taught us that every theory is underdetermined by the evidence. That is, many alternative theories, empirically equivalent to a given theory, are possible. From Larry Laudan we learned that the history of science gives us every reason to believe that even the best theories of a particular historical epoch will eventually turn out to be false. Stanford combines these insights and adds at least three further points. First, that we are never aware of the full range of real alternatives to contemporaneous theories and thus that eliminative induction is a problematic strategy for theory choice, even though inference to the best explanation remains "the central inferential tool of scientific inquiry" (p. 30). Indeed, in the history of science many real alternatives to established and important theories have eventually appeared. Stanford shows this for the case of three nineteenth century biological theories of generation and inheritance: Darwin's, Galton's, and Weismann's. In order to articulate alternatives, we don't have to rely on Craigian reductions or on far-fetched toy alternatives. Furthermore, other such real alternatives could and should also have been articulated. Thus, given the problem of the limited range of theoretical alternatives available to eliminative induction, shouldn't we believe that there are better, unconceived alternatives to our current best theories? Second, Stanford argues that unconceived alternatives are often radically and fundamentally different from standard, accepted theory. They may not even be in the same conceptual space. Third, Stanford convincingly argues that the alternatives must account for much of the same data, but need not be completely empirically equivalent at the particular moment of their appearance and certainly not during their respective careers. Indeed, one of them will eventually fare better empirically. The problem of unconceived alternatives is a real problem, and Stanford has done us a service by clearly diagnosing it. It is neither trivial nor false, nor both.

Real: I still fail to see the full depth of the problem. However, two things that you said worry me. The first is that the unconceived alternatives may be in radically different conceptual spaces. Does this mean that there will be no continuity in reference and formalism between them? The second is that empirical equivalence may just be partial. Doesn't any theory have to account for all the relevant and good data we gather from our investigation of the real world? But wait, Inst. Why don't you go see if our friend Scient is awake? She is a first-rate scientist and thinker who might be able to help us with some of our worries. Of course, since she is not a philosopher, we will have to translate and adopt her arguments in appropriate ways. Be gentle, she's a bit tired since she's been on a US National Science Foundation panel reviewing more than one hundred grant proposals in evolutionary genetics.

Inst: OK.

[Scientist walks onstage just as Inst starts to turn around.]

Scient: Good morning. I see that you the two of you are already awake. What are you philosophers discussing today?

Inst: Well, we've been reviewing some of the basic arguments of Kyle Stanford's new book and finding ample room for disagreement.

Scient: I see. Well, after reading it, I was particularly struck by the last chapter. I do not see exactly why there is or, more precisely, should be, such a big disagreement between you. Stanford draws the distinction thus:

> The characters traditionally identified as the realist and the instrumentalist both recognize theories that they strictly and literally believe to be true and theories that they think are merely instrumentally useful over a wider or narrower domain of nature. The instrumentalist simply assigns a much larger set of the theories we actually have to the latter category. (p. 205)

This seems to be a crisp and useful contrast. Do you agree with the characterization?

Real: Yes.

Inst: [a split second later] Yes.

Scient: Fine. But I do not actually understand the difference between these positions. How can the true be true without being useful, and how can the useful be useful without being true? I might concede that truth and utility are two different aspects or properties of a scientific theory, model or law, or propositional attitude more generally, but my fallibilist scientific stance inhibits me from separating them starkly. For me as a scientist, truth and utility go hand in hand.

Real: I am not sure that I follow. You mean to say that there is no difference that makes a difference between Inst and me?

Scient: Here's another way of making my point. Perhaps there is a difference between you and Inst that makes a philosophical difference. Even there, though, I am reminded of Arthur Fine's and Richard Rorty's nice deflationist or quietist arguments regarding the realism debate. More importantly, I am not sure that there is a difference between you that makes a scientific difference. Should any of us, scientists, philosophers, or the so-called layperson, believe in the entities and processes that theoretical science claims to be discovering? Well, some scientists do, some don't, and most don't really think too much about it. Yet, this diversity of ontological commitment to the entities and processes of a theory does not seem to make much difference to the development of new theories or the progress of science more generally.

Inst: [Annoyed] Surely it must make a difference! If a scientist doesn't believe in the entities or processes of the accepted theory, then she will be more open to other theories, with distinct ontological possibilities.

Real: I disagree, Inst. On the contrary, if a scientist believes in the entities and processes, then she will be more open precisely because she will want to test the validity of the postulated entities and processes by comparing them with alternatives.

