Psillos, S. (1999) *Scientific Realism: How Science Tracks the Truth* (London: Routledge).
Psillos, S. (2002) 'Simply the Best: a Case for Abduction', in A. Kakas and F. Sadri (eds), *Computational Logic: From Logic Programming Into the Future* (Berlin: Springer-Verlag), 605–625.
Reiner, R. and R. Pierson (1995) 'Hacking's Experimental Realism: an Untenable Middle Ground', *Philosophy of Science*, 62: 60–69.
Rescher, N. (1980) *Induction: an Essay on the Justification of Inductive Reasoning* (Oxford: Blackwell).
Resnik, D. B. (1994) 'Hacking's Experimental Realism', *Canadian Journal of Philosophy*, 24: 395–412.
Salmon, W. C. (1984) *Scientific Explanation and the Causal Structure of the World* (Princeton: Princeton University Press).
Sober, E. (1991) *Reconstructing the Past: Parsimony, Evidence, and Inference* (Cambridge, MA: MIT Press).

2

Optimism about the Pessimistic Induction

Sherrilyn Roush[1,2]

1. Why worry?

How confident does the history of science allow us to be about our current well-tested scientific theories, and why? The scientific realist thinks we are well within our rights to believe our best-tested theories, or some aspects of them, are approximately true.[3] Ambitious arguments have been made to this effect, such as that over historical time our scientific theories are converging to the truth, that the retention of concepts and claims is evidence for this, and that there can be no other serious explanation of the success of science than that its theories are approximately true. There is appeal in each of these ideas, but making such strong claims has tended to be hazardous, leaving us open to charges that many typical episodes in the history of science just do not fit the model (see, for example, Laudan, 1981). Arguing for a realist attitude via general claims – properties ascribed to sets of theories, trends we see in progressions of theories, and claimed links between general properties like success and truth that apply or fail to apply to any theory regardless of its content – is like arguing for or via a theory of science, which brings with it the obligation to defend *that* theory. I think a realist attitude towards particular scientific theories for which we have evidence can be maintained rationally without such a theory, even in the face of the pessimistic induction over the history of science.

The starting point at which questions arise as to what we have a right to believe about our theories is one where we have theories and evidence for them, and we are involved in the activity of apportioning our belief in each particular theory or hypothesis in accord with the strength of the particular evidence.[4] The devil's advocate sees our innocence and tries his best to sow seeds of doubt. If our starting point is as I say, though, the innocent believer in particular theories does not have to play the offense and propose sweeping views about science in general, but only to respond to the skeptic's challenges; the burden of initial argument is on the skeptic. The greatest strength of the realist attitude lies at this starting point, as Arthur Fine (1996) realized, and I will argue here that no one has given reason from the history of science to give it up. In particular, no pessimistic induction over the history of science has done what it needs to do in order to undermine our right to apportion our beliefs in particular theories to the evidence we have for them. To show this I am going to reconstruct what the pessimistic inductivist could possibly be saying when he tries to undermine our confidence, and show that there is one potentially threatening argument form he could use. However, this argument form requires much more than has been expected if it is to serve the pessimist, and pessimists have not actually ever offered what is needed.

For an example of how to start from the beginning, consider the realist of the sort I described, whom I will also call the 'optimist', who is typically confident that Quantum Theory has some very big things right. (As I will explain below, she does not have to know exactly which ones.) Her confidence in the claims of Quantum Theory is based on trust in the work of the scientific community, her understanding of the good evidence we have for this theory, and her own thought and knowledge about the physical world and logical and mathematical matters. There may be circumstances in which it would be helpful to her to have a further argument at her disposal that explains why all of this speaks to her having a right to her beliefs in Quantum Theory's claims. But the having of a right to a belief does not require her to have an argument available for defending just any challenge to the claim that she has a right to this belief. If it did then the justificational regress that follows would undermine our right to any of our beliefs.[5]

If this is where we start then the anti-realist will have to provide reason to believe that our innocent does *not* have a right to such beliefs. One way to do that would be to cast doubt on the optimist's evidence for Quantum Theory, meaning arguing about particular experiments and so on, and if those doubts were compelling then she would be obligated to respond to them. Of course, this has not been the strategy of the anti-realist, whose only other option, then, is to make a general argument for doubt that will apply to Quantum Theory *qua* theory, or *qua* theory of a particular sort. That is, the pessimist will have to find a fault or weakness that can be described without reference to the contents of specific theories. If he is to do this by an induction over history, then he must draw our realist into

reflection about all (many, most, typical, etc.) theories and scientists of the past, and show why those reflections are both relevant and damning for us. One cannot do a legitimate induction from properties of swans to properties of bananas.

The pessimist's tools have to be at a general level and apply to past theories, but the confidence of the optimist that the pessimist needs to undermine is in our particular theories, e.g. Quantum Theory, so the pessimist needs to display connections both between general and particular, and between them and us. If instead of or in addition to history he is going to invoke general concepts, such as underdetermination of theory by evidence, then the anti-realist will have to explain why these things have an upshot for belief in Quantum Theory. I don't say these things cannot be done – sometimes a theory exhibits a simple instance of an epistemologically relevant concept or phenomenon, and surely there are some similarities between our predecessors and ourselves. I am making the point that this is what needs to be done in order to provide a challenge to optimism that we need to take seriously. In the pessimistic induction over the history of science the relevance of the failures of our predecessors is supposed to come by way of likeness. We have some property or properties in common with them that are relevant to whether one achieves true beliefs, and since they failed at the project of getting true theories, we have insufficient reason for confidence that we are succeeding.

I will argue that for the pessimist's historical evidence base of failure to be relevantly similar to us, his argument must be formulated as a meta-induction, an induction over our beliefs in scientific theories and evidence on the one hand, and the properties of those beliefs – truth, falsity, or fallibility, for example – where these take the position that swans and white respectively do in the generalization to 'All swans are white.' There is one way that similarity at the level of these meta-facts – they had beliefs . . ., we have beliefs . . . – could be the basis of the induction the pessimist desires, I will argue. However, simply because of the way inductions work, two more elements are needed: to know which facts about their beliefs are relevant to claims about our beliefs and why – the contents of the beliefs are different after all – and to know why even a meta-claim about ourselves, a claim about our having certain beliefs and dispositions to believe, should affect our confidence in our object-level claims, say of Quantum Theory. Although I will explain why the pessimist is on good grounds in the second claim, I will also show that finding a similarity that could support a possible induction from our predecessors to ourselves in the first place, requires recourse to properties so general that the induction is easily undermined by more particular and relevant differences.

Could the similarity between our predecessors and ourselves be at the level of evidence? I will look for a similarity at the first-order level – the content of evidence claims – to show that there is no way to avoid putting the induction at the meta-level. The content of our set of evidence claims is different from that of our predecessors, but the pessimist might hope to avoid that problem; our evidence *set* is different from theirs, but they overlap over some propositions. However, even particular evidence claims that we may share with our predecessors can be relevant to whether our theories are true only if their content is relevant to the content of our theories in the right way. Typically it is not.

