Global and Local Pessimistic Meta-Inductions

Samuel Ruhmkorff
Bard College at Simon's Rock
84 Alford Rd.
Great Barrington, MA 01230
sruhmkorff@simons-rock.edu

The global pessimistic meta-induction argues from the falsity of scientific theories accepted in the past to the likely falsity of currently accepted scientific theories. I contend that this argument commits a statistical error previously unmentioned in the literature and is self-undermining. I then compare the global pessimistic meta-induction to a local pessimistic meta-induction based on recent negative assessments of the reliability of medical research. If there is any future in drawing pessimistic conclusions from the history of science, it lies in local meta-inductions, but these meta-inductions will not result in global distrust of the results of science.

1. Introduction

We are accustomed to evaluating agents and practices by their track records. Some have attempted to evaluate science negatively on the basis of its track record. At first blush, this is surprising, since science appears to be one of the most successful human activities. Room for pessimism is generated by the reflection that most theories in the history of science have been replaced by successor theories. Successor theories can be radically different from those they replace, and sometimes do not even recognize the entities countenanced by their predecessors. The global pessimistic meta-induction holds that according to current scientific thinking, most of the conclusions reached in the history of science have been false, and infers from this the likely falsity of currently accepted scientific theories. I argue that the global pessimistic meta-induction commits a statistical error previously unmentioned in the literature and is self-undermining. I then compare the global pessimistic meta-induction to a local pessimistic meta-induction based on recent negative assessments of the reliability of medical research. If there is any future in drawing pessimistic conclusions from the history of science, it lies in local meta-inductions, but these meta-inductions will not result in global distrust in the results of science.

2. Pessimistic Reasoning from the History of Science

There are two main ways that philosophers have reasoned pessimistically from the history of discarded scientific theories. The first, the pessimistic induction (PI), is Larry Laudan's attempt to refute the no miracles argument for scientific realism by attacking the link it alleges between success and truth in scientific theories (Laudan 1981). The second, the pessimistic meta-induction (PMI), concludes that currently accepted scientific theories are likely false because previously accepted theories have been overwhelmingly false (Putnam 1978, 25).[1] PMI can be deployed in local or global versions. Local versions make an inference either to the likely falsity of theories accepted on the basis of a specific methodology or to the likely falsity of theories in a restricted field of inquiry. Global PMI makes an inference to the likely falsity of all currently accepted scientific theories. The literature on PMI focuses on the global version. Henceforth, I use 'PMI' to mean this version unless otherwise specified.

Although it has been argued that Laudan intended only to defend PI (Enfield 2008; Park 2011; cf. Wray 2013), his infamous list outlining prominent examples of instrumentally

[1] There is not agreement among philosophers about how to use the phrase 'pessimistic induction' (see Wray 2013, 1723n). Some discussions use the phrase to refer both to PI and PMI (Doppelt 2007; Lange 2002; Psillos 1999; Worrall 1994; Stanford 2006; Lewis 2001; Fahrbach 2011a; Newman 2005); some, PI only (Hardin and Rosenberg 1982; Enfield 2008; Kitcher 1993; Elsamahi 2005; Ritchie 2008; Saatsi 2005; Lyons 2006; Fahrbach 2011b); some, PMI only (Magnus and Callender 2004; Hobbs 1994; Bishop 2003; Leplin 1997; Magnus 2010; Park 2011); and others do not specify (Chang 2003; Chakravartty 2004).

successful yet false theories in the history of science has served as a powerful articulation of the historical basis for PMI (Laudan 1981, 33). He writes: 'I daresay that for every highly successful theory in the past of science which we now believe to be a genuinely referring theory, one could find half a dozen once successful theories which we now regard as substantially non-referring [and hence false]' (Laudan 1981, 35). He also furnishes examples of several prominent successful past theories which, while referring, are not regarded as approximately true. Thus, according to Laudan, the ratio of false but successful past theories to true and successful past theories – as seen from the perspective of current theory – is greater than 6 to 1.

This historical claim supports the premise of the simplest version of PMI:

A1 The majority of theories accepted in the history of science are false.

A2 Therefore, by induction, currently accepted theories are likely false (see Putnam 1978, 25; Psillos 1999, 101; Stanford 2006, 7).

As Jarrett Leplin (1997, 142) has observed, this form of PMI is self-undermining in that A1 evaluates the truth of historical theories from the standpoint of current theory, while A2 concludes that current theory is likely false. To avoid this difficulty, PMI can be presented in reductio form:

B1 Assume that most currently accepted theories are true.

B2 Then most scientific theories accepted in the past are false, since they differ from current theories in significant ways.[2]

[2] B2 is from Peter Lewis' explication of PI (Lewis 2001, 373, following Psillos 1999, 102-103).

B3 Therefore, by induction, most currently accepted theories are false.

B4 Therefore, by reductio, most currently accepted theories are false.

This argument has been disputed in many ways. For example, it has been alleged that PMI commits the base-rate and turnover fallacies (Lewis 2001 and Lange 2002, respectively); that the inductive projection in B3 is blocked because past scientific theories are relevantly different from current scientific theories (see Section 3); and that the inference in B2 is invalid (Mizrahi 2012). In what follows, I argue that global PMI faces two objections that have not yet been raised by critics. I then state and evaluate a local PMI in the medical sciences based on recent negative assessments of this field by John Ioannidis and others. I argue that this local PMI avoids some of the problems that confront global PMI, but that local PMIs cannot be combined to support a global pessimistic outlook on science.

Before I begin, some terminological bookkeeping: by a scientific 'domain', I mean a singular scientific issue, e.g., the explanation of a particular kind of phenomenon such as the interaction between temperature and pressure. For each domain there is a partition of pair-wise inconsistent scientific theories exactly one of which is true. By a scientific 'field', I mean a collection of related domains, for example, those relating to the origins of species. Since scientific theories often attempt to explain phenomena from different domains or reduce phenomena in a given domain to more fundamental domains, there will be overlap in the partitions for different domains. For example, Newtonian mechanics will be part of the partitions for the domains of both earthly projectile motion and planetary orbits. It is still the case that exactly one of the theories composing the partition for a particular domain is true. Finally, it is usually assumed that PMI's proponents are antirealists and its target is realism. But there are some accounts of realism which are not committed to the approximate truth of currently accepted scientific theories (e.g., Godfrey-Smith 2003, 175-177). To avoid complexities regarding the proper construal of realism, I characterize the debate about PMI as between pessimists (those who believe that currently accepted theories are likely false) and optimists (those who think that currently accepted theories are likely true).

