The Rationality of Science in Relation to its History[1]

Many philosophers have thought that Kuhn's claim that there have been paradigm shifts introduced a problem for the rationality of science, because it appears that in such a change nothing can count as a neutral arbiter; even what you observe depends on which theory you already subscribe to. The history of science challenges its rationality in a different way in the pessimistic induction, where failures of our predecessors to come up with true theories about unobservable entities is taken by many to threaten the rationality of confidence in our own theories. The first problem arises from a perception of too much discontinuity, the second from an unfortunate kind of continuity, in the track record of science. I argue that both problems are only apparent, and due to under-description of the history. The continuing appeal of the pessimistic induction in particular is encouraged by narrow focus on a notion of method that Kuhn was eager to resist.

Kuhn's *Structure of Scientific Revolutions* richly displayed the relevance of historical considerations to questions of what it is rational for scientists to believe and why. He did not, as some have maintained, think his interpretation threatened to make scientists' beliefs look irrational, but his work did make an approach to the issue of rationality via rules of Scientific Method seem cartoonish. For good or ill, though today there are more Bayesians than falsificationists and we tend to avoid that two-word proper name "Scientific Method" that now seems so naïve, philosophers have not stopped investigating methodological rules and principles of rationality. And many philosophers continue to take the relevance of the history of science to be as a pool of cases to be used to illustrate and test our views of the general rules of rationality, which involves nothing essentially historical.

I will not apologize for either proposals of general theories of the rationality of science, or the effort to keep them on topic by means of examples (Donovan et al. 1992). While historians since Kuhn tend to be suspicious of generalizations about all science, ironically Kuhn himself

proposed an inner logic as general as any methodologist's to explain the rationality of scientists' beliefs and behavior, in terms of the cycle: paradigm – normal science – anomaly – crisis – revolution. But Kuhn's generalizations arose *out of* attention to the specifics, the differences over time, the challenges, conceptual and practical, of getting a theory to say anything about the world, out of what David Hollinger has called the "quotidian". [2] Though Kuhn was a philosopher in my view, he had a historian's eye, which most philosophers lack. In contrast with Kuhn's direction, with philosophers today, including myself, the path of discovery of generalizations about the rationality of science often goes from general epistemological considerations—such as the worthiness of betting in line with the axioms of probability—down, by deduction, to slightly more specific principles intended to explain scientists' behavior, such as that surprising evidence has more confirming power. Such principles are tested against cases antecedently regarded as sound or unsound, but the fact that some of those cases are from the past is not, per se, significant.

However there is another prominent pattern among philosophers of taking the history of science as relevant to its rationality that appears to flow in Kuhn's direction from historical facts to generalizations. For example, in some arguments for scientific realism—the view that we have reason to believe successful theories (or their essential structures) are 'true-ish', rather than just good predictors of observables—putative facts about the history of science are taken as premises. The predictive success, retention, and convergence of opinion over historical time of some theories has been taken to be a reason to think those theories are true-ish. Another case of this pattern, the one that will concern me here, is found in the pessimistic induction (PI), which has nagged at the consciences of philosophers, laypersons, and even some scientists since the 1980s. The premise of this argument is also the putative track record of science, but here it is the failures: a high proportion of past scientists' theories were successful, but by our lights false in what they said about unobservable entities, as in the cases of phlogiston and the luminiferous ether, goes the argument. Induction, broadly construed, is the method of science, and that invites an induction over this sad record to the conclusion that it is not rational to have much confidence that our current successful theories accurately represent the world we cannot see. The history of science thus presents a challenge to the rationality of endorsing our scientists' hypotheses about unobservables at what one might call face value.

The historical claim merely that there have existed successful theories whose claims about unobservables were false is not generally disputed because it is weak, too weak to establish a worrying pessimistic conclusion. That there were a few such cases or a small proportion of the total cases would give us at most a conclusion that our theories have some chance of being be false, and we already knew that. To make the historical claim strong enough for trouble, it would have to be that success with observables is not even a mark of truth. This would require knowing

[2] David Hollinger has recently discussed this theme in a lecture honoring Kuhn at MIT in December 2012. He originally developed it in a paper (Hollinger 1973) that was much appreciated by Kuhn, who distributed reprints of it to thirty of his acquaintances.

the proportion of false theories in the pools of the successful and unsuccessful theories, but it has been argued that we cannot know these two base rates without begging the question. (Lewis 2001; Magnus and Callendar 2004).

However, as I will argue, even if the pessimist could defend appropriate claims about the base rates, his argument would not enjoy smooth sailing. Those rates go to secure the premises of the pessimist's argument, which are general facts about history, such as that many successful theories have been false, or that our predecessors have often failed to conceive of conceivable hypotheses. The last step of the pessimist's argument, the "action" one might say, begins with the generalizations, not the particular cases that they are generalizations about. Thus though the pessimist does proceed *from* claims about history, the specifics of the failures he points to—the sort of thing historians have eyes for—quickly become apparently irrelevant to the argument.

This focus is natural—the conclusion that we should be less confident in our theories is general, so it is presumed that the relevant premises will be general too—but I will argue here that this presumption is false. In an induction, including a PI, for the most normative, general, epistemological reasons there is no point at which the specifics of cases cease to be relevant to the argument. I will use this to argue that even if we could establish damning generalizations about our predecessors, it would not be sufficient for the pessimist's argument to touch the confidences in our own theories that we have arrived at by the usual means of the particular science. On the basis of general points about induction I will argue that novelty and discontinuity over time, particularly a novelty of method that tends to go unappreciated, saves our science from the PI. Ironically, the relevance of novelty is that it renders most failures of the past irrelevant to the rationality of embracing our own theories. The novelty that has the potential for this effect on the pessimist's argument is revealed at every more specific level of description of the past and present cases.

I will conclude that no form of the PI can succeed in giving us doubts over and above those that competent scientists already address in their day-to-day work on particular hypotheses. I will give a diagnosis of the hold the pessimistic argument continues to have over us, and explain this problem as the dual of a familiar rationality problem that Kuhn gave to us by introducing the notion of paradigm shifts.

