*Synthese*. Final version available at Springer via http://dx.doi.org/10.1007/s11229-015-0947-2

**Unconceived alternatives and the cathedral problem**

**Abstract**

Kyle Stanford claims we have historical evidence that there likely are plausible unconceived alternatives in fundamental domains of science, and thus evidence that our best theories in these domains are probably false. Accordingly, we should adopt a form of instrumentalism. Elsewhere, I have argued that in fact we do not have historical evidence for the existence of plausible unconceived alternatives in particular domains of science, and that the main challenge to scientific realism is rather to provide evidence that there are likely not plausible unconceived alternatives. In the present paper, I contend that we may come to have such evidence in the long run of science. I then investigate the epistemic consequences of the claim that we presently do not have evidence for or against the existence of plausible unconceived alternatives, but that in the future we may come to have evidence against the existence of plausible unconceived alternatives. I argue there are prima facie reasons to endorse a form of voluntarism in this situation according to which scientists and others may rationally be more optimistic or more pessimistic about the truth of our best theories, on the grounds that the widespread acceptance of an obligation to be an instrumentalist threatens to disrupt the proper functioning of science, in part because the domain of application of the problem of unconceived alternatives is unclear.

KEYWORDS: unconceived alternatives, scientific realism, instrumentalism

## 1. Introduction

The problem of unconceived alternatives, as defended by Kyle Stanford (2006), contends that there are very likely unconceived alternatives to our best theories in fundamental domains of science that are at least as plausible as these theories given our present state of evidence. Since eliminative inference proceeds by our eliminating all but one of the plausible explanations of a given phenomenon, our recognition of the likely existence of plausible unconceived alternatives to our best theories undermines our ability to use eliminative inference to arrive legitimately at a high credence in our best theories. Accordingly, we should not believe these theories to be true, but rather adopt an instrumentalist attitude towards them.

In this paper, I begin by presenting a Bayesian model of eliminative inference. I then argue that, even if we do not currently have evidence that there are likely not plausible unconceived alternatives to our best theories about fundamental domains of nature, we collectively may very well come to have this evidence in the future—though not in our lifetimes. I contend that this possible scenario poses an important epistemic quandary that I label *the cathedral problem*: what should we believe about our best scientific theories if their validation against the possibility of plausible unconceived alternatives—if any such validation is in the offing—will come long after we are gone? I tentatively defend a form of voluntarism according to which scientists and others may rationally be more optimistic or more pessimistic about the truth of our best theories, on the grounds that the widespread acceptance of an obligation to be an instrumentalist threatens to disrupt the proper functioning of science. In so doing, I argue that the domain of application of the problem of unconceived alternatives is undefined by its own lights.

## 2. A Bayesian model of eliminative inference

The problem of unconceived alternatives is a challenge to eliminative inference. Eliminative inference proceeds by the elimination of available theories until one remains. The remaining theory is then inferred to be true. If we have reason to think there are plausible unconceived alternatives to the theories available to us in a given domain, eliminative inferences in that domain will not be warranted.

Since our elimination of theories is usually not an all-or-nothing affair, it will be helpful to sketch a probabilistic model of eliminative inference. We assign probabilities to available hypotheses and the catchall hypothesis (Hc)—the hypothesis that all available theories are false (Shimony 1970, 95). In eliminative inference, we learn evidence that lowers our probability for an available theory. The resulting probability freed up goes to the remaining theories, including the catchall if nonzero, through conditionalization. When this results in a sufficiently high probability being assigned to the best available theory, we can be said to believe this theory based on eliminative inference.

If Stanford is correct that we have reason to think there are likely plausible unconceived alternatives to our best theories about fundamental domains of nature, the mechanism of eliminative inference is not undermined. What is undermined is the idea that this mechanism will ever result in probability assignments to our best theories sufficient for full belief. If there are likely plausible unconceived alternatives, P(Hc) should be high. At a minimum, it should be approximately the value we assign to our best theory (Hb), because there is likely to be at least one unconceived theory as plausible as our best theory. The maximum probability after updating for an available hypothesis Hi without evidence against the catchall is $P(H_i)/((P(H_i)+P(Hc))$. If $P(Hb) \approx P(Hc)$ on the grounds that there is likely an unconceived alternative to our best theory at least as plausible as Hb, this value can never be greater than $\approx 1/2$. Moreover, we cannot get

evidence against the catchall, because (if Stanford is right) we have reason to think that this hypothesis subsumes at least one theory that is at least as plausible as our best theory in that domain, and it is hard to see how we could get evidence against an unarticulated theory. In sum, there are two ways the problem of unconceived alternatives prevents eliminative inference resulting in a high probability assignment to a theory: (1) it has us assign a high prior to Hc; (2) it prevents us from learning anything that decreases significantly the probability of Hc.