Scient: [Laughing] Ahem, I think I was interrupted, but I find it to be rather endearing that you can't even agree on the consequences resulting from an individual or a community expressing strong ontological commitment to the entities and processes of a given theory. Please permit me to return to the effectively null relevance of the diversity of ontological commitment to the actual practice of science. Remember that I am a scientist, so I should know, right? First of all, articulating unconceived alternatives depends less on particular ontological commitments and more on changes in, and interactions among, other factors: (1) material instruments and technologies, (2) modeling practices, (3) general, unifying theories, and (4) methods of data collection and analysis. Second, particular ontological commitments vary tremendously across a scientific community at a given time, and even over the career-path of specific scientists. Moreover, the commitments, even when stated explicitly in scientific texts, which seem to be all that you philosophers ever use as your data, often do not constrain our scientific work or do so only in very indirect and even puzzling ways. To summarize, the two of you can disagree as much as you want about whether scientists, philosophers, or laypeople should believe in the postulated entities and processes of the best scientific theories, as if there were single monolithic best theories. However, I don't think that ontological commitment makes much difference to actual scientific practice.

Real: You are shifting the terms of the debate!

Inst: Yes, you are changing the questions. Real and I have serious disagreements. I emphasize utility, he highlights truth. Utility and truth are different! We also care much more about the ontological status of the central theoretical terms of scientific theories, than about scientific practice. Moreover, as Stanford shows, what you call ontological commitments are indeed relevant during the history of science. That is, when scientists such as Darwin, Galton or Weismann constrain their respective, sequential views to particular classes of entities or processes, they cannot even consider alternative representations of heredity. For instance, Stanford shows that Darwin couldn't even imagine an explanation of heredity based on the continuity of the germ-line rather than on the collection of material particles (gemmules) from the entire body of the parents (Chapter 3). Thus, the ontological commitments of specific scientists lock them in. The danger then is that these extremely influential individuals feel that they have provided the only possible view on nature and that their theory is approximately true.

Scient: So what? Won't the massive diversity embodied in actual scientific communities liberate us from the ontic narrow-mindedness of single scientists? Why is the admittedly real existence of unconceived alternatives a problem? Won't science continue progressing? Yes, our best theories are fallible and imperfect, but aren't they useful and approximately true?

Real: [A big smile on his face] Well, well, well. Haven't we heard this before? I have to admit that although I think that you are missing the point by claiming that Inst and I don't have any real disagreements, I rather like the questions you pose. Can he, or Stanford, evade them? Sometimes I wonder, are we just stuck in a battle of intuitions and a constant shifting of the burdens of proof to the other?

[Inst scowls]

[From stage right a man appears. Constructivist is blindfolded and has a scarf covering his mouth. Real and Inst look at each other and snicker. Scient looks slightly horrified. She frees Const by untying the blindfold and the scarf.]

Scient: Const, I presume? Here we are in a rich rainforest ontology discussing Kyle Stanford's new book. No Quinean desert ontology here. Would you like to join us?

Const: Indeed. I have wanted to do so for quite some time, but somehow I often seem to be written out of the scientific realism debate in the philosophy of science. The usual contrast there is between realists and instrumentalists, or realists and empiricists, or realists and deflationists. I am practically never given a voice, or I am confused with the black sheep of my family, Social Const.

Scient: Well, maybe Real and Inst indeed have nothing to say to you, but I would like to speak and listen to you.

[Real and Inst stare off into space, shuffle impatiently, and start whistling.]

Const: While reading Stanford's book, I couldn't help but feel that a whole perspective on the realism debate was missing. We are told that there are two attitudes we can take towards a theory: believing that it is literally true, or merely believing that is instrumentally useful for predicting, explaining, and intervening. The instrumentalist holds the second attitude towards many more theories than the realist.

Scient: [Impatiently] Yes, we heard this earlier.