2 The Objective

To defend our science against the PI is not to make any positive, general argument about what kind of truths science does get us, but only to show that the pessimist's negative conclusion is not supported by his evidence. Also, for the pessimist to succeed requires him to give us a reason to withdraw confidence in our particular hypotheses, such as those about the mechanism of

chemical mutation in *E. coli.* DNA, and the convection currents and composition of the Sun.[3] This is what is at stake in the argument. Scientists have evidence and arguments for such hypotheses, apportioning their confidences in these hypotheses to what they judge to be the strength of their evidence for them, and the pessimist's argument must show why those things the good scientist already does are not sufficient to justify their confidences in particular hypotheses if he is going to show anything troubling.

One way to do that would be to offer counterevidence about the composition or convection currents of the Sun, or to offer meta-arguments casting doubt on the design of the *E. coli.* experiments. We know that providing evidence against particular scientific hypotheses or experimental designs is not what the pessimist is doing; he is providing an argument based on generalizations. That the PI is reflective does not excuse scientists from addressing it, of course. Scientists reflect on their procedures and arguments every day, but how reflective they are obligated to be is limited by the quality of the meta-level objections. So, the pessimist has the burden to give an argument of good quality that is distinct from those scientists already address, if we are to have a special problem that derives from the historical record. This, I will claim, is what he cannot do.

3 Induction

I will take the pessimist at his word that what he is doing is an induction.[4] The term "induction" includes any ampliative inference, one in which the conclusion contains more information than the premises, in which it is logically possible for the premises to be true and the conclusion false, whatever more particular form that inference takes. The argument I will make is not restricted to next-case induction or generalization, but applies equally well to, for example, inference to causes. My argument, which rests on a claim that relevance demands similarity, does not hold generally for inference to the best explanation, or abduction, but this is not a form of argument the pessimist is using, or I think can use.[5] Given all of this, I will use the term "induction" for the forms of ampliative inference for which my assumptions hold, in the hopes that any kind of ampliative inference for which my claim about similarity and relevance does not hold is also, like abduction, one there is no way for the pessimistic inductivist to exploit.

---

[3] When I write "particular" hypotheses I do not mean singular propositions – the mechanism of chemical mutation in *E. coli.* is a phenomenon with more than one instance – but rather propositions investigated in particular sciences.
[4] Larry Laudan is usually credited with the first PI (Laudan 1981) but his confutation of realism was not an induction. His argument took the form of historical counterexamples to strong realist claims of a sort I am not defending, for example that empirical success is a mark of truth. The first adumbration of the PI argument seems to have been given by Poincare (1905), who did not however develop or endorse it.
[5] It can be used to defend anti-realism more broadly, of course, in an argument proposing to explain the success of science without appealing to the truth of theories. I only say it does not look helpful in a pessimistic induction over history.

The first point is that induction needs a similarity base, a similarity between the subjects of the premises and conclusion that makes the premises relevant to the conclusion. [6] Prima facie, we may infer that all swans are white from the fact that all we have seen are white because our evidence and conclusion both ascribe the projected property to swans. By contrast we would not even consider inferring from the fact that all swans we have seen are white that all paper towels are white. Even if all paper towels are white it is not a claim supported by evidence concerning swans.

Secondly, even if there is a similarity base, as with the old swan example, an induction is not justified if there is a known property P that is plausibly relevant to the conclusion property, and P is not uniform between data- and target- populations. In the swan case there is such a property because the habitats of the swans you have seen may easily be different from the habitats of some swans you are projecting to, and there is often color variation within species in different habitats. Such a property P provides what Hans Reichenbach called a "cross-induction". [7] For an induction to be justified, the similarity base must not be undermined by available evidence of a further property of the subjects (here distinct habitats) that is plausibly relevant to the presence or absence of the projected property (here white color). [8]

Often, cross-inductions operate under the surface. A smart-aleck could point out that swans and paper towels do have a similarity base; they both occur in the United States, for example. What is wrong with an induction based on that similarity is that there are further properties of swans and paper towels that are plausibly relevant to color, such as that one is an animal and the other a cleaning item. Another way to think of the situations where cross-inductions are appropriate, and one that will be helpful here, is that under-description of the evidence and the target has concealed the irrelevance of the evidence to the conclusion. Sometimes, we know and apply a fuller description automatically and unconsciously, as with the paper towels; sometimes we discover the further properties for cross-induction in the course of time, as with Europeans discovering black swans in Australia; and sometimes we already know of the further properties but have failed to take them into account. The latter is the case with the PI.

4 Similarity between past and present science

The pessimist needs a similarity base between past and present science to make our predecessors' failures relevant to what we have a right to believe about the world. The pessimist

---

[6] This point goes through for inference to causes. When we infer that X causes Y, we do it on the basis of information about things that have property X and have or lack property Y. The evidence and conclusion are about things similar with respect to X.

[7] Cross-induction is an alternative way of describing the non-monotonicity, or erodability, of ampliative inference, namely that, in contrast to deduction, addition of evidence to the premises can undermine the legitimacy of the inference. Epistemologists call such an additional piece of evidence a "defeater" or "underminer".

[8] One might want to strengthen the requirement for justification by weakening the qualifier "available". It seems that there are some examples in which justification can be undermined by further evidence even if one does not possess that evidence. (Harman 1980) That stronger requirement is not necessary for my argument, and indeed would weaken the force of my conclusion, since it would weaken the requirement for a cross-induction.

would give us a challenge if he could convince us that we have less right to confidence than we think we do at the object level, at the level of beliefs about particular unobservable matters such as the mechanism of replication of the MERS-CoV virus, and this is the kind of implication the pessimist advertises. Can he find his similarity base at this object level?