The content of the evidence for or against most theories in history is not relevantly similar to the content of our particular evidence or theories. Obviously, whatever made the theory of bodily humors seem likely is not relevant to Quantum Theory in the right way. But the lack of appropriate relevance persists in more apparently plausible cases too. For example, whatever it was about the facts and the phlogiston theory that made scientists give up that theory is not evidence against our chemical theory or the parts of that picture that we might have retained. Our theory did not retain parts of that theory that we (and the pessimist) thought that evidence had falsified, so the evidence that falsified that theory does not falsify our chemical theory. The relevant evidence of our predecessors was supposed to be the evidence they had *for* their theories, one will say. But this fares even worse. Anything we did retain from their theory had not been falsified by any of our predecessors' evidence (or ours), so any positive evidence they had for their view is also positive evidence for that part of ours. The rest of the new theory we have about the subject matter would not be inconsistent with that evidence either, since we knew about that evidence when we formulated our theory. If the evidence for their theory told against our theory we would not be espousing our theory, since this is evidence that we knew about in choosing our own theories.[6]

These points depend on a comically idealized picture of the way transitions are made between theories over time. However, that fact does not help the pessimist. Consider the phenomenon of Kuhn loss, where a later theory fails to explain a phenomenon the earlier theory did. In such a case either our theory is consistent with that phenomenon (though shedding no light on it) or it makes predictions inconsistent with the phenomenon. If it is consistent, then that evidence shows not that our theory is false, but at most that it is incomplete. If our theory is inconsistent with that evidence, then there exists evidence against our theory. But the existence of evidence against Quantum Mechanics does not provide material for a pessimistic *induction*. There would be nothing essentially historical about the problem – at most history was used to discover a problem for a particular theory of ours, that there exists evidence contrary to it.

Or consider violations of what Feyerabend call the Consistency Condition, which says later theories must be consistent with the observational evidence for earlier theories and which I assumed above. As with Kuhn loss, such sets of evidence will not yield a pessimistic induction that casts doubt on Quantum Theory, since to find such a case is simply to find evidence contrary to our theory, not to find a phenomenon casting doubt on our theories that is essentially historical.

These points are sound when they concern particular theories and the contents of particular evidence, one might say, but what if Kuhn loss and violations of the Consistency Condition are recurrent phenomena in the history of science? In that case, the induction that may be possible is a *meta*-induction, since it concerns similarity of our predecessors to us in the believing of theories for which they and we had (have) good evidence, and in the existence of evidence against their theories that they were, and therefore we probably are, unaware of or stubborn about. A theory can be mostly or approximately true, however, while not being consistent with all evidence – indeed it can be exactly true despite some contrary evidence, provided only that we got an auxiliary a little bit wrong. And we can be legitimately highly confident in a theory while also being highly confident that something is wrong somewhere without knowing where, and thus legitimately confident that there is some contrary evidence somewhere too. I will explain these points when I come to the preface paradox below. If the pessimist could show that the phenomena of Kuhn loss and violations of consistency are large and rampant throughout the history of science, then he would be on his way to the meta-induction described. However, the facts about method that I argue below undermine the last possible pessimistic induction would also undermine that one.

These points and common sense tell us that the relevant similarity between us and our predecessors are the general facts that our predecessors *had* evidence that *supported* their theories. Since their evidence was inadequate to their presumed goal of having true theories, the similarly supportive evidence we have is also likely inadequate to the same goal for our theories. The similarity for a potential induction must be between how their evidence and beliefs related to their theories' truth values and how our evidence and beliefs relate to our own theories' truth values. Their beliefs were somehow justified by their evidence and false, and our beliefs are also justified by our evidence and despite this also probably false.

The problem with this approach is that in escaping the differences in particular content of theories and evidence from the past to our own by going to the general comparison, the pessimist will be making an induction that falls victim to those same particular differences in a different way, because of the way induction works. It is an elementary and familiar fact about induction, that even if all F so far are G, the inference to the claim that the remaining cases with F will be G is rendered illegitimate if the remaining cases are known to have a further property which we know is relevant to whether G arises. Reichenbach called the move where we take into account that further evidence 'cross induction' (Reichenbach, 1949: 430). Thus, the epistemological mistake of the man who jumps off the top of a building and at the 40th floor says 'So far, so good!' is not in making the induction in the first place, nor in being wrong in his conclusion, but in failing to use his knowledge of terrestrial gravity to undermine the induction with a cross induction.

There is straightforward material for a cross induction against the latest argument I have described on behalf of the pessimist. If the pessimist cites for the base property of inductive comparison the fact that our predecessors, like us, had (good) evidence for their beliefs, then two facts combine to counter the inference from this similarity and the falsity of their theories to the conclusion that ours are probably false too. One is that the content of our evidence propositions and evidence sets is different from the content of theirs; at the very least there is *more* of it. We have done more and different experiments, for example. Secondly, the content and quantity of evidence is relevant to whether the belief one bases on it is likely to be true. This is so, at least, unless no amount of evidence short of infallibility is sufficient to justify any degree of belief in a theory, and unless more or different evidence does not make a difference to whether your belief based on it is likely to be true. But the problem those possibilities pose is the general problem of induction, and if pointing to that was the pessimist's argument he need not have reported it anew or enlisted history to make it. Moreover, it is an argument that would undermine his own induction. Since our evidence set is relevantly different from our predecessors', it remains to be shown why their failure is a reason to believe in ours. The optimist is in a strong position here, for she has no burden to argue that our evidence is *better* than that of our predecessors. It is the pessimist's burden to say why it is not.

The first meta-induction I have reconstructed for the pessimist is undermined by a simple cross induction. However, there is another meta-induction available. In this argument we would take the base property again to be believing and having (good) evidence, but the inferred property will be not being wrong but being unreliable. If our predecessors' theories were usually false, as the pessimist supposes, then our predecessors were *unreliable*. If we, like them, believe on the basis of what looks to be good evidence, there would seem to be excellent reason to think we are unreliable too. The unreliability of belief in theories is different from theories being false because it is different from a belief being false. A belief can be unreliably formed and yet happen to be true. Thus, because the property of unreliability is distinct from that of falsehood, this holds out hope that there is a good pessimistic argument after all. What would follow about our rational confidence in particular theories if the pessimist succeeded in this way in giving reason to believe we are unreliable is again a separate question.

In what follows I will address the argument just described, and the proper relationship between beliefs in our reliability level and beliefs in our theories. For this purpose we need a more precise understanding of the properties of reliability and fallibility, and the proper relation of beliefs about whether we have these properties and rational confidence in our theories' claims about the world.

2. Fallibility

There is no disagreement today that our science is fallible, that is, that we *could* be wrong, that our evidence is not strong enough to imply the truth of our theories. The issue is not whether we are fallible, but whether given that we are we can nevertheless have a right to confidence that our theories are true. The pessimist just imagined tries to show us we don't by showing that we are not just fallible but *unreliable*. But one might think he does not need to go that far, for is there not just something wrong with admitting you are fallible on the one hand, and persisting in high confidence in your theories on the other?

If this is our reason for pessimism, and if it works, then we do not need an induction over the history of science. However, showing what is wrong with this idea in its various forms will yield an understanding of fallibility that will in turn allow us finally to evaluate the pessimist's induction as most recently described.

To see what looks wrong about acknowledging fallibility and remaining confident, suppose that admitting our fallibility means being confident that at least one of our claims or theories, $p_1, p_2, p_3, \ldots, p_{10,000}$, is wrong:

$$* \text{ not-}p_1 \text{ or not-}p_2 \text{ or not-}p_3 \text{ or} \ldots \text{or not-}p_{10,000}$$

How then can we simultaneously be confident in each of those claims

$$** \; p_1, p_2, p_3, \ldots, p_{10,000}$$

when their conjunction

$$*** \; p_1 \text{ and } p_2 \text{ and } p_3 \text{ and} \ldots \text{and } p_{10,000}$$

directly contradicts the disjunction * above? On the other hand, one can imagine what looks like a quite sensible way of having such an array of beliefs by imagining a person who confidently offers the claims of her book, yet also in the preface admits that a few of them are probably wrong. The difficulty here has come to be called the Preface Paradox. Philip Kitcher (2001a, 2001b) noted some years ago the resemblance of this paradox to imagining our beliefs in scientific theories as justified despite our acknowledgement of possible error, and sided with intuition. Of course the person writing the preface is saying something sensible, whatever precisely we eventually figure out it must be. She is not contradicting herself and we are not either when we do not take our fallibility to be a reason to give up our confidence in our theories.