3. Global PMI Ignores the Possibility of Attrition

PMI involves a statistical projection from the ratio of false to total theories accepted in the history of science to the ratio of false to total theories accepted currently. Basic statistical inference requires that the sample be selected randomly and that the relevant properties of the individuals in the sample and population be logically independent. It has been argued by a number of critics (e.g., Lipton and Worrall 2000, 204; Magnus and Callender 2004, 327; Stanford 2006, 10; Fahrbach 2011a, 2011b; Park 2011; Mizrahi 2012) that PMI fails the criterion of randomness because the selection of the inductive basis is limited by our position in the history of science.[3] Since it is plausible that, as the passage of time brings with it more accurate observations, greater sophistication in methods, and the accumulation of evidence, science is increasingly more likely to get at the truth, we should not project the reliability of science in the past to present-day science.

I contend that PMI also fails to satisfy the criterion of independence. Simple projection is inappropriate when predications of the properties involved to the entities involved are not logically independent. Suppose the first five suspects that Poirot examines have convincing alibis. He may then be said to have inductive evidence that no one committed the murder! But of course, he knows that someone did — and he will infer that ceteris paribus it is more likely, not less likely, that the next suspect he examines is guilty.

Scientific theories in the same domain are pair-wise inconsistent with each other. Even outside of a single domain, there are logical relationships among the theories within a field and in different fields, in part due to unification and reduction. Simple inductive projection about the ratios of truth and falsity of scientific theories is inappropriate because as long as there are a finite number of theories in a domain, the falsity of past theories will tend to increase rather than decrease our confidence in current theory. Since it ignores the logical relationships between successive scientific theories in the same and different domains, PMI ignores the role that attrition plays in eliminative inference.[4]

I will respond to two objections to my claim that PMI fails to meet the criterion of independence. To discuss the first objection properly, I must distinguish two versions of PMI. One is the version explicated above: most current theories are false. This supports a modest pessimism according to which we shouldn't believe current theories. The other is more ambitious: scientific theories accepted now and in the future are likely to be false. This supports a strong pessimism according to which we should be continually skeptical of the results of scientific theorizing.[5] Because modest pessimism is logically weaker than strong pessimism, it is the focus of my critique.

---

[3] There are statistical ways of handling not being able to sample from future members of populations, a phenomenon known as right-censoring. More on this in section 5.

[4] For more on eliminative inference, see the work of John Earman (1992), Philip Kitcher (1993), and Patrick Forber (2011).

[5] I am grateful to an anonymous referee for stressing the importance of this distinction. Although the general idea of strong PMI is easily stated — the population is expanded to include future as well as past and current theories — it is difficult to explicate. For example, the reductio in B1 - B4 does not work because it doesn't make sense to assume that current *and* future theories are true. I overlook this difficulty as my main concern lies with modest PMI.

The first objection is that there are enough unconceived scientific theories to make PMI's failure of independence insignificant. For example, it might be claimed that there are infinite scientific theories in each domain. Just as a detective with infinite suspects would not gain any confidence that a given remaining suspect is guilty by eliminating other suspects, no matter how many suspects she eliminates, optimists cannot claim that, ceteris paribus, the fact that a number of theories have been eliminated in the history of science makes it more likely that current theories are true.

In response, I am not convinced that for every domain of inquiry, there are an infinite number of scientific theories. As a point of logic, there are an infinite number of theories which have the phenomena in that domain as consequence. But this doesn't mean that all of these theories are scientific.

Even if we grant that there are an infinite number of scientific theories in each domain, not all of them would be equally plausible. There may be chunks of possibility space that are very implausible. For example, it is consistent with there being an infinite number of scientific theories in a given domain, that only ten of them are plausible while the others are wildly implausible. In this sort of scenario, attrition would be a factor as some plausible theories become refuted throughout the history of science (pace Lipton 1993, 94). Thus, one cannot infer from the claim that there are an infinite number of scientific theories in each domain, that pessimism about theories accepted now or in the future follows from a past track record of mostly refuted theories. In addition, one needs to provide evidence that there are enough plausible theories among these infinite theories to block attrition — something that has not yet been established by PMI. Where the burden of proof lies is important (Devitt 2011, 290n). To rebut PMI, optimists do not need to provide evidence that there are a limited number of plausible theories in each domain. PMI makes a positive claim about the likely falsity of current (or current and future) scientific theories. Therefore PMI must provide reason to think there are enough plausible theories in most domains to overcome the failure of independence.

How many theories are sufficient to overcome the failure of independence depends on the version of PMI in question. Technically, attrition can function so long as there are a finite number of plausible theories in a given domain. Yet if we somehow knew there were 500,000 plausible scientific theories in a given domain, attrition would be futile. We could hardly hope to generate and evaluate this many theories by the end of human civilization. So establishing a large finite number of plausible theories is sufficient to support strong PMI. Much less is needed to support modest PMI. The demonstration of the likely existence of even one or two unconceived theories as plausible as our best theories is enough to support the claim that our current best theories in a given domain are likely false.

I will consider two ways defenders of PMI might argue for the likely existence of one or many plausible unconceived theories in most domains, the first conceptual, the second empirical. First, the defender of PMI may contend that theories by themselves cannot be thought to be scientifically plausible or not. Rather, judgments of scientific plausibility concern triples: <theory, background theory, data>. Since plausibility judgments are indexed to background beliefs and available data, the number and identity of plausible theories will change over time. This means that there is no fixed set of plausible theories on which a strategy such as attrition might operate. Moreover, it means that there are a great number of theories that are plausible to some scientists at some time, and thus that attrition on human time scales will not generate any significant result.

However, the dependence of judgments of scientific plausibility on background theory and data does not establish a priori that there are enough scientifically plausible theories to threaten the meaningful use of attrition, because the other relata involved in these judgments cannot be shown a priori to be unstable. The issue is not whether, hypothetically, there might be a plenitude of scientifically plausible theories. It is whether there actually are such theories. If our current data turn out to be stable (meaning that, while new data are added, previous data are not overturned), and our background beliefs are also relatively stable, then there very well might not turn out to be many scientifically plausible theories over the course of the history of science despite the relativity of judgments of scientific plausibility. An additional consideration is that the bar for scientific plausibility is raised over time as successive scientific theories display the empirical and theoretical virtues to greater degrees (Doppelt 2007, 111). This lowers the number of plausible theories over time.