Similarity of this sort would be similarity in the content of theories or evidence. The PI is often intended to go way back and across subject matters, even to the theory of crystalline spheres and the theory of bodily humors. But the theory of bodily humors is not similar to the theory of quantum mechanics, not even in its subject matter, much less in its particular claims about its subject. And vast differences in the content of theories—what they claim about the world—is relevant to whether the theories are true.

Sometimes, though, the subject matter of past science is the same as that of ours. Newton's mechanics was declared universal in its scope, so it is not just the part of Newton's theory that we still think is approximately true that is the same subject matter as relativity and quantum mechanics. We do have a different particular theory from that strictly false Newtonian one so we might think that spoils the pessimism. But even in revolutions, physicists do not mutilate where it is not necessary, so there are similarities too, for example, the ones that structural realists argue are retained over the history of modern physics. There are also similarities in the evidence we have and they had for the similar contents because we retain that, too. So what if the similarity base is this: that part of the *content* of scientists' claims about the world that is similar?

This will not work, because if we think through similarity of content in theories and evidence carefully, we will see that the pessimist has a dilemma. Consider, first, the cases where our predecessors' theories were similar to ours in content. Those theories were either true or false. If their theories were false and we retained the false parts in our theories, then our theories are false too, but that is not an induction over history. If their theories were true and ours are similar in the respects that are true, then that is not grounds for pessimism. Similarity of the kind considered does bring relevance, but it does not support a PI.

Second, consider the cases where our predecessors' evidence was similar to ours in content. Any evidence is either supportive, counter-evidence, or irrelevant to a given theory of ours. If it is irrelevant, then it does not matter to how confident we should be in our theories. If our predecessors' evidence is supportive of our theories then that is good for our confidence. If it is counter-evidence, then it is reason to think our theories are false. But then our theories are seen as false because of particular counter-evidence to them, perhaps discovered by doing history, and not because of an induction over the history of science.

Thus, this object-level content strategy does not succeed. The argument reduces to something that either is not an induction or is not pessimistic. If these points are obvious, that is good for my argument. I have excavated them in order to show that the pessimist has no options at the object level. It is not just that the PI is often called and pursued as a meta-induction but that

because there is no appropriate similarity at the object level, the argument cannot both be successful and avoid that ascent to the meta-level.

The similarity at the basis of the apparently powerful argument must be a more general one between our predecessors and ourselves as investigators. We are doing the same thing that they did in some important sense. So how can we expect a different result? In particular, we are all doing science, and justified relative to the evidence that we have. Suddenly the theories of crystalline spheres and bodily humors seem relevant again. Our predecessors were unreliable in getting true theories. We are like them, so we are likely also often wrong. That justifiedness that they had and we have must of course be similar for the induction to proceed and that, in my view, is the weak link in the argument, which I will come back to.

But first, note that inducing over this property of justifiedness and to the property of unreliability—being often wrong—makes it a meta-induction in a precise sense. We are at the second order, meaning that the properties in question are properties of the scientist's beliefs, not of the world it is her primary aim to form beliefs about. In performing the pessimist's argument on ourselves, we are managing our beliefs about our beliefs. This has the important implication that the pessimist's argument must have two parts. For what he gets out of the induction over history, if he succeeds, is that we are likely often wrong in our theories, that is, our beliefs are unreliable. Recall that the position I am defending is not a general one about contemporary theories—that the successful ones are not too often wrong—but only the scientist's right to go on apportioning confidence in particular claims about unobservables, say about the interior of the Sun, to the strength of the usual kind of evidence she has for them, in the way a good scientist regularly does.

Unreliability is a property of *beliefs*, not of the convection currents in the interior of the Sun, and this presents an obstacle to the PI that has not been appreciated. Why should learning about our beliefs have an effect on what we think about the interior of the Sun? Facts about our beliefs are just not about the Sun, so how are they relevant? Put differently, what we believe about the convection currents in the Sun does not make a difference to what the interior of the Sun is doing, which corresponds to the fact that its apparent correlational relevance is screened off from claims about the Sun by other claims about the Sun. In gaining relevance of the past to the present by ascending to the meta-level, the pessimist has put in question the past's relevance to the confidences in particular theories that I am defending.

Any PI needs a justification not only for the inference from our predecessors' unreliability to ours (the horizontal inference), but also for the inference from our unreliability—being often wrong—to withdrawal of confidence in particular claims (the vertical inference). Having ascended to the meta-level to find premises that would not be irrelevant because of their difference in content, he must now descend if he is to deliver conclusions about our different content. Why should we think that an assumed general unreliability shows up as a falsehood here about muons, or there about quarks? How does any such inference go, and how is it justified?

I think second order beliefs do impose obligations at the first order, but why, and how it goes, are non-trivial questions. Elsewhere I defend a general answer to these questions, and so let pessimism live for another round. (Roush 2009, and ms.) The answer is that it is good to be calibrated, that is, for your confidence in proposition q to not only be appropriate to your evidence for q—say that it will rain tomorrow—but also to match your reliability in q-type questions—whether it will rain on day *x*—where reliability is an objective general relation between your believing q-like things and their being true. The paradigm case of calibration questions concerns a weather forecaster, whose reliability can be evaluated by track record, so the PI is well-suited to take advantage of this notion.

What is relevant here is that the rule I have defined (Roush 2009) demands a proportionality that explains why if we have evidence that the history of scientific failures is sufficiently relevant to our general reliability about things like the existence of muons and quarks, then our scientist does have a problem with particular claims about muons and quarks. This is because if a high fraction of our predecessor's theories about q-matters have been wrong, say 80%, then the fraction of q-matters we are likely wrong about is 80%, and only 20% are still to be considered right. The calibration norm then says to dial down the confidence in any such particular claim to 20%. This explains nicely part of the intuitiveness of the pessimistic argument, and shows that the vertical inference is defensible.[9] *If* the pessimist gives us reason to believe we are unreliable in q-like matters, then we should dial down our confidence in q.