Const: OK, but in both cases the data are taken as bedrock. The data and their production are rarely interpreted as philosophically problematic. Moreover, theory is understood as a representation, true or instrumentally reliable, of that data. I believe this family of views on theory and data to be mistaken. As a philosophical constructivist, I care about the theory-ladenness of observation, whose nature and extent varies from theory to theory. I am also concerned with the role general ontological assumptions, theoretical principles, and standards of evidence play in the production and evaluation of evidence. Let me mention two of my philosophical heroes. In the contemporary literature, Michael Friedman (1999) has important things to say about neo-Kantianism and the constitutive yet revisable "relativized a priori." This relativized a priori includes principles of coordination between abstract mathematical theory and concrete empirical data (e.g., *ibid*., pp. 79–80). Moreover, Ian Hacking develops the promising notion of "styles of scientific reasoning" which introduce "new types of: objects; evidence; sentences, new ways of being a candidate for truth or falsehood." (2002, p. 189; see also 2009) It is the constitutive and constructive role of these types of cognitively and socially embedded abstract assumptions, principles, and standards that I seek to explore. These are not even remotely mentioned in Stanford's book. I care about human, pragmatic agency!

Inst: Why should your assumptions and so forth be addressed by Stanford? What are they? How could they even help us understand the problem of unconceived alternatives any better?

Const: First of all, I accept that there are unconceived alternatives, but this is not exactly a problem because we do make progress in science…

Real: [Rudely] Yes, we have heard this before, even if I suspect that your reasons for holding these views will be different from mine. Let's move on, our time is running out. The sun advances. We must clean up our camping ground soon and start an exploration of the local fauna and flora, guided by Scient.

Const: OK. I suppose that this is the usual story: I don't even get my 15 minutes of fame, let alone attention! Well, here are two points of contact between my position and the problem of unconceived alternatives. First, if we accept that there is a problem with unconceived alternatives, then I think that we would should try to find solutions. Shouldn't philosophy be as much about solving problems intelligently as about discovering and diagnosing them? So I suggest that we should study the role our cognitive machinery, with its biases, heuristics, and norms, plays in the production of data and development of new theory. This cognitive machinery is itself trained by the theory within which it works. A study of these biases, heuristics, and norms should be both empirical and philosophical. How do we actually go about discovering and articulating new unconceived alternative theories, models, and laws at a cognitive level? Can we use computer science or psychology to model and understand this process? William Wimsatt (2007), for instance, has interesting things to say about this. More philosophically, what effect do cognitive biases and general theoretical principles have on the structure and confirmation of our best current scientific theories? How might the Semantic View or Structural Realism, or Bayesian confirmation theory be modified to take into account such biases, heuristics, and norms? Moreover, can we even separate empirical and philosophical questions in this context?

Scient: [Surprised] That's kind of interesting Const! What is the second point of contact between your views and the problem of unconceived alternatives?

Const: Well, I disagree with Stanford that there is no selective confirmation of different robust parts of our theories. He contends that:

> without some *prospectively applicable* and *historically reliable* criterion for distinguishing idle and/or genuinely confirmed parts of our theories from others, the strategy of selective confirmation offers no refuge for the scientific realist. (p. 169)

I agree with the conditional since it is a point of logic, but I think that such a criterion exists. Kuhn discusses it in the last chapter of his wonderful 1962 book: *the ability to solve problems*. There is a continuity and accumulation of problem-solving strategies across sequential theories. Working strategies become entrenched, idle ones don't. Genuinely confirmed problem-solving strategies are used in, and become part of, our laboratories, models, instruments, and theories. For instance, thermometers, differential equations, model organisms, supercolliders, and valence bond theory became standard, working parts of future scientific inquiry. So do certain entities and processes, even those of our so-called "fundamental theories of nature" (p. 32): genes, space-time, and molecules. These are also carried forward as effective, though fallible and flexible, strategies for solving problems. They can even be seen as constitutive and revisable principles for further scientific work. Moreover, vis-à-vis data and theory, I endorse a "top-down" rather than the "bottom-up" thesis Real and Inst propose. In an important sense, I think that scientists usefully reify the entities and processes they employ in their scientific inquiry.

Scient: Interesting again, although a bit opaque. I look forward to seeing the further development of some of these ideas in print soon. I sense, however, that neither Real nor Inst is impressed. Well, perhaps we should call it a day then. Gentlemen, it has been a pleasure to spend the morning with you discussing themes in Stanford's new book. The point here has not been to review each argument, but to show how different philosophical packages, represented by each one of us, would evaluate some of the central claims made. There is no question that Stanford's valuable book will continue to provide much food for thought. But perhaps Real and Inst have monopolized the discussion for too long? I feel that we should let in more voices in the realism debate. Perhaps we could then learn to be more scientifically honest and socially engaged? Moreover, maybe it would then be possible to finally achieve mutual understanding, and perhaps even peace, in this philosophical debate? Let us now turn to the rich ontology of our biological surroundings. Const, would you like to join us?