Might PMI be asserting that the history of science provides evidence that there is instability in the other relata involved in plausibility judgments such that it is likely that there are theories which are not deemed plausible now (either because they are considered and deemed implausible or simply not considered) but which will be deemed plausible in the future? This move also runs afoul of the criterion of independence. For example, the various background theories accepted at different points in the history of science are not logically independent, so the failure of previously accepted background theories makes it more likely that currently accepted background theories are correct and thus likely to be stably accepted.

The second way defenders of PMI might argue for the likely existence of one or many plausible unconceived theories in most domains is by appeal to Kyle Stanford's claim that the historical record establishes the likely existence of scientifically plausible unconceived alternatives to our best theories (2006).[6] However, as I have argued, Stanford's inference from the existence of instances in the history of science in which scientists have overlooked previously unconceived plausible alternatives to their best theories, to the likely existence of unconceived plausible alternatives to our best theories, is flawed. At most, it establishes that individual scientists are not reliable detectors of such alternatives when they exist. It does not give any reason to think there are such alternatives (Ruhmkorff 2011, 879-883), or that science as a corporate body cannot remedy the failings of individual scientists, a point made by Forber (2008) and Peter Godfrey-Smith (2008).

Even modest PMI needs to be supplemented with evidence that there are unconceived plausible alternatives to our best theories in most domains of science. The available conceptual and empirical arguments for this claim are lacking. Until PMI provides this evidence, it does not establish that its failure of independence is overcome by enough plausible unconceived alternatives to block attrition.

The second objection I will consider admits that the failure of independence in PMI opens the possibility for attrition, but maintains that the historical record gives us reason to think that the negative track record of science outweighs whatever gains there are from the process of attrition. This objection points to non-controversial cases where pessimistic reasoning and attrition are both operative, and where losses from pessimistic reasoning overcome gains from attrition. A simple example is a leaky urn. Suppose that we have an urn which initially has n balls, one of which is red and the rest of which are white. Suppose further that we know that this urn has a leak such that, at the beginning of our exploration, the probability that the red ball has leaked out (R) is p. The balls in the urn represent extant theories. The red ball is analogous to the

[6] It is unclear whether Stanford thinks there are enough unconceived alternatives for modest or

true theory in a given domain; the white balls, false theories. Since the red ball's leaking out models the truth's not being contained among extant theories, the potential leak affects the red ball only. We can calculate how much drawing and observing white balls would increase our credence in R by iterations of Bayes' Theorem. For example, if p = .1 and n = 100, and where P_i is the probability function after a sequence of *i* white balls have been drawn, then successive probabilities of R on a sequence of failures to select the red ball include: P_{10}(R) = .11, P_{50}(R) = .18, P_{90}(R) = .53, P_{99}(R) = .92. Here, the negative track record overcomes the force of attrition, and after drawing 99 white balls we should be fairly certain that the last ball drawn will not be the red ball.

When applied to the history of science, the leaky urn model has some virtues not possessed by PMI construed as a simple statistical projection. For example, it takes into account the logical relationships among theories in a given domain. While this model requires us to determine the prior probabilities of the scientific theories and the hypothesis that the true theory is not among extant theories (the 'catch-all' hypothesis; see Shimony 1970), this is a difficulty shared by at least some optimists. Most importantly, the leaky urn model demonstrates that even in a situation with attrition, the poor track record of a process can significantly raise the probability that it has not achieved the specified result.

Despite these virtues, there are at least two fundamental problems with applying the leaky urn model to the history of science. First, we do not typically begin our evaluation of scientific theories with a complete roster of all potential candidates. Rather, scientists explore the best available theory with half an eye out, more or less, for better alternatives. When these alternatives appear, they are typically previously unconceived. This means that there are not

---

strong pessimism (see Ruhmkorff 2011, 879-880).

priors in place for them, and the above kind of calculations do not apply. One might think that the probability for previously unconceived hypotheses should come entirely from the catch-all. But this would preclude the discovery of a new theory lowering one's confidence in extant theories. Since there are no priors for a previously unconceived hypothesis, and it can't simply be subsumed under the catch-all, it seems we must resort to the kind of heterodox model endorsed by Earman: 'the exploration of the space of possibilities constantly brings into consciousness heretofore unrecognized possibilities. The resulting shifts in our belief functions cannot be described by means of any sort of rule of conditionalization' (Earman 1992, 183). When we come across specific possibilities which we had not previously conceived, we may adjust the probabilities of the catch-all and of extant theories without conditionalizing.

Upon discovery of a new theory clearly superior to extant theories, PMI would have us (1) decrease our credences in the previous theories, (2) assign the new theory a credence which is at the least higher than our (now-decreased) credences in the extant theories, and (3) increase our credence in the catch-all because we now have another instance in which our theories turned out to be incorrect. This is in contrast to an optimistic reaction, which would involve (1) and (2) but also (3*) a decrease of credence in the catch-all representing the result of attrition. Thus the difference between a pessimistic and an optimistic reaction to uncovering a new, superior theory comes down to whether consequently one increases or decreases one's probability of the catch-all.

The fact that we can characterize the difference between a pessimistic and optimistic reaction to a typical case of theory change in science as different responses to a situation of throwing away the priors should make it clear that we are not going to have a Bayesian rule of inference that will force us to change our credences in one way or another. If we aren't forced to infer one way or the other, PMI does not compel us to take the pessimistic route.

The second problem with the leaky urn model is that even if scientists were to have candidate theories available to them ahead of time such that they had well-defined priors, the inferences they would make would depend on their initial prior probabilities about whether science will get at the truth. In the leaky urn model, whether we derive a pessimistic outcome from eliminating all but one ball is dependent on what p is. No matter how many balls there are, there are values for p which mean we shouldn't be pessimistic even if we have looked at every ball except the last with negative results each time. Likewise, in the case of science, how pessimistic we should be after observing a run of theories that are refuted depends on what our priors are for the true theory being one of the theories that scientists generate.

Whether we think (more realistically) of being in a situation of throwing away the priors upon discovering new theories, or (less realistically) of being in a situation in which we know all of the theories in advance and assign them priors, how pessimistic we are after observing a run of refuted theories depends on how we choose to revise our priors in the face of previously unconsidered theories (in the former case) or on our prior probability for p (in the latter case). Unlike the simple statistical model, the leaky urn model models attrition, but it does not accurately model science, and it fails to force the optimist into the pessimistic result desired.