5 Cross-Induction on Method

The horizontal inference from the past to the present is where the pessimist's stumbling block lies, in the question of whether our predecessors' unreliability is sufficiently relevant to our work to give us an induction to our own unreliability. For this, that in virtue of which we are justified must be sufficiently similar to the way our predecessors were justified. Why should we think this, and why do people actually think this? It is true that we are all doing induction in the broad sense, but we can say a little more and it has another name. We all, after all, used the Scientific Method.[10] Even though we avoid the language of scientific method, the rationality-philosopher's search for the most general rules of inductive reasoning does implicitly keep a focus on one method, in aiming for the minimum number of principles from which we could derive all of the

---

[9] Unfortunately for the pessimist, I have found almost no philosophers who agree with me about calibration and the role of re-calibration in assimilating evidence about ourselves, and despite a lively discussion of higher-order evidence going on in philosophy, no one has offered an alternative general account of how we are to take higher-order evidence into account or what justifies that. I maintain that the pessimist needs an account here, but since I believe there is one I grant him what he needs.

[10] We tend to be reluctant to extend this compliment to science before the modern era, but I think a pessimist believes there is enough failure to get true theories in the modern era to allow the pessimistic argument to proceed. Note that it is not necessary that one associate the similarity of justifiedness the pessimist attributes to us and our predecessors with scientific method, in order for my cross-induction below to succeed against it.

various more particular rules we see as being followed. And unabashed reference to the Scientific Method goes on as ever among scientists and laypeople. This focus on general method is a strong force I think in the grip the PI has on philosophers and others.

At least note this: since method is how we get from sensory irritations to beliefs general enough to be the conclusions of scientists,[11] the PI is maximally powerful when past and present scientists all use exactly the same method. Thus, we could blunt the PI by denying that there is any shared method at all. This is not an option for those of us who think there is, in the sense of basic forms of inductive inference and the demand for probabilistic coherence, and even some more sophisticated principles of evidence management. However, even granting the existence of shared method, or rules, or generalizations about sanctioned belief behavior, it is a mistake to think that that is enough to make the pessimistic induction work here.

This is because cross-inductions are available. A cross-induction does not require that the two populations be different in *every* way, but only in some way plausibly relevant to the projected property. And there are many more specific things to say about methods that are relevant to the effectiveness of our belief-forming practices at giving us true theories, that is, to our reliability. Those specific things are different for different contexts and questions, and even for the same subject matter, methods are different *between our predecessors and ourselves*. There is a lot more method than general philosophers of science tend to think about. Any procedure, tool, experimental design, protocol, instrument, is a method, because it has generality. It is repeated, held the same over cases of probing the world, and something has to be because we need sufficient sample sizes of evidence produced and evaluated in the same way to do legitimate positive inductions about particular matters. And though there is a great deal of retention of method at the general level, even there there are always new statistical methods, procedures, distinctions, and tools that are added to the methods we retain from past science. Statistics is itself a science, and it expands over time.[12] Thus, we will escape the PI on those particular occasions when cross-inductions on method are available, and they very often are when we compare ourselves to much of the history of science. Under-description of scientific method conceals irrelevance of the premise (here data about past scientists' justifiedness and failures) to the conclusion (here the claim that we are unreliable).

The specific, concrete differences of such belief-forming methods that you will find a good scientist counting as rendering previous failures irrelevant are differences that her evidence and background knowledge say are plausibly positively relevant to reliability. For a particular case, suppose your predecessor using chlorine in an experiment failed to get the expected answer,

---

[11] One might think the similarity between our predecessors and ourselves is that we all have inferred truth from empirical success. However, that is a method because it is a rule that we use to justify a move from results to conclusions. It is also an exceedingly general method, and more particular things there always are to say about methods can cross this method too in the way discussed below. For example, there are ever more and different ways of evaluating empirical success.

[12] See Glymour 2004 for some examples of this.

failed, in particular, to detect neutrinos at all the energy levels expected. Suppose you want to do the same experiment using gallium. A PI could say that since that same experiment failed enough times before, you are not justified in conducting it again. In this actual case, it was the same experiment to a significant degree, so there is a similarity base. But obviously, if a good scientist is proposing to do that experiment again with gallium that will be because he has reason to believe that the material—chlorine vs. gallium—could well make a difference to the results, good enough reason to make it possible to secure large amounts of funding for the quite elaborate operation. The failure using chlorine gives you good reason to believe the experiment will fail with gallium only if you do not have reason to believe the difference in material could make a positive difference to whether you detect neutrinos that might be there.

The pessimist is right this far: if you do exactly the same experiment a thousand times then you should not expect a different result the 1,001st time. Ten times would probably be sufficient. But though I am no expert in history, I am confident that the very same experiment over and over, with different personnel and freshly laundered lab coats but no changes plausibly relevant to reliability, is not what the history of science looks like.

Of course, everything is different from everything else. For every single experiment there will be some respect in which it is different from every other. Why does this not imply the ridiculous conclusion that all past failures of science are irrelevant to whether our work is reliable, and are legitimately ignored? Part of the answer is that the difference that crosses the induction to our likely failure has to be something we have reason to believe is relevant to the property the pessimist is projecting and the scientist is crossing, here unreliability. There is a fact of the matter, and often evidence about, whether a difference is relevant to reliability on question q, and many are not. For example, often the experimenter wears a different shirt when he runs the same experiment on a different day, but typically we do not think that will affect the results. If so, then yesterday's failure is not irrelevant just because of the different shirt. The fact that you use your method today or tomorrow usually is not relevant, although it will be if you are studying astronomical events like eclipses. You may have good reason to believe that whether you do it in Chicago or New York will not, per se, be relevant to the outcome, in which case you cannot ignore a failure the same experiment had in the other location just because it was a different location. Generally, which earrings I wear will not matter, but if they are made of heavy metal they might interfere with a magnet, so taking them off could make a difference, and if they are big and bright enough they might be distracting in a cognitive experiment with babies. Some properties are relevant to reliability on question q, and some are not. There is a fact of the matter that depends on the case, and that scientists make arguments about.