There is no clear answer about what we should do when we actually discover a previously unconceived theory. We cannot conditionalize because the relevant conditional probabilities do not exist. John Earman (1992, 195ff) distinguishes between "shaving off" (where the previously unconceived theory gets all of its probability from Hc) and a more radical resetting of all of one's probabilities. Both of these responses would allow in principle the assignment of a high P(Hb). However, according to the problem of unconceived alternatives, neither would be warranted in practice. Discovering one plausible unconceived alternative does not give us reason to think that there are not more. According to the problem of unconceived alternatives, whether we shave off from Hc or reset our priors, P(Hc) still should be sufficiently high to block high confidence in our best theory. The problem of unconceived alternatives may even suggest that, upon finding a new theory superior to existing alternatives, we should *increase* P(Hc) to take into account the inductive support given to the existence of plausible unconceived alternatives. Even realists might think that this is what thinkers in the early days of science should have done as they became aware of the phenomenon of plausible unconceived alternatives.

## 3. The cathedral problem

Stanford (2006) infers the likely existence of plausible unconceived alternatives in fundamental domains of nature on the basis of detailed historical investigations of episodes in which scientists made eliminative inferences to high degrees of belief in the theories they were defending when they had not exhausted the possibility space of plausible theories. The evidence that they had not exhausted this possibility space is that other scientists proposed and took seriously theories that the previously mentioned scientists had not conceived, to the extent of preferring these theories to their own.

Elsewhere (reference omitted), I argue that Stanford's evidence supports at most the claim that humans are not good at noticing plausible unconceived alternatives when they exist, not the conclusion that there likely are such theories, nor the stronger conclusion that there will always be such theories. I conclude that the main challenge to the scientific realist from the possibility of plausible unconceived alternatives is to provide evidence that there are likely not such alternatives.

Here I will argue that, even if it is true that the problem of unconceived alternatives makes it difficult for realists to provide evidence *now* that there are likely not plausible unconceived alternatives to our best theories, it does not follow that realists will not be able to provide such evidence *in the future*. Stanford must acknowledge that the question of whether fundamental scientific theories are vulnerable to the problem of unconceived alternatives is local and empirical. As Stanford admits, "the naturalistic spirit of our inquiry requires that we remain open to the possibility that we will find special reasons to doubt that some particular scientific theories remain vulnerable to the problem of unconceived alternatives that seems to attend our scientific theorizing quite generally" (2006, 203). Naturalism also suggests it is possible we will uncover evidence that applies *generally* to our scientific theorizing, giving us reason to think we

have likely exhausted plausible alternatives to our best theories concerning fundamental domains of nature. This evidence might take the form of long-term stability in the sciences combined with vigorous and dynamic attempts to generate plausible alternatives to our best theories, coherence among theories from different domains, and reasons to think that plausible unconceived alternatives would have been conceived if they existed (lack of conservatism, funding of non-dominant views, etc.—for more, see Stanford, this volume). Whether such long-term stability will come to pass depends on things those alive today will never know, such as the shape and existence of future human society, the resources devoted to scientific inquiry, the degree of freedom of scientific thinking, and the friendliness of the universe to our epistemic projects.

Karl Popper wrote: "All work in science is directed towards the growth of objective knowledge. We are workers who are adding to the growth of objective knowledge as masons to a cathedral" (1972, 121). The masons beginning work on a cathedral know they will not live to see the completion of the building. Similarly, scientists – and those whose beliefs are in part dependent on those of scientists – will not live to see the directions future science will take. This feature of science presents what I call "the cathedral problem": what are our epistemic obligations in situations in which the evidence for the accuracy of an inference can lag generations behind the inference itself?