[Inst embraces Real and gives him a kiss on the cheek. Scient approaches Const and makes small talk. The four exit stage left.]

Philosophy Department,
University of California,
Santa Cruz, CA,
USA

Author's Response

By P. Kyle Stanford

I first want to thank my commentators for the constructive spirit in which they have approached my work and the care with which they have considered it. In response I'll begin with some small but crucial points of contention about what the central argument of *Exceeding Our Grasp* really requires. The New Induction (NI) claims that at the time we have embraced any given scientific theory on the strength of a given body of evidence, there typically have been fundamentally distinct theoretical alternatives also well-confirmed by that evidence (often including those accepted by later scientific communities) that simply remained unconceived by theorists at the time.[2] Because this predicament has recurred systematically throughout the history of virtually every scientific field, we have every reason to believe and no reason to doubt that it is probably our current predicament as well.

But following Magnus (2006), Saatsi worries about how we can know that "the Newtonians, for example, operating in their cultural and scientific context, would have accepted the relativistic framework as a plausible alternative to theirs, given the data they had". To insist that any genuinely worrying unconceived alternative meets this demand, however, surely gets things the wrong way round. The NI suggests, after all, that we should expect to find presently unconceived theoretical alternatives that stand in just the relationship to present evidence as, say, General Relativity did to the evidence for Newtonian Mechanics or Mendelian genetics did to the evidence for Darwinian Pangenesis, and these previously unconceived alternatives would *actually come to be accepted* by later scientific communities. Thus, if features of the "cultural and scientific context" would have prevented such alternatives from being regarded as "plausible" at the time that they remained unconceived, this should lead us to worry about the stability of whatever features of the "cultural and scientific context" inform our *own* judgments of plausibility, not to conclude (quixotically) that genuinely threatening unconceived alternatives never really existed and therefore probably do not now!

To put the matter another way, insisting that Newtonians would have rejected General Relativity as implausible only helps to undermine the force of the NI if we *also* assume both (i) that this is what prevented Newtonians from conceiving of General Relativity in the first place *and* (ii) that there are no comparable changes in store for scientific communities of the present day in the background assumptions, range of evidence, or whatever other features of the "cultural and scientific context" are supposed to have grounded both the (hypothetical) judgment that General Relativity was not a plausible competitor and the (consequent) failure to conceive of it. After all, features of scientific and cultural context have *actually*

[2]Note that although a serious version of the NI strictly requires only that past unconceived alternatives were *not ruled out* by the available evidence (cf. p. 19n.), I sought to defend the stronger claim of equal confirmation for past unconceived alternatives in part to help deflect the suggestion that such alternatives remained unconceived only because they were poorly confirmed at the time (see below).

varied historically in ways that have undermined such previously entrenched judgments of implausibility, and we would seem to have little reason to believe that we are now at the end of this process or that our grasp of the relevant evidence is now substantially complete. So if indeed Newtonians would have dismissed General Relativity as "implausible" had they considered it, this more forcefully challenges the idea that we can straightforwardly rely on our own standards of scientific plausibility than the idea that the epistemic predicament of earlier scientific communities is also our own. And the fact that previously unconceived alternatives like General Relativity have been ultimately accepted is important to the NI not because it implies that earlier theories have been overturned, but instead because it shows that these theoretical possibilities were, in addition to being supported by the available data, scientifically serious and "plausible" in the only sense that really matters here.

Of course, in *Exceeding Our Grasp*, I tried to evade the need to debate any claim of relative fixity or privilege for relevant features of our own cultural and scientific context. My detailed historical investigation of nineteenth century theories of inheritance and generation sought to illustrate that we are not good at exhausting the space of well-confirmed theoretical alternatives *even within a single shared "cultural and scientific context"*. Thus I do suggest that Darwin, for example, would have regarded the fundamental mechanisms of inheritance later proposed by Galton, Weismann, and even Mendel as perfectly serious and plausible competitors to his own account had he managed to conceive of or consider these possibilities (which is not to say he would ultimately have accepted any of them). And if this pattern is indeed general, it gives us good reason to doubt that we are conceiving of all the well-confirmed theoretical alternatives that would count as "plausible" even by just the lights of our *own* "cultural and scientific context".