Whether the PI over the history of science has any doubt to contribute with respect to a particular hypothesis comes down to the question whether the method used to investigate that hypothesis is different from methods used in all past failures in investigating hypotheses, in a way sufficiently relevant to reliability on the scientist's current subject. As we have just seen, this is a type of question the good scientist addresses explicitly as a matter of course in investigating the

hypothesis. The question whether using gallium or chlorine is likely to make a difference to the results will be discussed thoroughly in any grant application for the type of solar neutrino experiment mentioned. So, the scientist addresses the PI over the history of science in doing the science itself. It follows that if the pessimist is going to give the scientist reason to doubt her hypothesis, he is going to have to argue not about theories and reliability in general, but about gallium, and whether its differences from chlorine are relevant to the energies of the neutrinos that can be detected with the given apparatus, and that will be a discussion with the scientist.

The pessimist might object that our scientist does not consider all of the cases included in the PI merely by a discussion of gallium. However recall that the question is whether this scientist's method is relevantly similar to those used by the scientists of the past who failed. His method is similar to those of Priestley with phlogiston and Lamarck with spontaneous generation and inheritance of acquired characters in roughly the way that paper towels are similar to swans. The relevant differences are so obvious that scientific journals economize on space by not requiring discussion of such comparisons. The scientist need not have explicitly considered them in order to take such cases properly into account.

One might suspect that I am begging the question, appealing to science to justify science. Surely the scientist is only justified in ignoring the general pessimistic induction if she is *justified* in thinking that the difference between chlorine and gallium, and that between the neutrino detector and a microscope, and between the neutrino detector and a bell jar, etc., are relevant to the reliability of detecting neutrinos. What right have we or they to think that? Moreover, why is my demand that the pessimist engage the scientist at the object level not gratuitously ignoring the issue?

I do not need to know whether the difference between chlorine and gallium and that between a neutrino detector and a microscope are relevant to reliability in detecting neutrinos, nor do the scientists' arguments about these matters need to be successful in order for my deflection of the PI to succeed. My point has been that these are the kinds of questions the pessimist's argument depends on and to which he needs negative answers if he is to succeed. They are questions about which particular differences of method are relevant to reliability of particular conclusions. They are particular because the scientist needs only one relevant difference of method in order to have a cross-induction against the relevance of a past failure. Every particular feature of a method in a particular case of our science is thus a potential threat to the pessimistic induction. All such features taken together exhaust the potential doubt a PI over the past could muster. It happens that every day scientists address, or are ready to address, particular claims about what methodical differences are relevant to the reliability of their particular results. Since every such feature is a threat to the pessimist's argument, his argument can only succeed if he takes the fight to the scientists themselves, and argues, for example, that the neutrino detector is not different from a microscope in a way that is plausibly relevant to reliability at detecting neutrinos. Perhaps the pessimist would succeed, but it would not be via an induction over history.

My claim is not that PIs *never* work, and therefore not that *all* past failures of scientists to get true theories are irrelevant to our confidences in our theories. Millions of PIs are good, and we do most of them implicitly, often without blinking. Some PIs—like the one from the whole history of science to a particular current hypothesis—are bad. To be a good scientist requires addressing, or being ready to address, questions of which particular similarities and differences to previous efforts are relevant to the reliability of one's particular results, and in doing so one addresses all of the doubts the history of scientific failures has to offer.

6 Rationality and History

Thus the specific differences and discontinuities of method over historical time—the things that historians are especially interested in—positively support, and are indeed essential to, the rationality of scientists' beliefs. It is useful to compare this situation with the difficulties that came for the rationality of science when Kuhn said there were global discontinuities called paradigm shifts that affect virtually everything about the way that a science operates—assumptions about the basic building blocks of the world, what are meaningful questions and sensible ways of going about answering them, and so on. The main, and familiar, problem about rationality is that if everything changes at once then there is nothing unchanged through a paradigm shift that could be a neutral arbiter between the before- and after- theories, to tell us why the change is rational or justified. The arbiter used to be observations, but these are theory-laden; what you see depends on which theory you already subscribe to. I will call this the *neutral-arbiter problem*. Obviously what I have said does not address this problem, but what I have argued has similarities to, and differences from, what Peter Galison said to this problem.

Galison points out that it is not true that all of science changes as a block (Galison 1997, 701-844). Go to a higher resolution and you will see that there are more levels than theory and observation. There is not only experiment testing theory, but material culture and computational methods among many other things, and the different cultures have quasi-independent inner logics driving them. So, layers typically change at different times according to their own needs and objectives, which are not always that of testing high theories. The continuity at one level can give you a vantage point from which to judge the wisdom of changes in another. And it does not have to be the observation level that is always the unchanged neutral party. Thus the intercalation of these layers is part of the epistemic strength of science.

Perhaps there is a non-trivial level of description at which everything changes at once, but for many purposes it is an under-description, and the under-description matters to rationality because it hides neutral arbiters. The move I made above has in common with Galison that you find your way out of skepticism by taking more specific facts into account, and explaining why they matter to rationality. However, Galison's argument rescues the epistemic strength of science by finding continuity over time despite the temporal discontinuities. In the rope metaphor that he uses, the existence of quasi-independent strands is crucial to a rope's strength because when one strand is

strained, the others are not breaking. I am addressing a different problem, which is a dual of the neutral arbiter problem because in the PI it is the continuity and similarity of method over time that appeared to create a problem. And the rationality of typical science is assured against this problem, I think, not despite differences over time but because of them. Earlier I pointed out that identical method in every instance of science would produce the most powerful possible pessimistic induction. The dual of that here is that a radical paradigm shift in which everything, even the more specific layers, really did change at once would be the ultimate weapon against the PI. It would prevent any induction from previous science to our own, whether negative or positive, because if everything were different there would be no similarity base at all. Of course, if we claimed a paradigm shift in that radical sense then an answer to the PI would come at the expense of an answer to the neutral arbiter problem.