This is right in my view, but it would also be good to know precisely what she is saying, so that we can put to final rest the sense of suspicion about our confidence in scientific theories that this paradox nurtures. Over the years a variety of proposals have been offered for resolving the Preface Paradox, but they carry high price tags. Some have involved denying that conjoining conjuncts preserves justification (denying closure of justification under conjunction). Others propose that the lesson of the Preface Paradox is that it is possible to rationally believe two contradictory claims – the conjunction and the disjunction – as long as we avoid conjoining them and thereby believing a contradiction. Fortunately I do not think we need to go this far, for if we take it that rational degrees of belief are required to satisfy the probability axioms, thus,

$$S \text{ rationally believes } q \text{ to degree } x \text{ iff, } P_S(q) = x,$$

where P_S is a probability function, then we will be able to see that there is no paradox at all. The facts I am about to discuss are familiar to probabilists, possibly so familiar that it explains why I have not been able to find this solution written down. But it is worth putting them down in a way that highlights their salience to our questions here.

I take it that the attitudes of the preface writer that we need to explain involve being highly confident in each of the claims of her book, and also highly confident that at least one of them is wrong. But it is simple to show that if one's degrees of belief conform to the probability axioms and they are not extreme – zero or one – then having a preface-writer set of attitudes is not just permissible, but obligatory. Suppose I am highly confident in each of my book's claims:

$$** \; p_1, p_2, p_3, \ldots, p_{10,000} \quad P_S(p_n) = .95, \text{ for each } n$$

Then I am rationally required also to have a very, very high degree of belief that at least one of them is false, i.e.:

$$* \text{ not-}p_1 \text{ or not-}p_2 \text{ or not-}p_3 \text{ or} \ldots \text{or not-}p_{10,000} \quad P_S(*) = .99 \text{ repeating}$$

This is because the probability of a conjunction, here

$$*** \; p_1 \text{ and } p_2 \text{ and } p_3 \text{ and} \ldots \text{and } p_{10,000}$$

is the *product* of the probabilities of the conjuncts.[7] Since the confidence I have in each conjunct is not perfect but .95, the required confidence in the conjunction is very low indeed. I would only need to have 59 such conjuncts in order for the required confidence in the conjunction to drop to 5 percent. 95 percent raised all the way to the 10,000th power, as in this case, makes the probability of the conjunction practically nil. It immediately follows that the probability of the negation of this conjunction, that is, of the disjunction of the negations of the conjuncts, must be very, very high. This is just how the degrees of belief of our preface writer are arranged, with her very confident that at least one of her claims is false.

One might object that on this picture the rational credence in the conjunction of claims of the book is still low, whereas the preface writer should surely be confident of that conjunction. But as long as we assume that her degrees of belief are not extreme, the probability axioms require her to have very low confidence in the conjunction despite her very high confidence in each conjunct, so this objection amounts to insisting that the preface writer should be probabilistically incoherent. The strongest of the author's reasons to believe each of the p's and their connections will anyway tend to be local, and to ask for her confidence in the conjunction we would have to ask whether she thinks *all* of the claims true. Given the obvious contradiction between this and the claim that at least one of the claims of the book is false, surely she would demur.

What if the preface writer's degrees of belief concerning these matters are extreme, zero or one? If her confidence in each of the conjuncts is full then probabilistic coherence requires her confidence in the conjunction also to be perfect ($1^n = 1$), maximal, and thereby also requires her to have zero confidence in the disjunction of negations of the conjuncts.[8] This may look like a problem, since for such a person to be probabilistically coherent, she must not have the slightest degree of belief that one of her beliefs is false; she must not acknowledge her fallibility, so she must not write that preface if she is to remain rational. But while this is true, it is not paradoxical, for the person described in this case is not plausibly someone who *would* write that preface in the first place. The person we are imagining has perfect confidence in each and every claim of her book. This means she does not leave open any possibility of being wrong, the formal expression of which is the fact that the maximal degree of belief (one) makes impossible any revision of that belief on the basis of change in other beliefs. The paradox of the preface does not arise merely from the fact that $p_1, p_2, p_3, \ldots, p_{10,000}$ are together incompatible with not-p_1 or not-p_2 or not-p_3 or … or not-$p_{10,000}$. There also must be reason to suppose that those propositions (or other appropriately contradictory relatives of them) are ones that the writer of the imagined, apparently reasonable, preface is confident of in an incoherent way.

One might object that how confident one is allowed to be in the disjunction of negations depends, on this picture, on the number of conjuncts asserted, and must be small if there are few conjuncts. The confidence of our preface writer, on the other hand, that she has made at least one error is surely high. The allowed confidence in the disjunction does indeed depend on the number of conjuncts on this picture. However, the preface writer's level of confidence that she has made at least one error also intuitively depends on how many claims she has made. If you had confidently made only three claims, how confident would you be that at least one of them is wrong? Surely not a great deal more than your hesitancy in each of them, in conformity with what the probability axioms require. In such a case it would be strange for the preface writer to express high confidence of error. Conformably, I have never seen an abstract or first paragraph of an academic paper announce the high likelihood that something in the paper was erroneous. The reason that the Preface Paradox is most pressing in the case of a book isn't just that books have prefaces, but also that books are long.

This straightforward resolution of the Preface Paradox depends on avoiding a threshold conception of justified belief. That is, we do not ask whether or not a person is justified in (fully) believing p depending on how high the probability of p was taken to be, or at what level of evidence quality a subject is justified in having full belief. Rather, we ask whether or not a person is justified in having a certain degree of belief in p, and this depends on whether that degree is probabilistically coherent with the rest of her degrees of belief in evidence – and other propositions. Thus, a person is justified in having *full* belief in p if and only if her evidence and other beliefs require her belief in p to be one (see note 5). The reasonableness of this solution is seen in how well it incorporates the phenomena of fallibility, which was what the threshold conception was trying to do in taking a non-maximal amount of evidence to justify a maximal degree of belief.

Just as it is rational to write the preface while remaining confident in each of the claims of your book, it is rational to acknowledge that some of the claims of your theories are probably wrong – you know not which – and yet to remain confident in each of those claims. A conception of rationality as requiring probabilistic coherence wouldn't have it any other way. In particular, we are well within our rights to be both confident in Quantum Theory and also confident that something of it probably mistaken, without needing to know where the problem is. Thus, confidence in our theories can coexist with confidence in our having made some mistake. This is another expression of the fact that the pessimist cannot win at the first order but must deal with claims we make *about* our beliefs and theories.

There is a richer way of representing the claim of fallibility than the confidence in the disjunction of negations of my beliefs, and it will give our pessimistic inducer another chance. It is of course not impossible for a maximally confident person to write the self-doubting preface; only, as we said, implausible. If we express fallibility in a more adequate way than we have so far, we will see that it is not even first-order probabilistically incoherent for the maximally confident person to write that preface. There is a broader irrationality involved in it though, that corresponds to our expectation that a person would not feel inclined to do that. The formulation of fallibility above implicitly took my expression of my fallibility about the topic of my book as itself in direct competition with my confidence in each of the propositions in question, equating my recognition of my fallibility with a lack of perfect confidence that all of those propositions were true. This is one way to understand the problem, and it generates the most direct version of the Preface Paradox that we just dealt with. However, what I am asserting when I say with high (not maximal) confidence:

$$\text{not-}p_1 \text{ or not-}p_2 \text{ or not-}p_3 \text{ or} \ldots \text{or not-}p_{10{,}000}$$

is not giving me or my audience any more information over my lack of perfect confidence that:

$$p_1 \text{ and } p_2 \text{ and } p_3 \text{ and} \ldots \text{and } p_{10{,}000}$$

(except the fact that with regard to these degrees of belief I am coherent and responsive to the relation between the contents of these propositions). My high, or at least non-zero, confidence in the disjunction of the negations of the conjuncts is not a statement of my fallibility. It is simply an expression of my less than perfect confidence about particular matters.[9]

When someone discovers that he is sometimes wrong about such things, what he discovers is unfortunate, his discovering it is fortunate, but neither is just the same thing as his lacking confidence in the particular propositions. Rather, it says he has a general property, a tendency among his beliefs; it is a statement *about* his beliefs, a statement which would include among its terms not only 'p' but also 'B(p)', meaning the subject believes p; as such, a statement of my own fallibility is second-order, expressing a belief about my beliefs. We feel that the maximally confident person should do something in response to this discovery about his beliefs, but it is not impossible as a matter of fact for him to fail to. It would not even make him incoherent at the first-order if he failed to respond with a reduction in his confidence in p (not without further assumptions). This way of thinking about fallibility thus avoids the first-order Preface Paradox too. However, intuitively, believing ourselves fallible surely places some kind of rational demands on us in the management of our first-order beliefs.