I will consider two responses to this question: the *evidentialist* and the *voluntarist*. The evidentialist response holds that we should never believe anything beyond our current evidence. If we do not have evidence that there are unlikely to be plausible unconceived alternatives to the best available theories about fundamental domains of nature, then evidentialism holds that scientists ought not to assign a high P(¬Hc). Voluntarism as I conceive it is modeled after Bas van Fraassen's view about epistemic obligation (1989). Voluntarism holds that it is epistemically

permissible to assign a range of values to ¬Hc despite the limitations of our current evidence. This needn't mean complete license. Voluntarism is compatible with holding that one should not assign P(¬Hc) = 1, or P(¬Hc) = 0, or even that one should not assign a sharp probability to ¬Hc, rather assigning it a vague range. What is essential to voluntarism is that it is epistemically permissible to assign probabilities to ¬Hc which encompass relatively low values, and permissible to assign probabilities to ¬Hc which encompass relatively high values, as far as the possibility of unconceived alternatives is concerned. (There may, of course, be other forms of evidence that obligate the belief that, in a specific domain, no existing theory is true.)

Stanford takes evidentialism and the consequent obligatory instrumentalism to be the appropriate response to being in situations characterized by the cathedral problem. I will call such instrumentalism 'evidentialist instrumentalism' to distinguish it from the position that adopts an instrumentalist attitude but does not hold such an attitude to be obligatory. Scientific realists who disagree with Stanford about whether we have evidence that the existence of plausible unconceived alternatives to our best theories is unlikely may agree with him about what we should believe when we lack such evidence. Thus there may be scientific realists who are evidentialists and hold that early scientists should have been instrumentalists about their best theories. In the next section, I raise concerns for both Stanford and evidentialist realists by citing a number of potential negative consequences of evidentialism when it is combined with the claim that in fundamental domains of nature, we do not currently have evidence that there are not likely plausible alternatives to our best theories. I conclude that it is not yet established that evidentialist instrumentalism is the correct response to the cathedral problem.

## 4. The difficulty of implementing evidentialist instrumentalism

I contend that even if we grant Stanford that our current evidence does not give us reasons to assign a high P(¬Hc), we should be cautious in assuming evidentialism and consequently endorsing obligatory instrumentalism. I have two reasons. The first is that Stanford's instrumentalism based on the problem of unconceived alternatives is extremely difficult, if not practically impossible, to implement. The second reason is based on three potentially negative consequences of the widespread acceptance of obligatory instrumentalism among scientists.

My first reason for cautioning against evidentialist instrumentalism is that adopting an instrumentalist attitude based on the problem of unconceived alternatives is extremely difficult, because it is hard to figure out what one's beliefs should be on such an attitude. Instrumentalism counsels a kind of agnosticism about our best theories concerning fundamental domains of nature, but it does not counsel agnosticism through-and-through. Rather, we should put a high credence in the observable consequences of our best theories (Stanford 2006, 200–204). Which consequences are these? Roughly, the consequences that are independent of theories regarding which we are instrumentalists. But which theories are these? Unfortunately, there are no marks by which we can distinguish the theories about which we should be instrumentalists and the theories about which we should be realists. In other words, it is unclear to which domains of science the problem of unconceived alternatives applies, and to which it does not. This means that the intellectual task of being agnostic in the right way will involve significant cognitive resources better devoted to the day-to-day practice of science.

Stanford's instrumentalism is selective: it is aimed only at the use of eliminative inference in theoretical contexts in which "*we have good reason to doubt that we can exhaust the space of alternative possibilities*" (Stanford 2006, 37). There are instances of scientific reasoning – and of eliminative inference – which are not susceptible to the problem of unconceived alternatives. It is

important to Stanford that his view not collapse into more general skepticism, or even a general skepticism about the results of science. Yet his view still counts as antirealism because he thinks that the conclusions of the targeted uses of eliminative inference are "virtually all of those fundamental theories concerning remote domains of nature that lie at the heart of the contemporary scientific conception of the natural world" (Stanford 2006, 37).