As Winther is centrally concerned with what is being left *out* of this story, I should perhaps say explicitly in this connection that I fully share the interest of his Constructivist in how the details of human pragmatic agency and cognitive machinery affect the constitution of evidence and the processes of scientific change more generally, but all of this seems to me to complement my aims in *Exceeding Our Grasp*, rather than compete with them. Indeed, what his Constructivist proposes is largely an empirical exploration of the various dynamical processes that help explain how and why particular unconceived alternatives remain unconceived by particular (human!) scientists and scientific communities. Of course, in *Exceeding Our Grasp* I also tried to avoid tying my exploration of the reality and consequences of our repeated failures to conceive of the full range of well-confirmed theoretical alternatives to any particular account of the (presumably heterogeneous and untidy) *sources* of those failures.

In any case, whatever the reasons for our repeated failures to conceive of the full range of well-confirmed theoretical alternatives even within a single historical and scientific context, it remains true that realists have sometimes sought to insulate present theories from invidious comparison with their predecessors in ways that, if successful, might help to undermine the NI as well as the PI. In this connection we should consider Saatsi's claim that I ignore the recent realist focus on predicting novel or surprising phenomena as the kind of success that genuinely requires the truth of a theory that enjoys it. This emphasis on novel predictive success is itself, of course, nothing new: it has been a staple of realist argument at least since Herschel and Whewell used it to defend what they saw as legitimate uses of the method of hypothesis against inductivist critics (see Laudan, 1981, p. 127ff). But notice that such an appeal would have to work quite differently against the NI than the PI. Against the PI, realists can plausibly suggest that it is simply unwarranted to draw conclusions about the ultimate fortunes of theories that did not enjoy any novel predictive success from the ultimate fortunes of those that do – after all, this is a difference between the two sets of theories that might well make it inappropriate to project inductively from one to the other. But the same appeal will not work in the case of the NI, for here we are instead projecting from the repeated failure of past theorists and scientific communities to conceive of the full range of well-confirmed alternative theoretical possibilities to the likely failure of present scientists and scientific communities to do so. The fact that some of the *theories* we have discovered along the way manage to successfully predict the existence of novel phenomena does nothing to show that the attempts of past *theorists* to exhaust the space of theoretical alternatives well-confirmed by the evidence are relevantly unlike the attempts of present *theorists* to do so or that there cannot be well-confirmed unconceived alternatives to a theory that enjoys success of this kind. So the appeal must be different. The claim must be that novel predictive success is a reliable sign that we have found a true theory *even if there are* well-confirmed alternatives to it that remain presently unconceived: we can safely ignore any such alternatives because only the (approximate) truth can (probably) have novel predictive success, so the novel predictive success of our present theory tells us that we have (probably) already found the (approximate) truth.

Even if this claim were true, of course, it would simply give up realism concerning theories in the many domains of science for which confirmation seems to come by way of broad explanatory scope and unifying power (important parts of the biological sciences, for example) rather than predictive success of any kind, much less novel predictive success. But in any case, *we already know that the claim is false*. Many predictions of novel phenomena have been made by theories that have turned out to be fundamentally mistaken, including the paradigmatic example of Fresnel's formulation of the wave or "undulatory" theory of light, which predicted that there should be a bright spot at the center of the shadow of a perfectly circular disc – a prediction famously treated as a *reductio* of the theory (by Poisson) until Arago bothered to actually perform the relevant experiment! Notice that despite Saatsi's suggestion that the PI "seems to play a major role in NI as well", and Psillos' suggestion that the NI places realism "in jeopardy only if the pessimistic induction were sound" (see also Chakravarrty, 2008), the difference between the PI and NI is crucial here: even if (as Saatsi suggests) the class of theories with novel predictive success that have subsequently been overturned is indeed too small to form a convincing basis for an inductive projection, it is nonetheless large enough to undermine the view that novel predictive success is a clear sign of (approximate) truth that therefore allows us to simply dismiss our independently motivated worries about the possibility and significance of unconceived alternatives.