What exactly those demands of rationality are is a general, non-trivial, issue, but this is a question the pessimistic inducer needs an answer to. For even if he succeeds in convincing us that the unreliability of our predecessors is a reason to believe in our own unreliability, we now see that we must ask *why* this should obligate us to reduce our confidence in Quantum Theory. Here I will do the pessimist a favor by filling in this blank in his argument by defining fallibility as a second-order property, and explaining why and how it places rationality constraints on first-order beliefs.

If I am asserting an explicit claim of my fallibility in the preface, I am not withdrawing confidence in my beliefs but making an observation about the probability that creatures like me, with dispositions like this, get it right about matters like this. The claim that q and the claim that I or those like me have a general tendency to imperfect beliefs on q-like matters, are distinct kinds of assertion.

It seems clear to me that the second-order property that is fallibility is an inverse of reliability, which is also a second-order property. We can say that a person's belief-forming process is *x percent reliable* (with respect to q) when:

$$PR(q/B(q)) = x$$[10]

That is, the probability that q is true, given that the subject has full belief in q is x, where for simplicity we are taking 'B(p)' to mean perfect confidence.[11] For the moment we may think of this probability as a rate of success, representing how *often* q is true when the subject believes it; an interpretation suitable in this context must in any case be objective, hence 'PR' instead of 'P'. Obviously, the probability that a person is right about q when she has perfect confidence in it is not necessarily 1; confidence does not imply reliability.

From this definition fallibility can then be expressed as the size of the gap between her reliability and perfect reliability. Perfect reliability means the rate, $PR(q/B(q))$, equals 1. Always, when she is sure of it, it is true. Thus, she is fallible (in her full beliefs) to degree y when:

$$1 - PR(q/B(q)) = y$$

To acknowledge one's fallibility is to acknowledge that $y > 0$, and that $PR(q/B(q)) = 1 - y$. That is, it requires a degree of belief in a statement about the objective conditional probability $PR(q/B(q))$. However, without further assumptions one's belief about this conditional probability does not constrain one's belief about q in any way.

When we say we might be wrong somewhere, we could mean this second-order claim. It does not immediately conflict with our first-order beliefs, which is good news for avoiding a new second-order Preface Paradox. The bad news that does suggest a new Preface Paradox is that it would be broadly irrational not to respond to the discovery of fallibility by doing something about our confidence in q. We will resolve this below in a way that reassures us in both of these directions. To make the connection between the two orders we would need to make an assumption about how a degree of belief about a property of our degree of belief in q should relate to degrees of belief in q. Since these two beliefs are distinct states of affairs at different orders, probabilistic coherence alone – conforming one's degrees of belief at a given order to the probability calculus – does not by itself give a constraint concerning that connection. For those in the know, note that this is explicitly acknowledged in the fact that $P(q/P(q) = x)$ is taken to be an independent constraint worthy of discussion and disagreement, and requiring argument.

3. Non sequitur

We now see how to express admission of fallibility in a way that is *consistent* with first-order assertion of particular claims, but is it not clear that the admission still undermines our justification for the first-order belief by an obvious and immediate inference? Wouldn't our serious fallibility as scientists, if shown, automatically give reason for an equally serious reduction of our confidence in Quantum Theory?

We should not conclude this hastily. Consider one consequence if we did think that the admission of fallibility immediately undermines our justifications for first-order beliefs, even domain-specifically. Most of us will have been frustrated at least once by a typical Creationist (or Intelligent-Designist) line of rhetoric one hears, which points out that the theory of evolution has not been *proved*. Therefore, it is concluded, opposing views (especially one in particular) have a right to be taught in public schools. We should have an 'open mind', and not be so confident in the views of the scientific mainstream that we present them as true. It certainly looks like the Creationist is appealing to the fallibility of science in this argument, and taking it to require no more than 50 percent confidence in evolution.

One obvious reply to this is that he is mistaking a small quantity of fallibility for a problem so large that it puts all views and frameworks on an equal footing of plausibility. 'Proof' is for mathematics and logic, not for empirical research where fallibility is just the way it is. The possibility of error does not justify an explosion to the night in which all cows are black. Our theories are fantastically better studied and verified than the opposing theories vying for a platform in public schools. But how can we justify this claim about the irrationality of the move to equal plausibility? And if the level of fallibility is significant, then what? The resemblance

of the form of the pessimistic induction over the history of science to this Creationist rhetoric is somewhat alarming. In charity we must suppose there is a difference, but it would be nice to know what it is. We will see below that when we understand what has potential in the pessimist's argument – by understanding the rational place of fallibility claims – we will also see what is fallacious about the Creationist argument.

The pessimist has potentially more to offer than the Creationist, if he can convince us that our predecessors were often or always wrong, meaning they had a low, or very low, reliability, and also that we are similar. But how should we count? What if there are more theories in the last 100 years than in the previous 5,000, and whereas all of the previous ones were false (or at least abandoned) a high percentage of those in the last 100 years are still not falsified (and are retained)? All of these theorists are our predecessors and the overall numbers wouldn't yet be shown to be bad on this quite plausible way of counting. It is not clear how anyone will decide non-arbitrarily how to count the historical cases, but this is a problem for the pessimist now since we have put the ball in his court again by describing one plausible way of counting that makes the fallibility of our predecessors low or unknown.

In charity to the pessimist I will suppose we can find a reason to think the supposedly large impairment of our predecessors is shared by us, in order to reconstruct his best possible argument. The current problem is why even if we successfully attribute a surprising, significant, general level of unreliability to ourselves, this should issue in a withdrawal of our confidence in specific first-order matters, such as particular claims of the Special Theory of Relativity. Suppose that q = the speed of light is not different in different frames of reference, where frame of reference is defined by velocity. The claims that are probabilistically relevant to that claim are the results of the Michelson-Morley experiment, the claim that the interferometer arms were equal in length, and so on. The first is relevant because if the experimental apparatus works as it should then the fact that no interference fringes show up in the experiment raises the probability that the speed of light is constant over reference frames. The second is relevant because length equality in the interferometer arms increases the probability that the apparatus works as it should, i.e., that fringes you may see are due to variation in the speed of light. Length equality is relevant to the hypothesis because it raises the probability of that hypothesis given the experimental outcome. Our being fallible is definitely not relevant to the Special Theory of Relativity in the way that typical evidence like this is.[12]

As pointed out above, it is a straightforward fact that the general level of unreliability we might attribute to ourselves,

$$PR(q/B(q)) = x,$$

is not probabilistically relevant to

$$q$$

and thus not to how much confidence we should have in q, without further assumptions. PR(q/B(q)) *is* related to PR(B(q)) via:

$$PR(q/B(q)) = PR(q \,.\, B(q))/PR(B(q))$$

But the question is what confidence we should have in q, given our credence for PR(q/B(q)) = x, not what confidence we should have in B(q) in that circumstance. How confident should I be in q if I believe that the probability that q is true given that I believe it is x?