4. Global PMI is Self-Undermining

I have argued that PMI fails to satisfy the statistical criterion of independence, leading it illicitly to ignore the possibility of attrition. I will now present the case that PMI is self-undermining. Although it is not uncommon for skeptical arguments to be vulnerable to the objection that they are self-undermining, the claim that PMI is self-undermining has not been discussed by critics of PMI. In his limited defense of PMI, Jesse Hobbs (1994) attempts to avoid the claim that PMI is self-undermining. I contend that PMI is in fact self-undermining, and that Hobbs' method of avoiding this result renders the conclusion of PMI harmless to the optimist.

The simplest version of PMI (A1 and A2) undermines itself because its inductive projection takes away our justification in believing the currently accepted theories used to establish the falsity of past theories (section 2). This problem is avoided by the reductio version of PMI. However, there is another self-undermining problem which the reductio version of PMI does not escape. PMI uses one of the methods of science — induction — to generate skepticism about conclusions reached by the methods of science. If currently accepted theories are false, then the methods by which we arrive at them, including induction, are not reliable. But induction is precisely what is used to arrive at B3 in the reductio step. It is acceptable for premises in the scope of a reductio assumption to be undermined by the overall conclusion of the argument, but it is not acceptable for a mode of inference that is independently assumed to be reliable and that is used in the scope of a reductio assumption to be undermined by the conclusion.

The most natural way to avoid this problem would be to differentiate between the methods used by science and the inductive projection from B2 to B3. If the argument of the preceding section is correct, we can identify the feature found in most scientific inference but not in PMI: the satisfaction of the criteria of randomness and independence. But, of course, this avoids the self-undermining objection only at the cost of admitting the fallacious nature of PMI. What else might distinguish PMI from the kinds of inferences commonly used in science? Since PMI lumps together conclusions generated by all the methods of science into its inductive basis, it does not seem able to separate its own method from these methods in a non-ad-hoc fashion. It is possible there is some mistake made by most scientific reasoning which is not made by PMI, but the carpet-bombing approach of global PMI does not give us a clue as to what this mistake might be. Peter Lipton considers the idea that PMI satisfies the tracking requirement. If it did satisfy this requirement, and many scientific inferences did not, that would be a candidate for the feature that blocks PMI's application to itself. However, Lipton has a powerful argument that PMI does not satisfy the tracking requirement (Lipton 2000, 200-204).

A more promising escape route relies upon the work of John Norton. Norton argues that the difficulties presented in characterizing and justifying induction are due to the fact that no general account of induction is possible. He contends that 'all inductive inference is local,' with its legitimacy grounded in the material facts of particular domains (Norton 2003, 647). The various inductive inference schemas presented by philosophers apply only in specific contexts and when backed by appropriate local facts. For example, inductions about the melting point of a substance are grounded in facts about its elemental nature (Norton 2003, 649). The defender of PMI could claim that induction fails in most scientific domains because we are mistaken about the facts in those domains, but that it is reliable when used by PMI, because PMI is grounded in facts about the past failures of scientific theories. This is an elegant solution to the self-undermining problem. However, if Norton's material theory of induction is correct, it is hard to see how global PMI can be persuasive. PMI projects past failure onto current theories. On Norton's view, there is no significant commonality between the inductive methods used by past and current theories unless the theories are grounded in the same purported facts — in which case current theories would not provide a negative assessment of past theories.

There does not appear to be a non-ad-hoc way of differentiating the reasoning used in PMI from the reasoning used in the scientific inferences it targets which preserves the inductive power of PMI. An attempt might still be made to avoid the self-undermining problem by reformulating PMI. A first attempt is to frame PMI as an inconsistency argument (as suggested by Leplin 1997, 142; Lewis 2001, 372; Stanford 2006, 146):

C1 In each scientific domain, there have been more than two pairwise inconsistent theories accepted by scientists over time.

C2 Therefore, most accepted scientific theories are false.

C3 Therefore, currently accepted scientific theories are likely false.

The inconsistency version of PMI appears not to run afoul of the self-undermining problem, because C1 is a combination of historical and logical claims, and C2 is a logical entailment from C1. Thus it does not appear to use scientific methods, strictly speaking. This conclusion, however, is not correct, because the inference to C3 involves the (questionable!) statistical assumption that currently accepted theories are a random sample of accepted theories.

Hobbs (1994, 179) attempts to preserve the global ambitions of PMI by modifying the reductio version of PMI. This might go something like:

D1 Assume that scientific method is reliable.

D2 Then induction is reliable and most currently accepted theories are true.

D3 Then most scientific theories accepted in the past are false, since they differ from current theories in significant ways.

D4 Therefore, by induction, most currently accepted theories are false.

D5 Therefore, by reductio, scientific method is not reliable.

This new reductio points out that scientific method contains within it a tension: if it is reliable, we have reason to think it is not reliable. Because induction is not independently assumed to be reliable, the conclusion's denial of its reliability is not self-undermining.

This new reductio avoids the self-undermining problem. However, it results in a new conclusion for PMI. The conclusion is no longer 'most currently accepted scientific theories are false', but rather 'scientific methods achieve inconsistent results'. This reframed conclusion is not as strong a claim as the original. Rather than asserting something definitive about currently accepted scientific theories, it tells us that scientific methods have not been reliable over time. This conclusion is compatible with currently accepted theories being mostly true and past scientific theories being mostly false. The defender of PMI faces a dilemma: either make the reductio about the truth of currently accepted theories and have the use of inductive method be undermined by the conclusion of the argument, or have the reductio be about the reliability of methods and be unable to assert anything about the likely falsity of currently accepted theories.

In a further attempt to avoid the self-undermining objection, Hobbs observes that 'the pessimistic induction asserts only that most — not all — scientific theories are false, so there is no inconsistency in supposing that it is one of the theories that is not' (Hobbs 1994, 179). True, but the objection at hand is that PMI is epistemically, not logically, self-undermining. PMI, if successful, undermines our justification in believing the results of scientific method. According to PMI, one of the results of scientific method is PMI. Therefore, according to PMI, we are not justified in believing PMI.

5. Local Pessimistic Meta-Inductions and their Prospects for Global Pessimism

In this section, I discuss recent meta-analyses in medicine which find that a disconcerting high percentage of prominent medical research findings are refuted by subsequent research; show how these meta-analyses can be developed into a local PMI; and argue that, while this local PMI is less problematic than global PMI, it is not possible to conjoin multiple local PMIs to support global pessimism.