In any case, it is not quite fair to suggest that my discussion *completely* ignores the recent realist emphasis on the importance of novel predictive success. I explicitly note that some of the theories of inheritance and development I discuss to which well-confirmed alternatives remained unconceived enjoyed such novel predictive success, such as Weismann's widely influential theoretical prediction of the need for a reduction division in the formation of sex cells, and the prediction made by teleomechanist thinkers that gill slits should be found in the course of human ontogenetic development. These examples further illustrate that novel predictive success is no proof of either the approximate truth of a theory that enjoys it or the absence of fundamentally distinct unconceived alternatives to that theory that are also well-confirmed by the evidence in support of it (including the ability to make the very prediction(s) in question).[3] Defenders of novel predictive success might try to seek criteria of novel prediction that would exclude such problematic examples, but down this road looms an unpleasant dilemma for the realist. The stricter her criteria for genuinely novel predictive success, the more of contemporary science she excludes from realist treatment while nonetheless failing to eliminate troublesome paradigmatic instances like the Poisson bright spot. But the more permissive her criteria of genuine novel predictive success become, the more conclusive evidence history provides that such success is no clear sign of truth, the absence of unconceived alternatives, or anything else that will help her case.

Of course, Saatsi makes it clear that his favored response to this sort of challenge is to weaken the epistemic entitlements upon which he thinks sensible realists should insist, and both he and Psillos think that I have unfairly evaluated the strategy of selective confirmation, in which realists argue that only those *parts or aspects* of theories centrally implicated in their empirical achievements (especially their achievement of novel predictive success) are to be believed. In *Exceeding Our Grasp* I argued that such selective realism was not genuinely supported by the merely retrospective judgment that the parts of past theories that were truly essential to their empirical successes are just those that have turned out to be accurate: this convergence is explained just as well by the fact that realists evaluate *both* what a past theory got right *and* what was truly necessary for its success using our present theoretical account of

[3]Such examples also introduce a version of what I have sometimes called the "threshold" problem for scientific realism: even if we concede that (some) contemporary theories have *more* novel predictive success than any ultimately rejected past theory, this gives us no reason to believe that we have now crossed over some kind of threshold in this respect such that these predictive powers are now finally substantial *enough* to ensure that there are no presently unconceived well-confirmed alternatives to those theories or that they must be true even if there are.

the relevant scientific domain. Such a strategy of analysis is virtually guaranteed to produce the convergence that the realist celebrates between the parts of past theories judged both true and essential for success whether the present theories used to make this evaluation are themselves even approximately true or not. In the absence of some *prospectively applicable* criterion for identifying the essential parts of a theory in advance of later developments, then, selective realism gets no credit for passing a test it could hardly have failed.

But Psillos, Saatsi, and even Winther's Constructivist insist that the required judgments could have been and can now be made prospectively: we can, they suggest, examine a theory in isolation to determine which of its parts, aspects, or elements are genuinely required for any (successful, novel) predictions it has made (or are involved in its successful "problem-solving strategies") and which are superfluous, and then proceed to ask independently whether those parts, aspects, or elements have been subsequently preserved in and/or ratified by the lights of present theories. If this claim is correct, of course, it introduces something of a puzzle: why did we (or the relevant scientific communities) *ever* believe more than those parts or aspects of past theories on which their empirical successes really depended? Saatsi concedes that, at least prima facie, the novel predictive successes of Fresnel's wave theory depended on radically false theoretical assumptions, but insists that the empirical successes of Newton's mechanics depended only upon the fact "that a rather large pool of data about moving bodies could be captured with outstanding accuracy by his laws of motion, [etc.]". If so, why did the many reflective and methodologically scrupulous Newtonian scientists ever believe anything more than the laws of motion themselves? And why did Priestly, or Maxwell, or Lavoisier, or Darwin, or Galton, or Weismann, not to mention their contemporaries, ever believe more of what their own theories proposed about nature than those parts or aspects genuinely required for their successes?

Saatsi admits that there is no general recipe for prospectively identifying the aspects of a theory that ground its success, but also denies the need for one, noting in agreement with Psillos (1999) that "scientists themselves evaluate the 'working'/'idle' status of their theoretical posits every day". This claim is surely right, but the case I made against Psillos in *Exceeding Our Grasp* involved showing that scientists' own judgments about which parts of their theories are crucial to their successes and/or most likely to be preserved in later theories are routinely *mistaken* in central cases (including, it turns out, in the two central historical cases Psillos offered in support of his claim).[4] Presumably our own prospective case-by-case judgments of what is really required for the success of a theory or what is most likely to be preserved from it are not likely to be better or more reliable than those of scientists themselves, but the historical evidence suggests that theirs are not nearly reliable enough to bear the epistemic weight that selective realism would need to lay upon them.