But if the problem of unconceived alternatives is properly motivated, its own domain of application is uncertain. If humans are bad at knowing when all plausible alternatives have been considered, how can we know which instances of eliminative inference are affected by the problem of unconceived alternatives? And if we don't know which eliminative inferences are affected by the problem of unconceived alternatives, we don't have evidence that in a particular instance there likely are plausible unconceived alternatives.

Stanford states:

> I remain cautiously optimistic that there will be something general to say about our systematic vulnerability to the problem of unconceived alternatives (and we had better hope that there is, since our intuitions in particular cases concerning whether we are vulnerable to the problem or not have turned out to be so spectacularly unreliable. (2006, 203)

As I see it, there are two candidates for marks by which we can distinguish targeted and non-targeted instances of eliminative inference that are suggested by Stanford's writing:

1. *Appearances*. Stanford contrasts situations in which there appears to be an inchoate possibility space and those in which there is a clearly stated, exhaustive partition (2006, 133, 32). The problem of unconceived alternatives applies to the former, but not the latter.

2. *History*. Stanford suggests that the extent of the application of the problem of unconceived alternatives will be revealed by historical investigations in various domains of science (2006, 37, 46).

These considerations seem to work together. The appearances of the possibility space direct the investment of our efforts at historical investigation. Historical investigation reveals where we in fact have reason to doubt, and give us a more detailed understanding of the openness of the possibility space. The appearances of the possibility space give us support for interpreting past failures as defeaters for current and future inferences. They also give us reason to consider seriously a more general scientific antirealism in advance of historical investigations (2006, 37).

Despite this mutual reinforcement, I take issue with both of these ways of distinguishing between targeted and non-targeted instances of eliminative inference. Take the appearances of the possibility space first. Stanford gives us a number of situations in which scientists took themselves to be reasonably sure that they had considered all plausible alternatives, but were wrong. These situations give us reason to think that there are possibility spaces which appear to be fully explored, but are not. This means that we can't use the appearance of the possibility space to us as a mark of whether inferences involving that possibility space are targeted or not. The whole point of the problem of unconceived alternatives is that we cannot trust ourselves in our judgments concerning whether a given possibility space has been exhausted.

There are two ways we might end up being mistaken about having exhausted a given possibility space. The first is by failing to notice that the disjunction of available theories does not exhaust that space. This happens frequently in the historical cases that Stanford considers (Stanford 2006, 67f, 86-90, 111-126). It is also illustrated by Ian Hacking's (2000) contention that scientists' conceptual frameworks shape their understanding of theoretical possibilities,

making the results of scientific inference not inevitable, but rather dependent on historical contingencies.

The second way of being mistaken about having exhausted a given possibility space happens when the partition of the space is in fact exhaustive, but is composed of some theories whose high level of generality obscures details relevant to theory comparison, preventing a full scientific evaluation. I will give two examples of the second situation. First, the heliocentric hypothesis states that the complete explanation for the apparent motion of the sun through the heavens is the rotation of the earth about its axis. The heliocentric hypothesis has specific observable consequences when combined with appropriate auxiliary hypotheses, e.g., stellar parallax. Now consider the trivially generated partition {H, ¬H}. ¬H here functions as the catchall: it covers all ways in which the other hypothesis might be false. Because it is at a high level of generality, it does not have specific observable consequences. On some ways that ¬H would be true, there would be no stellar parallax. On other ways, there would be stellar parallax, as in scenarios where the apparent motion of the sun is explained by the combination of the motion of the earth and the sun. Even in this relatively simple example, there is not a straightforward way to evaluate the catchall. Because we cannot evaluate the catchall as a scientific theory, our success in logically exhausting the space does not mean that we get epistemic credit for doing so. My second example comes from the study of human origins. In recent decades there has been a bitterly divided contest between the theory of recent African origin (humans evolved in Africa and migrated to other parts of the globe) and multiregionalism (humans evolved independently in different parts of the globe and interbred with different archaic populations) (Stringer 2012, 264). As more genetic data have become available, the debate has transformed into competition between increasingly complex hypotheses proposing

different amounts of hybridization between modern humans and archaic populations and the assimilation of archaic populations into different regional groups of modern humans. The neat logical partition (humans evolved in Africa but not elsewhere ∨ humans evolved elsewhere but not in Africa ∨ humans evolved in Africa and elsewhere ∨ humans didn't evolve) has been replaced by a spectrum of theories positing different degrees of contributions from different populations of archaic humans. This is another scenario where we have found that we cannot make a fair scientific comparison of theories that neatly partition a given possibility space.