A number of researchers, most notably Ioannidis (e.g., 2005a, 2005b), have been examining the rate at which results are overturned in medical studies. For example, a 1991 study appeared to show that hormone therapy for women greatly reduces the risk of coronary artery disease, while later, more powerful studies in 1998 and 2002 showed that such therapy actually leads to worse coronary outcomes; and a 1993 study appeared to show that vitamin E reduces the risk of coronary disease, while a later, more powerful study in 2000 showed that it has no effect (Ioannidis 2005a, 223). There are a number of dimensions along which these refutations have been examined, including determining their frequency among highly cited articles (e.g., Ioannidis 2005a), explaining why they occur (e.g., Ioannidis 2005b), measuring the responsiveness of the literature to refuted findings (e.g., Tatsioni et al. 2007), and examining the relative reliability of different kinds of studies, for example, randomized clinical trials versus nonrandomized epidemiological studies (e.g., Balk et al. 2002, Ioannidis et al. 2001). I focus on the first two of these dimensions by discussing an influential study by Ioannidis.

Ioannidis (2005a) found that 16% of highly cited articles (>1000 citations by August 20, 2004) in journals with the highest impact factors were contradicted by later, more powerful studies, and an additional 16% were found by later, more powerful studies to have effect sizes of less than 50% of those originally claimed. 24% of the studies were not addressed by subsequent research. This means that 41% of the results for which replication was attempted were contradicted or found to be inflated by a factor of at least two. Since, from the standpoint of the philosophy of science, a claim's having an effect size of less than half usually counts as being refuted,[7] this means that about 41% of highly cited results were contradicted by later research.[8] A control group of less frequently-cited articles from the same journals and time period fared slightly better, with 33% of tested results being contradicted by later research (Ioannidis 2005a, 220). These results give us significant reason to be skeptical of the results of medical studies, even those which are published in the best medical journals and highly cited.

Ioannidis' diagnosis of this high rate of contradiction is complex. Contributing factors include: bias in research (Ioannidis 2005b); nonrandomized trials (of which 5 out of 6 were contradicted; but cf. Ioannidis et al. 2001, and Ioannidis 2005a, 225, where he notes the possibility that nonrandomized studies are more likely to be challenged sooner and thus are likely overrepresented among contradicted studies); smaller rather than larger sample sizes in refuted studies (Ioannidis 2005a, 224); and publication and time-lag biases (whereby studies with highly significant and potentially aberrational positive results are over-represented among published articles in major journals and are published more quickly than other articles) (Ioannidis 2005a, 224). Particularly intriguing is the idea that large-scale features of the structure of medical and biological inquiry contribute to the high contradiction rate. Having a number of distinct working groups looking at the same problem increases the chances that at least one of them will find something statistically significant, especially if they are looking at a wide array of

[7] See below for a caveat to this claim.
[8] In my calculations, I leave out theories unchallenged by subsequent studies. Ioannidis (2007a, 225) points out that sometimes studies are not challenged because they are seen as definitive, and ethical and other considerations bar attempts at replication. My percentages may be slightly more

possible relationships (Ioannidis 2005b, 697-698). The computational power and richness of data sets available to researchers increases the chance that some of them will be successful in achieving statistical significance, even when no real relationship exists (Ioannidis 2005b, 701).

There are some aspects of Ioannidis' work that are problematic. First, in his study there is right-censoring, the phenomenon where relevant data is unable to be collected due to the observer's position in time. Ioannidis' position in time when conducting the study prevented him from knowing both whether eligible studies would turn out to be contradicted after 2004 and whether some studies in the time range will turn out to meet the criterion of having at least 1000 citations after August 20, 2004. If the average time it takes for a research finding to be subjected to rigorous testing is longer than the average time it takes to reach 1000 citations, then the sample would have a tendency to include some studies not yet subjected to rigorous testing, and thus tend to understate the refutation rate. If the average time it takes for a research finding to be subjected to rigorous testing is shorter than the average time it takes to reach 1000 citations, there would be a tendency for some refuted studies which eventually meet the criterion of being highly cited to be excluded from the sample; again, the refutation rate of highly cited studies would tend to be underestimated.

Second, the compatibility of studies on Ioannidis' analysis depends on the order in which they are published. Ioannidis sorts hypotheses into three categories relative to the claimed effect sizes of the initial study (E_i) and the subsequent study (E_s): $E_s \leq 0$, $0 < E_s \leq 1/2(E_i)$, $1/2(E_i) < E_s$ (Ioannidis 2005a, 219).[9] The breadth of these categories and the fact that they are defined relative to the initial study make for perverse results. For example, if the initial study finds that

pessimistic than is warranted, as I am excluding some studies which are more likely to be reliable than a random member of the sample.

rosuvastatin 5mg/day reduces LDL cholesterol 10% and the subsequent study finds that the reduction is 65%, then the initial study counts as confirmed by the subsequent study — a strange result in itself. But if the initial study finds cholesterol is reduced by 65% and the subsequent study finds that the reduction is 10%, this counts as a refutation. The publication order of two studies should not be a relevant factor in determining whether the results of those studies are categorized as inconsistent, yet Ioannidis' methodology allows this to happen.

Third, because Ioannidis' categorization scheme sorts studies by the proportional relationship of their results to those of the initial study, essential information about the absolute distance between the results of the studies is lost. For example, Ioannidis' categorization treats results of a 5% and a 2.4% reduction of cholesterol in successive studies in the same way as it would a 50% and a 24% reduction of cholesterol: both are cases of initially stronger effects. Yet it is not accurate to treat these cases in the same way because the results of the first pair of studies are more congruent than those of the second pair and the distances between the results (2.6% and 26%, respectively) would likely bear a different relationship to the studies' respective margins of error.