On what, then, will Saatsi, Psillos, or any of Winther's backpackers rely in determining which parts of our theories we are entitled to believe in light of their successful prediction of novel phenomena? Even our retrospective judgments do not pick out any single feature or aspect of theories that is invariably implicated in any predictive success they might enjoy. In the case of Newton's Mechanics, Saatsi tells us, the theory's empirical successes required only the truth of the laws of motion, but the realist cannot hold that such phenomenological regularities are what we should generally expect to find preserved in theoretical transitions: this would simply be constructive empiricism by another name. In the case of Fresnel's wave theory Saatsi seems happy to allow that a correct identification of the "structure" of nature produces the theory's novel predictive success. In still other cases it seems by present lights to have been the postulation of particular entities that was crucial to a theory's novel predictive success, as is perhaps evident in the case of atoms and Brownian motion (often treated as a "novel" prediction because Brownian motion was no part of what the atomic theory was originally developed to account for). Winther's Constructivist points us towards an even broader and more heterogeneous array of features that can be preserved from the "problem-solving strategies" of a theory to its successors, but all this simply testifies to the fact that no *one* feature or aspect of scientific theories is invariably implicated in any novel predictive success it enjoys

[4] In more recent work (2009) I have argued that scientists' confidence that particular aspects of their own theories will be preserved in their successors is often explicitly grounded in the fact that they cannot conceive of any other possible cause for a given phenomenon; if so, the problem of unconceived alternatives helps to explain why scientists' judgments of essentiality or likely persistence are themselves unreliable.

and/or can therefore be confidently predicted to survive further theoretical upheaval or replacement by presently unconceived theoretical alternatives.

Perhaps Saatsi, Psillos, and other suitably modest realists would be content with the bare and unimprovable insistence that *something* from any sufficiently successful present theory will *somehow* be preserved *somewhere* in their successors, but this seems rather a slender reed on which to hang the realist banner. Indeed, I myself do not doubt that the empirical successes of our best scientific theories (or their predecessors, for that matter) obtain in virtue of *some* complex and interesting interconnections between those theories and the world, and I certainly believe that there will be *some* systematic relationships and important continuities between present theories and their even more empirically powerful successors. But history seems to teach us that we are in no position to say with confidence what those interconnections are or how such continuity will be realized in any particular case of fundamental scientific theorizing. And if we cannot even specify in advance which parts or aspects of our best scientific theories are true and/or will be preserved in any of their successors, we seem to have given up what the realist cared most about all along.

This, of course, brings us to Psillos' concerns about the form of instrumentalism I advocate as the appropriate epistemic attitude towards many of our fundamental scientific theories. He suggests that like many illustrious predecessors this view sets aside a special group or class of beliefs as those towards which an instrumentalist attitude simply cannot be adopted, but this description strikes me as obscuring more than it illuminates. On the view I advance, we could in principle adopt an instrumentalist stance towards any particular belief or set of beliefs we have, treating them simply as powerful conceptual tools for mediating among other beliefs we are not treating in this same way. The point is not that any particular type of belief is automatically privileged or special, but instead that we cannot coherently adopt such an instrumentalist stance towards all of our beliefs at the same time. And as it turns out, we have good epistemological reasons for taking such an instrumentalist stance towards some of our beliefs and not others.

The instrumentalism I advocate is strictly modeled on the realist's *own* attitude towards a theory like Newtonian mechanics, which she thinks is fundamentally false but nonetheless serves under a wide range of conditions as a powerful and reliable instrument for mediating her engagement with a variety of mechanical phenomena (the motion of rockets, moons, cannonballs, tides, etc.) to which she takes herself to have some independent route(s) of epistemic access. For the realist herself, the instrumental utility of Newtonian mechanics *consists in* the truth of many of its implications concerning rockets, moons, cannonballs, tides, etc. And for those who are prepared to adopt the very same attitude towards a much wider range of successful fundamental scientific theories, there will still remain an extremely wide range of beliefs (including most of what we sometimes call our evolving and scientifically educated common sense about the world) for which we simply do not have the same rationale for adopting instrumentalism that we have in the case of many fundamental scientific theories. The former beliefs (e.g. if I drop this cannonball out of this window, it will fall and hit the ground after X number of seconds…) remain available to serve as those with respect to which we think the latter (e.g. the gravitational attraction between two massive bodies creates a force…) function merely as powerful instruments for prediction and intervention, for they are either not supported by eliminative inferences at all, or the eliminative support we have for them is of a sort for which we have no historical evidence that the prospect of radically distinct well-confirmed unconceived alternatives is anything more than a speculative possibility. It is these former beliefs that I suggested an instrumentalist could treat as "strictly and literally true", though Psillos is surely right that this is a poor choice of words: what I meant is simply that we take them to be true in the same straightforward sense that the scientific realist thinks the claims made by Newton's radically false theory about the behavior of the rockets, moons, cannonballs, and tides of our everyday experience are more-or-less just plain true. But this is emphatically not to say that there is some special *domain* of objects, events, or phenomena concerning which our beliefs must all be strictly and literally true: after all, a cannonball is also a matter field. So when we adopt an instrumentalist stance towards a given theory we do not believe claims about objects, events, and phenomena *when* they can be understood independently of that theory, but instead *as* they can be understood independently of that theory and any others towards which we adopt an instrumentalist stance.