This illustrates a fundamental aspect of the problem of unconceived alternatives as Stanford himself characterizes it: the catchall is not a scientific hypothesis, and neither it nor other kinds of general descriptions of hypotheses are conducive to evaluation in the way that available scientific hypotheses are (see Stanford 2006, 53; 2009, 266f). If we are faced with a partition {T1, T2, T3, T4}, even if it is logically exhaustive, it may be at the cost of a high level of generality which prevents us from making scientific comparisons among the theories. This means that we cannot use our sense of the completeness of our exploration of a possibility space as a mark of our full exploration of that space. And this means that the quest for a coherent instrumentalist attitude towards our theories is exceedingly difficult, if not practically impossible.

Stanford might claim that, at the least, we can infer that, if there is the appearance of an inchoate possibility space among the theories in a given domain, the problem of unconceived alternatives applies to that domain. Yet even if this claim is granted, our inability to eliminate certain domains from the problem of unconceived alternatives would make us uncertain about how to adopt the instrumentalism that Stanford claims falls out of it, as we would be uncertain what the observable consequences of our best theories are.

The second possible criterion, that historical investigations will reveal the domains to which the problem of unconceived alternatives applies and does not apply, is also problematic. Stanford's historical research uncovers, for example, failures in the ability of 19th century scientists to conceive of plausible alternatives to extant theories in the science of heredity. Does this research raise skeptical doubts specifically for modern genetic theory (as opposed, say, to chemical theory)? How can that be when modern genetic theory is so different from the 19th century theories? It is hard to know how the results of the historical investigations should generalize. Therefore, this criterion cannot serve as a marker for identifying when we should be instrumentalists about a theory.

It might be objected that historical research uncovers features of scientific practice that correlated with past failures to consider plausible unconceived alternatives; and that the problem of unconceived alternatives applies to inferences generated from science characterized by similar features. Examples of such features might be mechanisms of peer review, funding, and graduate training (for more on how features of scientific practice might inhibit or encourage the search for plausible unconceived alternatives, see Stanford, this volume, and Kidd, this volume). On this line of thought, we don't project from past failures of eliminative inference to present and future failures of eliminative inferences based on content area, but rather based on variables characterizing the structure of scientific inquiry.

John Ioannidis has made this kind of argument in the field of medical research (2005a; 2005b). He has identified factors that contribute to the high refutation and non-confirmation rate of the results of medical studies published in top journals. These factors include features of publication practices that lead to publication bias and time-lag bias, features of research practices including the use of non-randomized trials and small sample sizes, and global features of the

scientific community such as the existence of many research teams doing high-powered statistical analyses on thousands of variables, increasing the likelihood of spurious significant results.

Elsewhere (reference withheld), I have defended Ioannidis' use of this method as an appropriate means of generating skepticism about a field of science on the basis of historical failures, while arguing that it cannot result in global skepticism about the results of science. Since Stanford does not intended to generate global skepticism, my argument in that paper does not prevent him from making an Ioannidis-style defense of the problem of unconceived alternatives. Yet I do not think that Ioannidis' method can be used to defend this problem as it has been presented to date. Stanford has claimed that the problem of unconceived alternatives applies to scientific results concerning fundamental domains of nature without the careful identification of variables present in the historical cases which are also present in contemporary scientific research into fundamental domains of nature—other than that the scientists are human (2006, 45). Any attempt to characterize these variables will face Ludwig Fahrbach's (2011a; 2011b; this volume) challenge based on the distinctiveness of science as it has been practiced since 1950. While applying Ioannidis' method to the problem of unconceived alternatives is a promising direction, the magnitude of a full-blown application of this method to scientific inquiry concerning fundamental domains of nature is such that anyone who is skeptical about the results of science in fundamental domains of nature should be skeptical about the results of this application. Again, there is no imminent promise of a way of distinguishing between eliminative inferences that are targeted by the problem of unconceived alternatives and those that are not so targeted.