4. Descent

I will call this the problem of 'descent' because we have to say how and why a second-order discovery about our tendency to believe should affect our confidence in specific first-order claims. Surely a discovery that, say, we fall into a class of people who have a 20 percent error rate on theoretical claims doesn't mean we need to drop our confidence in the relativity of simultaneity to 50–50, meaning that all bets are off. But it is reasonable to think it means *something*.

A notion that would give the pessimist's intentions their due is that of calibration: we are calibrated when we match our confidence to our accuracy (reliability). We can see an example of the concept of calibration in the way that an eyewitness's confidence on the witness stand relates, or should relate, to his other properties. We onlookers may have no information about the person's track record in making face recognition judgements, but there is a track record and it appears to be epistemically good, in some sense, if the witness is no more confident in his judgement now than the percentage of times he's gotten such judgements right in the past. His track record is a clue to how reliable he is. If he's been right 99 percent of the time it would be epistemically good if he were very highly confident. (Otherwise in trusting him only to his level of confidence we are deprived of the information he likely has.) At the other end of the spectrum, if he were right 50 percent of the time in past such

cases, we hope that he would say 'I don't know.' He may have no inkling of his bad track record – he may have acquired amnesia since his last such judgement – and if we don't know about the amnesia or the track record we are out of luck. It is clear what would be most epistemically sound and helpful in these situations, from imagining it present and imagining it absent.

There is psychological evidence that jurors use confidence of a witness as an indication of accuracy by default, effectively assuming calibration without any information pertinent to the property. That is, on first encounter with a witness, a juror tends to believe what a confident witness says in direct proportion to the witness's confidence. Since confidence implies nothing without calibration, this appears to be an epistemic disaster for the system of trial by jury, and was greeted as such on its discovery. This illustrates both that calibration is a non-trivial property, i.e. that confidence and accuracy are distinct properties that do not necessarily line up in naturally occurring beliefs, and that it matters to us; it would be good for us if a witness's testimony were calibrated before we form any particular degree of confidence in what he says, even if, as a matter of fact, we tend to be poor judges of calibration. Incidentally, and in keeping with the idea that we do demand calibration, it has recently been discovered that (mock) jurors do make use of any further evidence of error by that witness that comes available, to reduce their judgement of his calibration and thereby of his credibility when he is confident (Tenney et al., 2007). The jurors' default assumption of calibration is defeasible. We not only care about this property of calibration but implicitly reason with it continuously in dealing with new evidence and evidence about those who deliver it.

So, suppose calibration – having your confidence match your accuracy – is a good thing, and that by 'accuracy' we mean reliability, probability of being right when you believe, the property defined above that we can get information about by looking at track record. This would imply that our confidence in our particular theories should match our reliability in making such judgements, and that if we aspire to that then our confidence should match what we *believe* our reliability is.[13] We have hereby uncovered the basis of an intuition that makes us feel a pull from the pessimistic induction: if the pessimist can show us that the (supposed) low reliability of our predecessors is good evidence that *we* have a low reliability, then we should reduce our confidence in our particular theories to the level of reliability we have now come to believe we have. If we believe our reliability level as scientists is 50 percent, then we should be no more confident than not that our theories are true. Thus, we can see how the pessimist can finish his argument on us, if he manages to convince us we are unreliable. But this is the only good news for the pessimist. In what follows I will discuss how the historical evidence the pessimist gives us *could* be a reason to believe we are unreliable, but argue that it is not.

Note that this way of viewing the relevance of the second-order property of fallibility to what our first-order confidence in particular theories should be does not license the Creationist move. This is because the rule I have described does not imply that fallibility implies equal plausibility of all theories. The fallibility will, speaking in idealization, be a given number, the reliability being 1 minus that number. Suppose, for illustration, that the fallibility level of contemporary science is 20 percent; then the reliability level is 80 percent. The rule says match your confidence to your reliability, so in this case our confidence should be 80 percent. Discoveries about your reliability level in q-like matters may obligate you to push your confidence in q up or down by a small or large amount, but they do not automatically mandate, or even license, a leveling in which every view is as good as every other. This view also shows why the amount of fallibility matters – our allowable confidence varies with it in a way that conforms to the fact that .02 fallibility is quite a bit better than .3 and almost as good as 0 (perfect). In so doing it also resolves the second-order Preface Paradox, by showing that you can rationally both believe you are fallible and maintain confidence in your particular theory, in the right proportions, of course.

5. What have they to do with us?

Suppose our predecessors had low reliability, meaning that they had a pretty high rate of believing things that were false. If our reliability level is the same as theirs then, because we should be calibrated, we should have low confidence in our theories. The question is whether our reliability level is the same or similar, and whether theirs being low is a *reason* to think ours is so. If I have been right above, then this is the best way of seeing the issue at stake in the pessimistic induction. The relevant question is not whether their theories were false and whether that should matter to us – as we saw above the pessimist can't get where he wants to go that way. He will instead argue that because they justifiedly believed and were unreliable our justified beliefs are unreliable too. Note that the more particular fact that we have *different* evidence for our theories than our predecessors had for theirs, which undermined the pessimist's induction above, will not do the job of saving our optimist here. The content of evidence claims cannot be depended on to be relevant to whether our beliefs are reliable since we could easily churn out different evidence in each round of research, but maintain the same miserable level of unreliability, for example because we use the same sorry ways of gathering evidence. However, there are particular second-order facts that will undermine the new induction, which involve differences in the use of methods over time.

Let us suppose that we are licensed in concluding that our predecessors are unreliable by looking at their track record. If we are to block the inference to the conclusion that we are too, it would seem that we need to know something relevant to our track record. And one might wonder how we could investigate our own reliability level without already knowing whether our theories are true. To see our reliability by rate of true theories we would need to do a count among the theories we believe, and see how many are true. But theories that we know are false we no longer believe, and asserting the truth of any or all of the ones we believe would seem to be begging the question here. If there exists any discernible track record for us, must it not be in the broad sense of 'us', the set of human beings doing science, in which case, our having granted the premise that our predecessors were unreliable would mean the pessimist wins after all.

What this misses is that the thing we need an estimate of is not track record *per se* but reliability. Tallying a track record is one way of estimating reliability but it is not the only way. For example, we can reasonably judge a machine to be reliable at doing its job through knowledge of its mechanism, or we can justifiably judge it as more reliable than another machine via our knowledge of the two mechanisms and of how good such mechanisms could possibly be at delivering the result we want. Another kind of simple example to make the point: we know the valences of hydrogen and carbon. Say we know the valence needed at the active site of any enzyme that is going to catalyze a chemical reaction we want to know how to produce. Suppose carbon has that valence and hydrogen does not. A given molecule with a carbon sidechain may not work, but we can see that its chance of doing so is fantastically higher than that of a molecule with only hydrogen sidechains.

We see an analogous thing in comparing our science to that of our predecessors, if we think of methods as mechanisms for leading us to believe the truth and avoid falsehood. Over historical time, our reliability is potentially and probably different from that of our predecessors because we use different methods. Here I use the term 'method' broadly, so that it includes, for example, techniques that do not take the form of rules, and also includes techniques that are specific to a given subject matter. The greater the historical time between us and them, the greater the probable difference in methods between us and our predecessors. If so, then this will be the material for a cross induction of the latest pessimistic induction if I can show that methods, in particular those that have actually developed over historical time, make a difference to reliability.

It is emphatically not necessary for the optimist to show that our methods are better – if we show that our methods are different and that methods are relevant to reliability, then it will become the burden of the pessimist to show that our methods are not better, and that thereby we are subject to the induction from our predecessors' unreliability. However, we can, I think, argue for the general superiority of later methods to earlier methods in many cases, and this will be the easiest way to show the weaker claim we need, that method makes a difference to reliability.