However, just as Galison did not need to claim that all of the layers of science remain the same over time in order to address the neutral arbiter problem, so, too, no claim of difference all the way down was necessary for my defense of the typical work-a-day scientist given above, since crossing a PI does not require that the method of investigating a hypothesis be different in *every* way from the method that tried and failed before. It only needs to be different in some way that we have reason to believe is relevant to reliability. There is always a pool of similar episodes that are relevant to whether a scientist should trust the results of what she is doing now, and she must—and I say typically does—take their failures into account. However, that pool gets smaller the more fully what she is doing now is described. Consistently with Galison's view, both similarity and difference, continuity and discontinuity, are actual and necessary to the rationality of scientists' beliefs.

7 Application

I will illustrate how this argument goes for a recent version of the PI, before going on to reply to objections. In Kyle Stanford's PI, the similarity base is not explicitly method, but the induction can be crossed by means of facts about method. His argument uses as a similarity base the fact that we and our predecessors are subject to unconceived but conceivable alternative hypotheses about unobservables that are equally compatible with, and explain, our evidence (Stanford 2006). We know that our predecessors were subject to this because we have since conceived relevant alternatives they did not. There is no reason to think we are different in this, so we can expect our successors to conceive of alternative explanations for our evidence that we have not. Our predecessors were often wrong about unobservables[13] and this was connected to their failure to

---

[13] Note the dependence of this premise on a high base rate for our predecessors' failures. This is necessary to come to a conclusion that we are *likely* to be wrong, and, as mentioned earlier, has been argued to be impossible to assign. (Lewis 2001, Magnus and Callendar 2004)

conceive of conceivable possibilities. We can therefore also expect to be shown wrong in our hypotheses about unobservables because of our similarity to our predecessors.

I do think we can expect to be shown wrong, but the significance of that generic fact is questionable, as I will discuss below. The first flaw in Stanford's pessimistic argument is that it does not take into account changes in methods, specifically methods for ruling out alternative hypotheses about unobservables. The kinds of examples Stanford deals with are hypotheses about the mechanisms of heredity and they illustrate nicely how limited our intuitive imagination is. But we have statistical methods for ruling out alternative hypotheses that do not require intuitively imagining the mechanisms or objects that could be involved (Roush 2005, 218-221; 2010; Glymour 2004). Unconceived does not imply not ruled out. Thus, past scientists lacked many methods that we have for ruling out unconceived alternatives, and ruling these out was presumed by the pessimist to be relevant to reliability. Induction crossed.

Since we never have full evidence, we cannot suppose there is ever a stage at which we have ruled out all unconceived conceivable alternative possibilities, even if we employ different methods every time. But this remaining similarity between our predecessors and ourselves is not as significant as many suppose. That there exists at least one alternative explanation of one's evidence means that it is possible one is wrong, but says nothing about how plausible that is, and thus nothing about the degree or extent of our unreliability. One might respond that though the mere existence of such hypotheses at every stage is not a problem the number of them surely is. But for this gambit the pessimist will need to argue that the number of remaining alternatives is always high, and I do not see how he knows that. It is common even to suggest that the sea of remaining conceivable hypotheses that give explanations of our evidence will be infinite no matter what we do. But the evidence of history used by Stanford gives no argument for these claims, since we have (intuitively) conceived of only finitely many possibilities that our predecessors did not, and a small number at that.

If we grant the infinity of that set of unconceived conceivables for the sake of argument, the idea that it is a problem is based on some misconceptions. One source of the response is the fact that any finite number divided by infinity is zero. Thus, ruling out any finite number of further alternative hypotheses does not constitute progress on the alternative hypotheses problem because it does not raise the fraction we have ruled out and so does not raise the probability of our original hypothesis. However, while nineteenth century scientists, even physicists, ruled out hypotheses seriatim, modern statistical methods allow us to rule out classes containing an infinite number of unconceived alternative hypotheses in one stroke (Roush 2005, 218-221). And even supposing that there remain an infinite number of alternatives there can be a clear probabilistic sense in which the proportion remaining has been decreased. Mathematically, this only requires that all hypotheses are assigned finite non-zero weights that sum to one, which can be done using any convergent infinite series of fractions that sums to one.

We can rule out possibilities without conceiving of them, an infinite number at a time, but if we must always suppose some possibilities remain, this suggests another intuitive problem. We will never get to the end of this space, let us suppose, and so it seems we cannot span it or take its measure. Thus, how can we ever legitimately estimate how far we have gotten? This is a vague thought that corresponds to two real questions, but to neither of them are failures in the history of science more relevant than scientists already take them to be. The first question is how our scientist spans that space to come up with a particular probability for a hypothesis about the convection currents of the Sun. But objecting to that will require arguing with him either about the details of his evidence for that particular hypothesis or about his estimation methods, and will not require or be helped by a PI over history. The other possible argument that could be attempted here is a general one: namely, that scientists cannot possibly have grounds for evaluating the catch-all term[14] and thus not the probability of the hypothesis itself. That would be a conceptual matter to take up with a statistician, or a statistically inclined philosopher, or the scientist so inclined herself. The point is that it would not require, or be helped per se, by reference to the general fact that scientists have failed in the past.

The pessimist might protest that it is unobservable claims that are at issue, that history shows there is something especially recalcitrant about them. Many successful theories were wrong in their generalizations about observables too, but I will put that aside. This distinction makes no difference to my argument because I do not have to show that method changes make a difference to reliability about unobservable matters, even in particular cases. The unobservability of neutrinos in the example above made no difference to the fact that the issue of whether scientists are justified in believing things that they do about neutrinos depends on whether there is a cross-induction via the method used to establish things about neutrinos, which depends entirely on things like whether changing the material to gallium plausibly makes a difference to the result. This is an issue, and a kind of issue, that scientists address explicitly, so it is the scientist whom the pessimist must challenge.