Does all this involve, as Psillos suggests, a "double standard" in confirmation? If so, it is a double standard I think we can and should embrace. Are atoms and amoebae really on epistemological equal footing? Although we can point to a glowing blue dot in a suitably prepared photograph and say "see, that's an atom," virtually all of what we think we know about atoms comes from the role they play in a highly elaborate fundamental theory we have adopted because its empirical accomplishments are so much more impressive than those of any competing account we know of concerning the fine structure of matter. But quite a lot of what we know about amoebae (how fast they move, what they eat, how often they reproduce, etc.) does not come to us in this way, but in a variety of other ways by means of which we routinely gather knowledge about the world around us (even if this knowledge is also ultimately "theoretical" in character). If there is a double standard here it simply recommends, as all good epistemological double standards do, that we treat beliefs differently when there are important differences in the kinds of evidence we have in support of them.

Of course this picture does allow, just as Psillos suggests, that we are accumulating more and more knowledge about the world all the time. The iterative progress of scientific inquiry on which Saatsi rightly insists allows us to rule out more and more candidate theoretical accounts of the various domains of nature, and in the process develop increasingly powerful conceptual tools for mediating our engagement with nature. Along the way we steadily add to that part of our knowledge that does not rest on suspect eliminative foundations. Thus, the question isn't whether science is great, whether its iterative methods are a marvel, or even whether we make progress – it is, they are, and we do – but whether we are epistemically entitled to believe the descriptions offered by our best scientific theories (or even some systematically and reliably identifiable part or aspect of those descriptions) of otherwise inaccessible domains of nature – and we're not. As the judo masters know, the song says it true: there's no success like failure, and failure's no success at all.

Department of Logic and Philosophy of Science
University of California, Irvine
Irvine, CA
USA

References

Chakravarrty, A. "What You Don't Know Can't Hurt You: Realism and the Unconceived", *Philosophical Studies* 137 (2008), pp. 149–158.

Friedman, M. *The Dynamics of Reason* (Stanford, CA: CSLI Publications, 1999).

Hacking, I. *Historical Ontology* (Cambridge, MA: Cambridge University Press, 2002).

Hacking, I. *Scientific Reason* (Taipei: National Taiwan University Press, 2009).

Laudan, L. *Science and Relativism. Some Key Controversies in the Philosophy of Science* (Chicago: University of Chicago Press, 1990).

Laudan, L. "The Epistemology of Light: Some Methodological Issues in the Subtle Fluids Debate", in *Science and Hypothesis*, Chapter 8 (New York: Taylor and Francis, 1981), pp. 111–140.

Magnus, P.D. "What's New About the New Induction?", *Synthese* 148 (2006), pp. 295–301.

Psillos, S. *Scientific Realism: How Science Tracks Truth* (London: Routledge, 1999).

Saatsi, J. "Reconsidering the Fresnel-Maxwell Case Study", *Studies in History and Philosophy of Science* 36 (2005), pp. 509–538.

Saatsi, J. and P. Vickers. "Miraculous Success? Inconsistency and Untruth in Kirchhoff's Diffraction Theory", forthcoming.

Stanford, P.K. "Scientific Realism, the Atomic Theory, and the Catch-All Hypothesis: Can We Test Fundamental Theories Against All Serious Alternatives?", *British Journal for the Philosophy of Science* 60 (2009), pp. 253–269.

Wimsatt, W.C. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (London: Harvard University Press, 2007).