There is something to the stereotype that the Ancients preferred speculation to observation, and that speculation combined with observation is better than speculation alone. There is something right about the idea that Bacon's interventional experimental method, and his rules for safeguarding against psychological prejudices, get us more information than mere casual observation, at least, if any method gets us anything at all. There is something to the idea that following the refinements of experimental methods like those described by John Stuart Mill and William Jevons gives us more accurate and safer beliefs. The twentieth century brought a massive number of discoveries and refinements in statistical methods, and the new century has brought further ones already. Prima facie it looks pretty good for a claim that in reliability we are not as bad off as our predecessors. A more detailed look bears this out.

Consider some examples of methods, e.g. simple induction and cross induction. We may be able to show that induction brings us truths in the long run if any method (e.g. counterinduction) does (Reichenbach, 1949; Salmon, 1967). We may be able to show that the straight rule of induction is faster than any other method of ampliative inference (Juhl, 1994). But it does not seem that we can show flatly that induction is likely, or even *more* likely than other methods, to bring us true beliefs. It may be that none are getting onto the world. However, as we have noted, if the pessimist is going to hang his case on the fact that every scientific enterprise faces the problem of induction generally, he could have left off his induction over the history of science. We can show that problem through one toy case. Moreover, the pessimist usually does want to admit that science gets us some more restricted kind of knowledge than that of theories, since otherwise he has a hard time explaining the worth of what scientists are doing.

Let us assume that the world is susceptible to inductive procedures. I will argue that some methods are more reliable than others – where reliability is as defined above, the probability that q is true given that one has come to believe it by that method. In particular, some of the methods we use are more reliable than any of the methods our predecessors used. To show that a method is more reliable than another method, it is sufficient to show that there are potential errors that the first method is more likely to catch than the second method is, and that there are not errors that the second method is more likely to catch than the first method is. If this is to be sufficient for defending ourselves against the foibles of our predecessors, we must assume that the application of the first method is just as competent as the application of the second. However, the pessimist hasn't shown that our application of method to the world is worse than that of our predecessors, so we are free to make this assumption.

First let us consider methods pairwise on a question about a connection between two properties, F and G. We could construe the question as whether the one causes the other, or simply whether the correlation will continue as we look at more cases with F. We will apply both methods to a finite data set in which every case with property F also has property G. Now consider induction by generalization. This simple rule says that if all the cases you have seen having property F have property G too, then infer the generalization All F-cases have G. This rule is subject to the simple kind of error of failing to make an available cross inductive move such as, for example, in the case of the flying man above, using his knowledge of gravity. It is hard to imagine a human being so uninformed about the world that he doesn't know independently of his jump that the sequence of correlations he is seeing must shortly come to an end. But to get himself to a warranted belief about whether or not he will go splat, it is not enough to know this; he must also use the information. The method of induction by generalization did not tell him to do that. Adding the method of cross induction would. And if you use any information you have that is the basis for another induction that cuts across the trend you see and shows that it cannot, or cannot be expected to continue, there are many cases in which you will more likely avoid a false belief than your counterpart who does not supplement induction by generalization with cross induction.

The error that cross induction guards against is precisely one that would reduce our reliability in the sense defined above, for it would be a case of the thing you believe (that your fall is not dangerous) being false. The point here generalizes since any induction is subject to potential cross inductions due to the non-monotonicity of ampliative evidential support; unlike a deductive inference, an inductive inference is erodable by addition of premises. It is true that any cross induction is also susceptible to undermining by further evidence, but that is a new error whose discovery will also require employment of the method of cross induction. It does not immediately appear that there can be potential errors of the method of induction plus cross induction that are caught by the method of induction alone, since the former method incorporates the latter, and the interaction between the two parts of the former method does not seem to create new error possibilities.

One might object that any cross induction adds one more piece of evidence to the premises, which will make the premises as a group have strictly more potential error than the inductive inference that evidence was supplementing.[14] And surely it cannot be proven in general that that added error potential is less than the error the cross induction allows us to avoid. This is true, but the comparison to be made is between a method in which the next use of evidence is always for making a straight induction or the next use of evidence could issue in the making of a cross induction instead. On both sides of the comparison you are adding a piece of evidence. The evidence added in the two types of case could possibly have different potential error themselves, but we assumed above that we are no worse at *applying* our methods than our predecessors were at applying theirs so we may plausibly assume a symmetry between what they and we would achieve in error potential of our evidence.[15] The pessimist could try to argue that we are worse at applying our methods, but that wasn't a position I have heard a defense of. Thus, induction with cross induction is more reliable than induction alone. Moreover, the idea that this is so only by our own lights, a favorite phrase of pessimists, has a hollow ring.

A contemporary Bayesian approach to method incorporates the cross-induction rule in the following way. An idealized agent's beliefs are represented as a total function from all the propositions of a language to numbers between zero and one that represent strength of belief, that is, as a total probability function. This means that this ideal agent has degrees of belief about all matters, and most importantly here, that all the beliefs that any agent actually has are expected to be taken into account in any application of any rule that falls out of the probability axioms. The probability calculus thus incorporates a general, rigorous version of the cross-induction rule, and it brings more. For example, abiding by it allows us to expose equivalence relationships between different formulations of the same and related information, thus helping to identify cross inductions we might otherwise have missed.

The quantitative aspect of the Bayesian approach also yields rules for weighing how far a cross induction undermines an induction. In the case of the flying man the undermining is complete. Not so in many other cases. In coming to a conclusion in such cases, we are better off if we can tell whether the first induction or the cross induction is *more* powerful, and this approach provides techniques for doing that. This is only to mention some specific features of Bayesian method that are relevant to the execution of cross inductions. Advantages of further features and methods would take several books to describe, but we don't need to consider those to make the point that some methods are more reliable than others, and in particular that some of our methods are better than some, and in some cases any, of those of our predecessors.

Another problem with simple induction noticed probably first by Francis Bacon, and dealt with increasingly rigorously and expansively by John Stuart Mill, Charles Sanders Peirce, Neyman and Pearson, Sir Ronald Fisher, and many others who have followed, is that in the investigation of how potent an observed association between F and G is, one must deal with the possibility that the effect G was brought about by something other than F that one was not aware of. This is now addressed through an enormous and growing list of techniques: methods for insuring a representative sample of F's, the setup of a control group that doesn't have F in order to see whether F was really what made the difference to G's presence, the randomization of the control group to attempt to insure that variables other than F are not causing any correlation seen. And the foregoing are primitive descriptions of old techniques for squeezing information out of experiment and observation, that have been enriched and supplemented in the last fifty years by entire fields of researchers, in departments of statistics for example, and even among philosophers and computer scientists in the case, for example, of causal net programming. To bring the relevance of method to reliability full circle, scientists have increasingly acquired more, and more sophisticated, methods for evaluating the reliability of their applications of method.

In many of these cases we do not need to appeal to track record to determine whether one method is more reliable than another at achieving a particular aspect of the epistemological goal. Even when the statistical community has not settled a matter – e.g. between Bayesians and classical statisticians over whether a randomization procedure provides a better control group than mere matching – it is not imagined to be helpful to try to settle the dispute about method by pointing to a track record of scientists getting the right theory when they use it. Thus we do not need to already know whether or not our scientists are finding true theories in order to evaluate the quality of the methods they are using compared to the methods of their predecessors. The bottom line for our argument here is that these sophisticated methods are used today and were not used by our predecessors in earlier centuries; method is relevant to reliability – as we have shown by showing that some later ones are more reliable than some or all earlier ones, and this is material for a cross induction against the inference from our predecessors' unreliability to ours.