Must, that is, unless the pessimist can make an argument that it is, in principle, impossible for method differences to change one's success at claims about unobservables. This might be argued on general empiricist grounds, though I think that is unsuccessful (Roush 2006, Ch. 6). But those arguments are not a PI, and if successful would not need a PI. However, perhaps a new PI could be made about method itself. We see that we have relevantly different methods from our predecessors for going at claims about unobservables, but they had apparently relevantly different methods from their predecessors too, and little good it did them for they still came up with theories that are false if our theories are true. [15] This PI is aimed at all of the factual claims—that this or that method change *is* plausibly relevant to reliability—that a cross-induction could rest on.

---

[14] This is the probability of your evidence given all logically possible alternatives to your hypothesis.

[15] Thanks to Bill Talbott for this argument. Note once again the dependence of its premise on a high base rate of falsehood of past successful theories.

However, this gambit is untenable. For what justified the pessimist's doing an induction here rather than a counter-induction? A counter-induction would have taken us to the conclusion that having failed in the past, this time the new and apparently more reliable methods *are* relevant to reliability about unobservables. Since the argument's conclusion is about unobservables, in particular about the relevance of method changes to reliability about them, the inference that was made is justified only if we think induction is a more reliable method than counter-induction at getting true conclusions about unobservables. But then this argument's conclusion that no method is relevant to reliability about unobservables undermines its own justifying inference rule. The argument's conclusion prevents legitimate inference to that conclusion. Another caution is in order with this conclusion, of course, since we could have gotten to it by Hume's argument that no inference or method makes it more rational to believe this versus that about the unobserved or the unobservable. Thus, if we want to establish that conclusion then reference to history is an unnecessary detour.

8 Too Good to be True? The Size of Potential Error

By now it may seem that my conclusion is just too good to be true.[16] Intuitively it seems that there must be something right about the PI. There is something right, though as I will argue it too is already taken into account in good scientists' particular judgments. That there is something right comes from the fact that induction is not deduction. Just as inductive support comes in degrees, so does every cross-induction come with a degree. There is a degree to which you are justified in believing the cross-induction property is present, and a degree to which you are justified in believing it is relevant to your reliability on q, those two combining for a cross of a certain degree. Some relevance of the past to the present remains because there are still similarities between our predecessors and ourselves. The fact that we are all human beings doing the scientific method in the most abstract sense is the scaffolding for the thousands of more particular features of our methods, so its relevance is not zero. Thus, something of the general PI remains, which raises the question how strong it is. It is easy to see from what I have argued above that the degree of legitimate PIs and crossings by novelty of method will co-vary with the degree of similarity and dissimilarity of method. I will now argue that both of these co-vary with

---

[16] Thanks to Catherine Z. Elgin for the very helpful objection addressed in this section. There is another apparent way of arguing that a history of past failures has got to make a difference that our particular judgments do not already pick up on, which is imagining the same current situation of evidence and theories but preceded in one case by a history of failures and in the other case by a history of successes. Surely that makes a difference to the confidence we are entitled to have in our theories. (Thanks to Shelly Kagan for this objection.) The problem with this argument is its premise. If our theories have all successes, then there cannot have ever been evidence inconsistent with our theories, since it would be part of our evidence pool. So if the same amount of evidence was collected in the two pasts then there is much more positive or neutral evidence for our theories than there would be with a history of past failures. That means that it is not possible to have the same current theory-evidence situation with the two different histories.

logical strength of hypotheses. This means that to be sensitive to the logical strength of hypotheses is already to be sensitive to the degree to which pessimistic inductions work.

First, it is surprisingly rarely appreciated that the admission that a theory is very likely false is thoroughly compatible with being highly confident of each of its particular claims that it is true. At least it is rarely appreciated that this is a good thing.[17] To illustrate, consider the example of the Standard Model of particle physics combined with auxiliary assumptions in comparison to a particular claim that follows from them, such as the existence of the Higgs boson. On the basis of successful experiments, physicists may be, and some are, quite confident though not certain that the Higgs boson exists, while also being confident that the Standard Model is false. The rationality of this can be seen if we represent the rational confidence of a subject as a probability. Then we would formulate the claim that the subject's degree of belief in q is x as $P(q) = x$, the probability of q is x. If so, then the fact that a scientists' degree of belief, x, is 99% instead of 100%—confident but not certain—makes a very big difference to what probability rationally requires when a single claim like the existence of the Higgs boson is conjoined with many others.

If one is certain in a claim, then probabilistically one cannot coherently revise it, meaning that one regards it as impossible that the claim is mistaken. It also requires that one be certain of its conjunction with other claims one is certain of. In contrast, if one has even a sliver of a doubt about individual claims, and those claims are independent, then one's confidence in the conjunction must be exponentially lower than that in any of the individual claims. If there are 16, let us say independent, claims of existence of elementary particles, then even if the scientists are extremely confident about each, say 99%, rationality requires them to have a 15% confidence that at least one of them is wrong. This doubt in the conjunction grows with the number of conjuncts: with 40 independent claims, the required degree of doubt in the conjunction is 27%. The degree of required doubt increases even faster the lower the confidence in the original individual hypotheses. If one is 95% confident in each of the 16 claims about the particles—only 4% lower than just supposed—then one will be required to be *more confident than not* that at least one of them is wrong: 57%. (Starting with 99% it takes 100 claims to get to more likely than not that the conjunction is false.)

To see what this tells us about theories, we have to say what a theory is. If we idealize, then a theory can be written as a set of a few independent law-like generalizations. However, that will have no empirical consequences for this world, the actual world, without adding a lot of auxiliary assumptions about this world. For example, Newton's theory of mechanics can be compactly expressed as three laws, but one must specify where the massive objects actually are at a given time, how massive they are, and so on, in order to figure out what this theory says about where

---

[17] An exception is Philip Kitcher (2001). Many take this probabilistic fact to indicate a "paradox of the preface," mistakenly in my opinion. (Roush 2010)

they are going to be at a different time. If seen as a proposition, then a universal, substantive theory is a huge conjunction, and with the number of claims increasing at this level, the required confidence that the conjunction is wrong increases dramatically. If we have a million claims, then even if we have 95% confidence in each we are required to be more sure that at least one of them is false than we are sure of any individual one of them. In our example, if we are 95% confident that the Higgs boson exists, then to remain rational we must be 96% confident that the Standard Model is false. From the other side, even if we are 96% confident that the Standard Model is false, very high confidence in the existence of the Higgs boson does not make us irrational.