Before I state the local pessimistic meta-induction, it is important to be precise about what is established by Ioannidis' results. He is concerned with studies satisfying certain criteria (henceforth 'M-studies'). These criteria include being highly cited, using contemporary research and statistical methods, and being among the first studies to investigate the question at issue. The reason for this last constraint is that, as Ioannidis (2005a, 224-225) acknowledges, skeptical conclusions about early results cannot be projected onto later situations where more evidence has aggregated. M-studies tend to state their results in terms of absolute or relative risk reduction

---

[9] Things are not always so tidy; for example, 'diminished effects' also includes shorter duration

among the subjects in the sample. In their conclusions, nonrandomized M-studies have a strong tendency to assert the existence of an association rather than a causal connection, while randomized M-studies have a tendency to make causal claims (e.g., 'We conclude that in patients with angiographically proven symptomatic coronary atherosclerosis, α-tocopherol treatment substantially reduces the rate of non-fatal MI' (Stephens et al. 1996, 781). Ioannidis (2005a, 218) asserts both that in a certain percentage of M-studies in his sample, subsequent research provided evidence either against the existence of the claimed association or effect, or against the stated magnitude of the claimed association or effect, and that his sample is representative of M-studies. Although a more exact determination of the disconfirmation rate of the studies in Ioannidis' sample would involve accounting for right-censoring and establishing a new metric for disconfirmation, I assume that the core of Ioannidis' message would survive this process: 'contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes'; therefore 'evidence from recent trials, no mater [sic] how impressive, should be interpreted with caution, when only one trial is available' (Ioannidis 2005a, 218, 225).

These conclusions can be formulated as a local pessimistic meta-induction in the field of medicine (PMI-M):[10]

E1 41% of the associative or causal claims made by M-studies in the sample were inconsistent with the results of subsequent published studies either (1) because the later studies provided evidence against the existence of the association or effect; or (2) because the later

---

of effect rather than effect size. I do not think this detail matters for my point here.

[10] I am grateful to an anonymous reviewer for recommending a fuller exposition of PMI-M and its differences from global PMI.

studies provided evidence that the magnitude of the association or effect was significantly different.

E2 Therefore, we can expect approximately 41% of the associative and causal claims made by M-studies to be inconsistent with subsequent published studies.

E3 When we learn that the results of a study are inconsistent with the results of other published studies, we should reduce our credence in the claims made by that study.

E4 Therefore, we should reduce our credence in the associative and causal claims made by M-studies.

Note that initial subsequent studies are M-studies. It is only after there is significantly different information in the form of greater statistical power or multiple subsequent studies that we can start to form reasoned judgments about where the truth lies (Ioannidis 2005a, 224); studies based on this significantly different information are not M-studies. One result of this is that PMI-M generates a mixture of modest and strong pessimism. Relative to a particular question under investigation—e.g., the relationship between Vitamin E and coronary disease—the pessimism is modest in that skepticism is generated about current but not (many) future results. Relative to M-studies, the pessimism is strong in that skepticism is generated about these studies now and in the future.

There are other important differences between PMI-M and global PMI. First, PMI-M concerns itself with particular associative and causal claims, not theories-writ-large. By itself, it will not throw doubt on, say, the germ theory of disease. Second, PMI-M does not apply to findings produced in relevantly different research environments. For example, it does not license skepticism about studies performed before modern computing power allowed statistical analyses of massive scope. Third, the sample of global PMI contains theories that were broadly accepted and empirically successful over a substantial period of time. Their success often included making novel predictions as well as describing the phenomena. By contrast, PMI-M deals with associative and causal claims which were not empirically descriptive for very long and which did not make novel predictions. In some cases, there was confirmation from subsequent studies, as with Vitamin E and coronary disease; in other cases, the success of the claims did not extend beyond the data set used to generate them.

These differences support the claim (to be defended later) that local pessimistic meta-inductions are limited in their ability to support global pessimism. But these differences do not render PMI-M unimportant. PMI-M concerns itself with claims made by the initial studies of questions in medicine that are published in top journals and highly cited. There is a prima facie strong case for taking such claims seriously. Moreover, it sometimes happens that, because of limited resources or ethical concerns, no later studies are conducted (Ioannidis 2005a, 225). In these cases, PMI-M helps guide our final considered judgments on the associative or causal claims in question.

There are some clear advantages of PMI-M over PMI, for example:

1. Taking published research findings to be the individuals in the sample and population avoids the problem of how to individuate theories (Lange 2002, 283; Magnus 2010, 807).

2. Taking citations to be a proxy for acceptance allows for a quantitative measure of the threshold of acceptance (or at least influence) of a theory in a field.

3. Since there were no landmark changes in method during the time period studied (1990-2003), the failure of the statistical criterion of randomness is mitigated (see section 3).

4. It does not illicitly assume that there is no attrition because it applies only to initial findings concerning a given question.

5. Ioannidis' attempts to identify the sources of the errors he discovers support the projection to other cases in which identified sources of error are present, making the pessimistic inference more compelling than one based on negative results alone.

The case can be made that PMI-M is self-undermining. Ioannidis has published many research findings using statistical methods which critique the accuracy of published research findings grounded in statistical methods. Moreover, he identifies studies with small sample sizes and nonrandomized studies as being particularly suspect. The median sample size for studies inconsistent with later studies was 624, while the median sample size for confirmed studies was 2165; his own nonrandomized study has a sample size of 45 (Ioannidis 2005a, 220, 224). It looks as if E4 would have us reduce our confidence in the inference from E1 to E2, and hence undercut our reasons for holding E4.

However, there are two ways PMI-M can avoid the self-undermining problem. The first is to differentiate Ioannidis' studies from M-studies in non-ad-hoc fashion. This can be done by reference to the number of variables involved. Ioannidis (2005b, 698) has identified the phenomenon of chasing statistical significance by extensive analysis of a great number of variables as contributing to the high refutation rate of published studies. Ioannidis' own studies do not look at a great number of variables, limiting their focus to refutation status and a few others such as number of citations. Another basis for contending that Ioannidis' studies are not M-studies relies upon Norton's material theory of induction. The network of causal relationships among health interventions and outcomes in humans is very different from the conceptual, mathematical, and causal relationships underpinning the mechanics of medical research. This makes the kinds of projective inferences in the two cases distinct.

The second way for PMI-M to avoid the self-undermining problem is to employ Hobbs' reductio strategy (section 4). I rejected this strategy for global PMI because the conclusion that scientific methods over time have generated inconsistent results is not sufficient to support modest pessimism. Global PMI applies to a variety of methods over centuries of scientific inquiry. It is not implausible to suppose that the false theories are concentrated among the earlier theories. This supposition blocks the inference to modest pessimism. But PMI-M deals with studies within a narrow range of time and methods. Let us grant that Ioannidis' studies are M-studies. The conclusion of the reductio version of PMI-M is that M-studies have generated inconsistent results. This is sufficient to support skepticism about claims made by M-studies because it is implausible to suppose that there is a relevant difference between earlier and later studies using similar methodologies in a relatively brief time period. Whether or not Ioannidis' studies are M-studies, the self-undermining problem is avoided. Furthermore, even if the above arguments fail to persuade, it is at least possible for a local PMI to avoid being self-undermining, since Ioannidis' methods can be applied to a clearly distinct field of inquiry, say, paleontology.