6. More is all we need

More, and more specific, investigation along these lines of comparing methods over the history of science would be valuable for judging where we stand, and this, I suggest, is what the discussion of realism and anti-realism might constructively give way to, or at least give a place for. At a gross level we know that our predecessors were not using certain methods before a certain point because those methods did not exist at that time. But there are many questions about what they actually were doing in each case, and how it compares to what we do for those same topics or related ones. We need not be motivated by an effort to defeat the general pessimist about scientific theories – he is already defeated, or at least will be by the end of this chapter. We are free to be motivated by the intrinsic interest of the relation between reliability and method.

Even if better methods are available today, are they being used, and used well? And is our confidence actually tailored to their reliability? And even if better general methods are available today might they not, for substantive reasons, have no added reliability value for a given topic of investigation? In such a case the calibration imperative would counsel no more confidence in our conclusions than we think our predecessors had a right to. The task of determining the reliability or comparative reliability of our methods looks more difficult the more subject-specific the method under consideration is, since usually one would need already to know or make assumptions about the subject to know whether such a method is generally any good. (Although if we cannot make *any* such assumptions, we are back to the basic problem of induction.) Content-independent arguments like those given above about general method go the distance for these methods.

There are several general reasons why this isn't a problem. One is, as explained, that now that we have shown that method is relevant to reliability, the burden is on the pessimist, not us, to show in particular cases that our methods, of whatever sort, are no better than those of our predecessors, if he wants to cast doubt on our beliefs in particular theories. And it is confidence in particular theories – Quantum Mechanics – not some general claim about such confidence in all of our theories, that the optimist has the goal of maintaining. Another is that what matters here is the comparison to our predecessors, not the absolute level of our reliability, for the question is whether their reliability problems are problems for us. This requires argument, but it is one that can be made.

We can see how the argument works with methods that require substantive assumptions by considering four cases, a two-by-two matrix over sameness and difference of substantive assumptions and method. In the first case we share substantive assumptions with our predecessors and our overall method is the same too, partly in consequence. In this case if we think they were wrong, then we will think we're wrong too, because on assumption of equally competent application we have the same conclusions! But that means those results will not be the matters at issue that we are currently confident about. (We also do not actually have the same conclusions so are unlikely to be in this case.) In the second case our substantive assumptions are the same and our method is different (which must be due to other aspects of it). This is the case we dealt with above: method is in general relevant to reliability, so this is material for a cross induction and moves the burden of proof to the pessimist.

In the third case our substantive assumptions are different but the overall method they are part of is the same as that of our predecessors. Unlikely, but in any case if our overall method is the same while the substantive assumptions are different, there must be some non-substantive part of the method that is different, and the central argument above applies. Once again the ball is in the pessimist's court. If our substantive assumptions are different and our overall method is, partly thereby, different, then since difference in method undermines the pessimistic induction, the burden of proof is with the pessimist. The fact that methods often involve substantive assumptions particular to a domain or a question, does not pose a problem for our response to the pessimist.

The pessimist may reasonably protest that even if we show our methods to be different from our predecessors', many of our predecessors also had methods that were different from their predecessors'. A lot of good it did them since they ended up wrong a lot of the time. This won't be enough to revive our naysayer's argument, though. The fact that our predecessors often had different methods from their predecessors' is material for undermining the induction over their predecessors to *them*. Thus, our predecessors' predecessors' failures did not render our predecessors' confidence unjustified either. So too for us, by the only induction that is left to make. This may also be why we tend to think of those of our predecessors who used methods that are a subset of ours as being in some real sense justified even when we also think their conclusions were wrong.

One might spy another negative induction in the offing here: it is true that our predecessors had different methods than their predecessors but they didn't succeed any better? Does this not provide inductive evidence that method is *not* relevant to reliability, which was a crucial premise of our cross induction?[16] This line is also doomed, for we showed non-inductively that method *is* relevant to reliability, even if methods involve substantive assumptions. Thus, there is something we know, relevant to the property this argument is inducing to – namely, the irrelevance of method to reliability – that is (maximally) relevant to whether that property can be expected to be there in the yet unexamined cases. This induction too is undermined by a cross induction.

The argument of this chapter also stands against a new pessimistic induction developed recently by Kyle Stanford (2006; see also 2000a, 2000b, 2003). The inference he describes involves conceivable alternate theories for explaining our evidence. Stanford points out that there were alternative possibilities to our predecessors' theories that were conceivable, and that showed our predecessors' theories false, but which they did not conceive of. We know these possibilities were conceivable, because we have since conceived them. There is no reason to think we are different from them in this respect; there must also be conceivables we have not conceived. And, roughly, we can't show that they don't undermine our theories because we're not conceiving of them.

Stanford focuses not on the challenge posed by beliefs of a certain sort, those about unobservables, but on ways at arriving at beliefs in science, in particular whether we can eliminate alternative explanations of phenomena. Since method is directly relevant to this, the argument of this chapter is relevant too. To perform Stanford's new induction, the cases of our predecessors and ourselves must be relevantly similar in the basis property, which here is proposed to be the affliction by unconceived conceivables relevant to the theoretical question the scientists in each case are investigating. This property is second-order and affects our predecessors' reliability, because as we know their affliction was partly responsible for their believing false theories. In fact their ways of dealing with the problem left them unreliable, as we can see because their theories were false. Since we are afflicted by the same problem, we can expect that we are unreliable. However, granting that we also are subject to unconceived conceivables, the question is whether our predecessors' unreliability that was due to this property really says anything about us.[17] For it to do so there must not be any other properties to show that we may be different from those predecessors in our reliability when facing possibility spaces of alternative theories.

Unfortunately for the pessimist, there have been a lot of discoveries of new methods for dealing with large theoretical possibility spaces. In particular, before the early twentieth century possible alternative explanations for phenomena and experiments were ruled out seriatim as they were conceived of. The early twentieth century saw the blossoming of a first round of techniques for ruling out large classes of theories without *conceiving* of their members (Roush, 2005: 218–223), Since then, of course, there have arisen more, and more sophisticated techniques.

Once we point out that our methods for dealing with unconceived conceivables are different in a way relevant to whether our theories are likely to be true, it is the pessimist's burden to show that despite these differences in method we are no more reliable than our forebears. The question is not, as Stanford sometimes suggests (Stanford, 2006: 131, 133), whether we have a method good enough to rule out all possible alternative theories. Since the pessimist is the one doing an induction, the question is whether our predecessors' faults should make us think we are in the same boat, so that we should dial down confidence in our particular theories. He has not shown this.

Thinking carefully about the pessimistic induction over the history of scientific failures shows that it must be a meta-induction if it is to make past failures *relevant* to us. The Preface Paradox is no paradox at the level where the optimist resides, with first-order confidence in particular theories. The pessimist must appeal not merely to the falsity of our predecessors' theories but to the unreliability of their ways of coming to their beliefs (as confirmed by their repeated false conclusions). But even if we grant their unreliability, nothing follows from this about whether we have a right to our confidence in our particular theories unless it is shown (1) that their unreliability is a reason to think we are unreliable and (2) that our believing we are unreliable forces a revision of our first-order beliefs on us. I have argued via our desire for calibration that (2) is true. However, (1) has not been shown, since the manifest difference in methods between us and our predecessors undermines the legitimacy of the pessimist's induction. The pessimist seems to be in even worse shape than having the burden of proof unloaded on him, since it is unclear how to show that method is not relevant to reliability without appealing to the general problem of induction.