One might object that the magnitude of this admission that high theories are likely false is inadequate to address the point of the pessimist. This is merely confidence that at least one of the million claims is false. What we see in the historical record is cases where a big part of the big idea was wrong in a big way. Should logic alone make us confident of the same about our own theories, and if so, then would that not amount to winning the battle but losing the war against the pessimist? One might mean several different things by "big", but we can address the objection by measuring size of error as how many or what fraction of one's claims were wrong.[18] This would capture, for example, the idea that a big claim has many implications. Consider a theory with one million independent claims as earlier, and suppose that we are 85% sure of each of them. Then we will be obligated to be 97% confident that the theory is false somewhere. But also, at this 85% level it only takes five claims to be required to be 56% sure that one of them is false.[19] That, as like as not, one of every five of my claims is false is a substantial admission: one fifth of one million is two hundred thousand. Yet this does not prevent the rationality of my 85% confidence in each one of the one million claims. This is as it should be because the big admission that two hundred thousand of my claims are more likely than not to be false, and the reference to the "big" mistakes of history, give us no hint of which of my claims are at fault. Recall that it is scientists' right to go on with their practice of making this or that particular claim in keeping with their usual evidence and arguments that I am concerned to defend, and our example of a big admission does not impose a confidence drop below 85% on any particular claim. A general theory is equivalent to a huge conjunction, so we should be well aware without looking at our predecessors that it is very likely to be false, and it is clear on reflection that that does not impose a low confidence for any particular claim.

Logical strength of a hypothesis also makes it more susceptible to a PI by making it harder to rescue by a cross-induction on method. Consider the 16 particles of the Standard Model.

---

[18] Is the structure of a theory the big part, or is it the types of entities the theory takes to exist that are big? Is the hypothesis that the ether exists big because the ether was supposed to cover the entire universe and affect every motion? Or was it small because it was a single proposition that, as it turned out, is independent of much of the rest of the theory it was housed in? Answering these questions would do much to carve out one's brand of realism or anti-realism, which is not my purpose here.

[19] If my confidence in each is .75 then I must be .76 confident that one of every five claims I made is false.

Verifying all of those particles supports the theory to some degree, and the whole set of these verifications gives stronger support to the theory than any subset of them would. However, these particles were verified using a very wide array of types of methods, i.e., particle detectors. The bubble chamber is different from the spark chamber, and neither is much like the Large Hadron Collider. To defend the Standard Model against a PI, we should appeal to differences between the method we used to test it and those our predecessors used on their theories. But if we were to state the method by which this theory was tested or supported, it would have to be a quite generic saying because it would have to be true of all of those methods that went into the verification, and relatively little is. There will be some statistical methods that all of our particle experimenters have used and their predecessors did not have, but what is common in the verifications of the particles does not go a great deal beyond that. By contrast, someone who used the bubble chamber to detect kaon decay would have a great deal to say about how his method was relevantly different from those of his predecessors who tried to detect the unseen. We necessarily have less material for a cross-induction on method on behalf of a theory than we will on behalf of particular, logically weaker, claims that follow from the theory. Logically weaker claims are not only easier to support by evidence (other things equal), but also more resistant to the PI. Notice once again that the distinction between observable and unobservable made no difference to the arguments of this section. Even supposing it exists, the Higgs boson is unobservable.

9 Conclusion

The history of science appears to pose two threats to the rationality of science, one due to radical paradigm shifts, the other due to our predecessors' track record of failures to get true theories. These challenges are duals in that one rests on the consequences of too much discontinuity, the other on those of too much continuity. However, these apparent problems are illusions due to under-description, and they disappear when with more specific descriptions of the history we find additional continuity and discontinuity respectively. It appeared that history cast doubt on the rationality of science, but the rationality of science is saved by an eye for detail that is characteristic of the historian.

In particular, the pessimistic induction over the history of science is powerless to create justified doubt about our particular hypotheses that is not already addressed, or prepared for, in good scientists' arguments about particular conclusions. For scientists to become doubtful about particular hypotheses on the basis of a general induction over history would be for them to double-count the evidence, and to ignore the relevance that difference of method has to reliability. It would be as if, like characters in an Ionesco play, they were to infer that paper towels are white from seeing white swans, because they saw them in the same country.

References

Donovan, A., L. Laudan and R. Laudan (1998), *Scrutinizing Science: empirical studies of scientific change*. Kluwer: Boston.

Galison, P. (1997), *Image and Logic: A Material Culture of Microphysics*. University of Chicago: Chicago.

Harman, G. (1980), "Reasoning and Evidence one does not Possess", *Midwest Studies in Philosophy* 5(1): 163–182.

Hollinger D. (1973), "T.S. Kuhn's Theory of Science and its Implications for History", *American Historical Review* 78(2): 370–393.

Kitcher, P. (2001) "Real Realism: The Galilean Strategy", *Philosophical Review* 110 (2): 151–197.

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

Lewis, P., 2001, "Why the Pessimistic Induction is a Fallacy", Synthese 129: 371–380.

Magnus, P.D. and C. Callender (2004), "Realist Ennui and the Base Rate Fallacy", Philosophy of Science 71: 320–338.

Poincare, H. (1905), *Science and Hypothesis*. Scott: London.

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

Roush, S. (2009), "Second-Guessing: A Self-Help Manual", *Episteme* 6.3: 251–268.

Roush, S. (2010), "Optimism about the Pessimistic Induction", in Magnus, P.D. and Busch, Jacob, eds., *New Waves in Philosophy of Science*. Palgrave-MacMillan.

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