Because of the advantages of local over global PMIs, a global pessimist may wish to carry out local PMIs in (most) every domain of science. In developing this idea, the pessimist can build on the advantageous features of Ioannidis' methodology. For example, Ioannidis solves the worries about individuation and acceptance through the use of journal publications and citation indexes. If some of the local PMIs draw on the earlier history of science, there will not be the neat individuation — with database access! — available to Ioannidis. However, it would be possible to use historians of science as consultants. In a specific field, these historians could be surveyed regarding: which theories were accepted when, for how long, and how widely throughout the scientific community; the methods by which these theories were generated; and the time, if any, at which the refutation of these theories became accepted.[11]

Another refinement of method would partly address the issue of the non-random selection of the inductive basis. Just as survival studies are able to use partially censored data — e.g., a subject who is diagnosed with cancer and survives until the end of the study — in their determinations of mortality rates, a pessimistic meta-induction could include in the sample theories which have not yet been refuted with appropriate use of regression analysis (see, e.g., Hosmer, Jr. and Lemeshow 1999). This would avoid the problem of current theories being disqualified from counting in the inductive basis because they can't be assumed to be true (lest the question be begged by the optimist): these theories can in fact be included by using standard techniques to deal with right-censoring.

These would be significant improvements for PMI. Looking at multiple areas of science with different methods, using advanced statistical techniques appropriate to the complexity of the data, and using survival analysis techniques to deal with the problem of right-censoring would yield interesting results. Surely, different domains of science would have varying degrees of success and failure, making for a more nuanced, potentially more accurate conclusion.

However, the results of such a process, even if quite negative, would fall short of global pessimism. Hobbs' reductio maneuver avoids the self-undermining problem only when the

[11] Mizrahi (2012) sketches one way PMI might look if it were to use random sampling. He samples theories from those labeled as such in texts like the Oxford Dictionary of Biology. The motivation to develop PMI in a more sophisticated manner is admirable, but the result that most theories in his sample remain unrefuted does not tell against PMI because three of his four sources are not primarily historical in their focus. The agendas of scientific reference works may not include presenting theories accepted in the past as comprehensively as currently accepted theories, and so the latter may be overrepresented.

method in question is narrowly construed. There could not be a series of local PMIs, all of which function by reductio, that together raise doubt about all methods used by science, because there are many scientific methods inappropriate to the scientific study of the reliability of science. To raise doubts about the results of these methods, there will have to be local PMIs that are not in reductio form. No matter how bleak the verdict of a series of local PMIs, there will be part of science — viz., that part involved in the evaluation of those scientific methods that cannot evaluate themselves — that is not thrown into question. Other relevantly similar parts of science will also not be thrown into doubt.

6. Conclusion

In recent philosophy of science, there has been a trend towards the consideration of local rather than global issues (see, e.g., Stanford 2001, Norton 2003, Magnus and Callender 2004, Turner 2004). I have contended that, in the case of PMI, this trend is welcome. Global pessimism cannot be motivated either by global PMI or by a conjunction of successful local PMIs. Those who employ pessimistic reasoning from the history of science must acknowledge the reliability of the scientific study of the accuracy of scientific methods that cannot evaluate themselves, as well as (for consistency's sake) the reliability of other applications of these methods.

PMI-M is local in virtue of being confined to medical findings produced by specified methods used within a limited time frame in the history of medical research, and also in virtue of not projecting the results to cases where there is substantially more aggregated evidence. Even within these limited bounds, there is pressure to evaluate methods at an even more local level: 'at the level of the field, disease, mechanism, or question' (Goodman and Greenland 2007, 15). Those engaged in carrying out careful local PMIs will likely become more fine-grained in their focus over time as they seek to isolate features of the methods they evaluate that contribute to or detract from those methods' reliability.

By uncovering features that lead to unreliability, the very project of engaging in local PMIs can lead to the improvement of scientific methods (Faust and Meehl 2002). Even if it is found that scientific methods in many domains are inaccurate, optimists do not have to claim that the scientific process is accurate in the way that an archer is accurate. The goal is not never to be wrong – it is to get the right answer once in each domain and remember that it is right. One of the great forces of science is its historical record of past experiments, dead-ends, and breakthroughs. For all that global PMI has shown, as evidence accumulates over time, as more and more initially plausible theories are discarded, and as the standards for plausibility are raised, the force of attrition makes it more likely that we are on the right track, even if we have a considerable track record of failure. These factors are present in local PMIs as well, so even when we reach a pessimistic local result, we should be cautious in projecting past results onto present and future outcomes.

Acknowledgements

I am grateful for helpful feedback from Ludwig Fahrbach, Nicholas Horton, Brian Kierland, Bradley Monton, David Soergel, Aaron Smith, Bob Snyder, the philosophy department at Boise State University, and the anonymous referees of this paper.

Note on contributor

Samuel Ruhmkorff (Ph.D. Michigan, 2001) is Associate Professor of Philosophy at Bard College at Simon's Rock. His research focuses on scientific realism and antirealism, probabilistic epistemology, and religious pluralism.

References

Balk, Ethan M. et al. 2002. "Correlation of Quality Measures with Estimates of Treatment Effect in Meta-analyses of Randomized Controlled Trials." *Journal of the American Medical Association* 287: 2973-2982.

Bishop, Michael A. 2003. "The Pessimistic Induction, The Flight to Reference and the Metaphysical Zoo." *International Studies in the Philosophy of Science* 17: 161-178.

Chakravartty, Anjan. 2004. "Structuralism as a Form of Scientific Realism." *International Studies in the Philosophy of Science* 18: 151-171.

Chang, Hasok. 2003. "Preservative Realism and its Discontents: Revisiting Caloric." *Philosophy of Science* 70: 902-12.

Devitt, Michael. 2011. "Are Unconceived Alternatives a Problem for Scientific Realism?" *Journal for General Philosophy of Science* 42: 285-293.

Doppelt, Gerald. 2007. "Reconstructing Scientific Realism to Rebut the Pessimistic Meta-Induction." *Philosophy of Science* 74: 96-118.

Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Elsamahi, Mohamed. 2005. "A Critique of Localized Realism." *Philosophy of Science* 72: 1350-1360.

Enfield, Patrick. 2008. Review of *Exceeding Our Grasp*, P. Kyle Stanford. *British Journal for the Philosophy of Science.* 59: 881-895.

Fahrbach, Ludwig. 2011a."How the Growth of Science Ends Theory Change." *Synthese* 180: 139-155.

——. 2011b. "Theory Change and Degree of Success." *Philosophy of Science* 78: 1283-1292.

Faust, David and Paul E. Meehl. 2002. "Using Meta-Scientific Studies to Clarify or Resolve Questions in the Philosophy and History of Science." *Philosophy of Science* 69: S185-S196.

Forber, Patrick. 2008. "Forever beyond Our Grasp?" Review of Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives, by P. Kyle Stanford. *Biology and Philosophy* 23:135–41.

——. 2011. "Reconceiving Eliminative Inference." *Philosophy of Science* 78: 185-208.

Godfrey-Smith, Peter. 2003. *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: The University of Chicago Press.

——. 2008. "Recurrent Transient Underdetermination and the Glass Half Full." *Philosophical Studies* 137:141–48.

Goodman, Steven and Sander Greenland. 2007. "Assessing the Unreliability of the Medical Literature: A Response to 'Most Published Research Findings are False'." *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working Paper 135. Accessed at http://biostats.bepress.com/jhubiostat/paper135 on July 5, 2012.

Hardin, Clyde L. and Alexander Rosenberg. 1982. "In Defense of Convergent Realism." *Philosophy of Science* 49: 604-15.

Hobbs, Jesse. 1994. "A Limited Defense of the Pessimistic Induction." *British Journal for the Philosophy of Science* 45: 171-191.

Hosmer, Jr., David W. and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data.* New York: John Wiley & Sons.

Ioannidis, John P.A. et al. 2001. "Comparison of Evidence of Treatment Effects in Randomized and Nonrandomized Studies." *Journal of the American Medical Association* 286: 821-830.

Ioannidis John P. A., and Thomas A. Trikalinos. 2005. "Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials." *Journal of Clinical Epidemiology* 58: 543-549.

Ioannidis, John P. A. 2005a. "Contradicted and Clinically Stronger Effects in Highly Cited Clinical Research." *Journal of the American Medical Association* 294: 218-228.

——. 2005b. "Why Most Published Research Findings are False." *PLoS Medicine* 2: 696-701

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusion*. New York: Oxford University Press.

Lange, Marc. 2002. "Baseball, Pessimistic Inductions, and the Turnover Fallacy." *Analysis* 62: 281-285.

Laudan, Larry. 1981. "A Confutation of Convergent Realism." *Philosophy of Science* 48: 19-49.

Leplin, Jarrett. 1997. *A Novel Defense of Scientific Realism*. Oxford: Oxford University Press.

Lewis, Peter J. 2001. "Why the Pessimistic Induction is a Fallacy." *Synthese* 129: 371-380.

Lipton, Peter and John Worrall. 2000. "Tracking Track Records." *Proceedings of the Aristotelian Society, Supplemental Volumes*: 74: 179-235.

Lyons, Timothy D. 2006. "Scientific Realism and the Stratagema de Divide et Impera." *British Journal for the Philosophy of Science* 57: 537-560.

Magnus, P.D. 2010. "Inductions, Red Herrings, and the Best Explanation for the Mixed Record of Science." *British Journal for the Philosophy of Science* 61: 803-819.

Magnus, P.D., and Craig Callender. 2004. "Realist Ennui and the Base Rate Fallacy." *Philosophy of Science* 71: 320-338.

Mizrahi, Moti. 2012. "The Pessimistic Induction: A Bad Argument Gone Too Far." *Synthese.* doi:10.1007/s11229-012-0138-3.

Newman, Mark. 2005. "Ramsey Sentence Realism as an Answer to the Pessimistic Meta-Induction." *Philosophy of Science* 72: 1373-1384.

Norton, John D. 2003. "A Material Theory of Induction." *Philosophy of Science* 70: 647-670.

Park, Seungbae. 2011. "A Confutation of the Pessimistic Induction." *Journal for General Philosophy of Science* 42: 75-84.

Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth.* London: Routledge.

Putnam, Hilary. 1975. *Philosophical Papers*, Vol. 1: *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.

-----. 1978. *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Ritchie, Jack. 2008. "Structural Realism and Davidson." *Synthese* 162: 85-100.

Ruhmkorff, Samuel. 2011. "Some Difficulties for the Problem of Unconceived Alternatives." *Philosophy of Science* 78: 875-886.

Saatsi, Juha T. 2005. "On the Pessimistic Induction and Two Fallacies." *Philosophy of Science* 72: 1088-1098.

Shimony, Abner. 1970. "Scientific Inference." In *The Nature and Function of Scientific Theories: Essays in Contemporary Science and Philosophy*, ed. Robert G. Colodny, 79–172. Pittsburgh: University of Pittsburgh Press.

Stanford, P. Kyle. 2001. “Refusing the Devil’s Bargain: What Kind of Underdetermination Should We Take Seriously?” *Philosophy of Science* 68: S1-S12.

-----. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Stephens, Nigel G., Ann Parsons, Peter M. Schofield, Frank Kelly, Kevin Cheeseman, Malcolm J. Mitchinson, Morris J. Brown. 1996. “Randomised Controlled Trial of Vitamin E in Patients with Coronary Disease: Cambridge Heart Antioxidant Study (CHAOS).” *The Lancet* 347: 781-786.

Tatsioni, Athina, Nikolaos G. Bonitsis, and John P. A. Ioannidis. 2007. “Persistence of Contradicted Claims in the Literature.” *Journal of the American Medical Association* 298: 2517-2526.

Turner, Derek. 2004. “Local Underdetermination in Historical Science.” *Philosophy of Science* 72: 209-230.

Turner, Derek. 2005. “Misleading Observable Analogues in Paleontology.” *Studies in History and Philosophy of Science* 36: 175-183.

Worrall, John. 1994. "How to Remain Reasonably Optimistic: Scientific Realism and the 'Luminiferous Ether'." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994: 334-342.

Wray, K. Brad. 2013. "Success and Truth in the Realism/Anti-realism Debate." *Synthese* 190: 1719-1729.