Notes

1. This chapter has benefited greatly from discussions with Fabrizio Cariani and members of the Philosophy Department of the University of Washington, especially Arthur Fine, Bill Talbott, Andrea Woody, and Alison Wylie.
2. This work has been supported by NSF grant 0823418.
3. Recent efforts of this sort can be found in Leplin (1997), who argues for a link between novel predictive success and truth, and Psillos (1999), who argues for a link between referential continuity and descriptive accuracy.
4. This stance is of course somewhat similar to Arthur Fine's 'natural ontological attitude' (Fine, 1996). However, besides my focusing more on epistemology rather than ontology, there is also the difference that here it is a starting point against which, I argue, there is a compelling form of argument for ascent from our object-level beliefs to reflection on them, and from there descent to withdrawal of confidence in particular theories. It's just that, as I show below, it does not appear possible to make such an argument successful.
5. The claim I intend here does not assume as much externalist epistemology as it may seem. I grant that the optimist needs to have an ability to give reasons for her beliefs in theories, but the believer I have described can do this by citing her evidence for them. My claim that she need not have an argument ready for just any challenge to her right to her beliefs in theories is consistent with an obligation to take seriously any prima facie good argument against that right if she is to retain her status as justified. My point is that whether there are good arguments against her right is something that needs investigation before she loses her justified status by not having responses to them. Canvassing the space of such arguments is what I do in this chapter.
6. I am making a crude 'one drop' assumption about falsification here. But the pessimist must also be assuming the past theories are false or his negative evidence from the history of science disappears, since the property he is inducing to is 'probably false'. We can use here whatever sense of 'falsification' he is using there, and make the same argument. The pessimist could instead be arguing that past theories are false in virtue of being inconsistent with our theories, but that is not a first-order option in which the evidence for past theories is pessimistically relevant to our theories, which is the form of pessimistic argument considered in this section. Any second-order argument in which the falsity of their theories is supposed to be a reason to reduce confidence in our theories would fall victim to the central argument of this chapter, which follows.
7. If the p's are not independent, then there are conformably fewer sources of potential error than the number of p's, and the case is like a conjunction with fewer conjuncts. The comparison of few and many conjuncts in the preface situation is treated below.
8. Thus, a maximal level of justification is closed under conjunction. What is not closed under conjunction is rational *degree* of justified belief. The failure to distinguish these two facts, and the intuitions that correspond to them, causes much confusion about these kinds of cases.
9. This is also one problem that afflicts Andy Egan and Adam Elga's (2005) argument that one can't coherently believe one is unreliable without withdrawing the beliefs in question. There fallibility of a given degree has simply been formulated as with the original Preface Paradox here, as the disjunction of the negations of the propositions I believe, which automatically cancels that same degree of confidence in the claims in question. The second-order formulation of fallibility I give and the solution I just described provide two independent resolutions of their problem.
10. This is not the only way to express reliability, and hence fallibility, though I think that any adequate way of expressing it as a property must be second-order, not first-order. I have switched to considering the reliability of full belief for ease of exposition. The same argument can be formulated for rate of success for a given level of partial belief in q.
11. One might be worried about the lower end of the spectrum of degrees between 0 and 1. What could it mean to say that someone was 20 percent reliable? It means, simply, that they get it right 20 percent of the time. They wouldn't be a very useful source (except in reverse, in a yes–no query), but that isn't the problem here.
12. One might notice that the second claim is the denial of a Type II defeater – a claim not that the purported evidence claims are false but that the purported support relation between the evidence and the hypothesis is weaker than thought. And one may wonder whether we don't need something higher than first-order probabilistic relevance to model that relevance. If we need second-order claims and these are intuitively relevant to the truth of Special Relativity, then we don't need any special explanation of why fallibility claims are relevant too. But first-order relevance accounts for (first-order) Type II defeaters handily. C is a Type II defeater of the claim that A supports B iff $Pr(B/A) > Pr(B/A.C)$. A Type II defeater *screens off* the relevance of A to B. The relevance of all the specific claims to the hypothesis that the speed of light is constant over reference frames can be represented at the level of first-order probabilistic coherence. The relevance of claims about fallibility cannot.
13. The obligation to *keep* yourself calibrated (as opposed to being in that state) can be formulated as a rule for updating our beliefs on discovery of information about our reliability level, so: $Pr_f(q) = Pr_i(q/[Pr_i(q) = x. PR(q/Pr_i(q) = x) = y]) = y$. That is, the final degree of belief you ought to have in q given that your current degree of belief in q is x and that you think the objective probability of q when you believe q to degree x is y, is y. So, for example, if you know that in the past when you responded to the weatherman saying there was a 30 percent chance of rain on a given day by having a 30 percent confidence in rain that day, and it actually rained 60 percent of the days, then you should update today's weatherman-induced 30 percent confidence in rain to 60 percent. This scheme is an extension of the Bayesian constraints on rationality, which relaxes certain idealizations in order to accommodate the fact that while it is unfortunate to discover you are less reliable than you expected, the only irrationality would be to fail to respond to that discovery. I am currently developing the details of this rationality constraint. Defense of the formal constraint is a very complicated matter, but my ground for endorsing calibration here is the empirical evidence that it is beneficial.
14. I am grateful to Arthur Fine for this objection.

15. There is a serious problem in this direction, first noticed by Hume, that comes from the potential error added in moves where we try to catch our errors. See Vickers (2000). I am able to escape it here because the situation allows me to make felicitously comparative judgements.
16. I owe this objection to Bill Talbott.
17. I say the property in question is reliability. We could start with an argument where our predecessors faced unconceived conceivables and ending up with *false* theories, but we already know that such an induction is undermined by our having different evidence sets from them.

References

Egan, A. and A. Elga (2005) 'I Can't Believe I'm Stupid', *Philosophical Perspectives*, 19(1): 77–93.

Fine, A. (1996) *The Shaky Game: Einstein, Realism, and the Quantum Theory*, 2nd edn (Chicago: University of Chicago Press).

Juhl, C. F. (1994) 'The Speed-Optimality of Reichenbach's Straight Rule of Induction', *British Journal for the Philosophy of Science*, 45: 857–863.

Kitcher, P. (2001a) *Science, Truth, and Democracy* (New York: Oxford University Press).

Kitcher, P. (2001b) 'Real Realism: the Galilean Strategy', *Philosophical Review*, 110: 151–197.

Laudan, L. (1981) 'A Confutation of Convergent Realism', *Philosophy of Science*, 48: 19–49.

Leplin, J. (1997) *A Novel Defense of Scientific Realism* (New York: Oxford University Press).

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

Reichenbach, H. (1949) *The Theory of Probability* (Berkeley: University of California Press).

Roush, S. (2005) *Tracking Truth: Knowledge, Evidence, and Science* (Oxford: Oxford University Press).

Salmon, W. C. (1967) *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press).

Stanford, P. K. (2000a) 'An Anti-Realist Explanation of the Success of Science', *Philosophy of Science*, 67: 266–284.

Stanford, P. K. (2000b) 'Refusing the Devil's Bargain: What Kind of Underdetermination Should We Take Seriously?', *Philosophy of Science*, 48 (Supplement): S1–S12.

Stanford, P. K. (2003) 'Pyrrhic Victories for Scientific Realism', *Journal of Philosophy*, 100: 553–572.

Stanford, P. K. (2006) *Exceeding our Grasp: Science, History, and the Problem of Unconceived Alternatives* (New York: Oxford University Press).

Tenney, E. R., R. J. MacCoun, B. A. Spellman, and R. Hastie (2007) 'Calibration Trumps Confidence as Basis for Witness Credibility', *Psychological Science*, 18: 46–50.

Vickers, J. (2000) 'I Believe it, but Soon I'll Not Believe it Anymore: Skepticism, Empiricism, and Reflection', *Synthese*, 124: 155–174.

First published 2010 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN 978–0–230–22263–2 hardback
ISBN 978–0–230–22264–9 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
19 18 17 16 15 14 13 12 11 10

Printed and bound in Great Britain by
CPI Antony Rowe, Chippenham and Eastbourne

Contents

New Waves in Philosophy
Series Standing Order ISBN 978–0–230–53797–2 (hardcover)
Series Standing Order ISBN 978–0–230–53798–9 (paperback)
(outside North America only)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBN quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

New Waves in Philosophy of Science

Edited by

P. D. Magnus
University of Albany, USA

and

Jacob Busch
University of St Andrews, UK

palgrave
macmillan