I conclude that we must be deeply uncertain about the domains of scientific inquiry to which the problem of unconceived alternatives applies. If this is correct, then even if we are able to decide voluntarily to adopt a broad epistemic attitude (Ratcliffe 2011), we cannot — or can only with great difficulty — adopt an instrumentalist attitude in the way required by Stanford. On Stanford's view, instrumentalism has us place a high credence in the observable consequences of the theories we accept. Which consequences count as observable depends on which theories we accept without an instrumentalist attitude, which in turn requires knowing which inferences are targeted by the problem of unconceived alternatives and which are not. At the least, sorting out one's attitudes towards various theories will be a difficult business without clear payoff, a business that will likely interfere with other scientific pursuits. At the most, this task will be impossible to implement, making a putative epistemic obligation to do so questionable.

## 5. Potential negative consequences of evidentialist instrumentalism

My second reason for wariness about evidentialist instrumentalism has to do with the consequences of widespread acceptance among scientists of evidentialist instrumentalism. Of course, people can have all sorts of concerns about the consequences of the behavior and attitudes of scientists. For example, one consequence of the advances in technology made by scientists has been a growing wave of extinctions and deleterious changes in climate. However, what I have in mind are epistemic consequences connected to the goals of scientists themselves *qua* scientists. Whether scientists primarily aim for knowledge, true theories, understanding, or useful technologies, I am concerned that widespread acceptance of evidentialist instrumentalism will result in a less efficient mechanism for achieving these goals.

Stanford claims there is very little practical difference between having instrumentalist and realist attitudes towards our best theories. The sole difference he cites is the relative willingness of instrumentalists to entertain the search for new theories (2006, 211). This proposed difference seems right. Scientists less committed to the best available theory will likely spend more time reflecting on alternatives, or at least be more open to them when proposed by others. I will assume this is a good thing. Funnily enough, the widespread adoption of instrumentalism on the basis of the problem of unconceived alternatives would increase scientists' ability to uncover plausible unconceived alternatives! However, this positive consequence of evidentialist instrumentalism can also result from voluntarism, since presumably some scientists will choose to be instrumentalists. In addition, there are several potential negative consequences of evidentialist instrumentalism that threaten to outweigh this positive outcome. Because of these potential consequences, evidentialist instrumentalism is not an obvious implication of the establishment of the claim that we do not currently have evidence that eliminates the existence of plausible relevant alternatives in fundamental domains of nature.

The first potential negative consequence of the widespread acceptance of evidentialist instrumentalism is that adopting an instrumentalist attitude towards our theories would put some of us at a cognitive remove from the content of these theories. Consider a meditator who, through strict spiritual discipline, has managed to pierce the veil of Maya and understand that all duality is an illusion. This person may have extensive withholding of belief about the individuation of physical reality. However, when that person is ducking a snowball, or bird-watching, or cooking breakfast, it would be beneficial for her to think of snowballs and birds and eggs as beings real and distinct from her. Cognitive effort is required to maintain the distinction between contexts of practical manipulation and contexts of avowal. It is easier just to think of things as things in all

contexts rather than as things in some contexts and as-ifs in others. For some scientists, it will be more efficient to think of ¬Hc being likely true rather than considering an available theory 'the best that we have so far, likely false, but still the theory that is most worth investigating at this time' in contexts of avowal. Stanford acknowledges the superiority of scientific investigations using the conceptual resources of theories rather than their observational consequences (2006, 197f). I maintain that it is a real possibility that some people will be better scientists if they go in for full belief in the theories they accept without making a distinction between contexts of use and contexts of avowal, especially as the distinction between these contexts is scientifically inert except insofar as it supports the generation of previously unconceived plausible alternatives.

The second potential negative consequence of the widespread acceptance of evidentialist instrumentalism is motivational.[1] It is certainly possible for people to be motivated by abstract considerations such as 'I am dedicating my life to the study of this probably false theory which will almost certainly be surpassed by an as-yet unconceived theory because it is only by my and my colleagues' so doing that the unconceived successor theory may come to be discovered, which in turn will almost surely be surpassed…'; and perhaps more likely that some scientists are motivated by being part of a multigenerational epistemic project. However, I suspect that even more people will be motivated by considerations such as 'dedicating my life to the study of this theory is important because this theory is approximately true and I will learn much more about the world as I refine it'. Science requires extensive commitment and personal sacrifice. The motivation of scientists is connected with peer recognition, puzzle-solving, instrumental

[1] Arthur Fine (1996) attributes what he calls "motivational realism" to Einstein; Karen Darling (2003) attributes the same view to Duhem. According to motivational realism, a metaphysical realist outlook, while not cognitively defensible, is helpful or even necessary for engaging in the practice of science. The view I am gesturing at here has important commonalities with motivational realism, but involves a different brand of realism: the view that our best scientific theories are approximately true.

reliability, and maybe, to some degree, money. But it is implausible to think that there are not a significant number of scientists whose motivation is to find out something true about the world, and who would accordingly be less motivated were they to adopt thoroughgoing instrumentalist attitudes. For these reasons, it is plausible that there are some scientists who would be more dedicated, effective researchers if they assigned a relatively high P(¬Hc). The widespread acceptance of evidentialist instrumentalism would reduce the motivation of these scientists, or perhaps even prevent them from being scientists in the first place.

The third potential negative consequence of widespread evidentialist instrumentalism is that it is in tension with a disposition for credulity shared by many humans. I believe it is plausible that humans, when given a task to search for something that they were told was unlikely to be found in their data set, would have a tendency to raise the probability that it was in their data set if there were promising indicators of puzzle-solving success, even if it had been explicitly stated that these promising indicators would accompany their data set no matter what. If this is right, scientists meeting with empirical success would have a tendency to raise P(¬Hc) beyond their evidence. Correcting for this tendency would result in an epistemic discipline irrelevant to the practice of science, except insofar as it would encourage consideration of new theories. The cognitive effort thereby expended could be better spent on other tasks more germane to the practice of science. Furthermore, if there is a tendency for dedicated, effective researchers to raise P(¬Hc) beyond their evidence when engaged in successful, committed problem-solving, then there is a question of whether there is a strong epistemic obligation not to do so.

One objection to my claim that evidentialist instrumentalism has potentially negative consequences unexplored by Stanford is that humans are able to work within a conceptual

framework with a high degree of facility without having the associated beliefs. Examples are provided by literary scholars and Stanford's navigators. We can quickly adopt a pretense and discard it or modify it at will. Literary theorists who make progress on thorny issues of interpretation in a challenging novel are not thereby disposed to believe that the novel is veridical. And people often display a high degree of motivation in exploring fictional universes, as evidenced by the popularity of fan fiction.

An important difference between cases such as these and science is that propositions about the fixed earth and the intentions of Mr. Darcy are known to be false, and so navigators and literary scholars don't waste time thinking about whether or not they are true. But if we do not know what is true and false in a scenario, and are in part trying to figure out what we should believe is true or false, then these potential negative consequences may obtain. A navigator is less encumbered by speculations on the reality of the fixed earth than a literary scholar who discovered an apparent memoir would be by the question of whether or not it was written as a work of fiction. In addition, the navigator's use of the fixed earth theory represents a small fraction of her total theorizing activities. The potential negative motivational consequences on scientists of instrumentalism concerning theories regarding all fundamental domains of nature are on a different scale.

A second objection to my caution against evidentialist instrumentalism is that our epistemic obligations should not be influenced by the consequences of conforming to these obligations. Is it legitimate to condone belief in a high $P(\neg Hc)$ beyond people's evidence because it would result in greater instrumental reliability, especially in situations in which we may, for all we know, be using eliminative inference correctly, but where the full evidence for this correctness will not be available during our lifetime? This is an important question about

epistemic consequentialism beyond the scope of this paper. My goal here is to caution against a leap to evidentialism and subsequent evidentialist instrumentalism in situations characterized by the cathedral problem. Engendering a debate about the viability of epistemic consequentialism in conditions characterized by the cathedral problem is sufficient for this purpose.

A third objection is that it is illicit to hold a philosophical position – evidentialism – responsible for the potential negative consequences of its gaining widespread acceptance among scientists. It is a position primarily intended for philosophers of science. And widespread acceptance of evidentialist instrumentalism among philosophers of science is unlikely to undermine scientific practice. This maneuver would reduce the force of the above considerations at the cost of introducing a two-tiered epistemology, one for the enlightened folk (philosophers of science) and one for the common folk (scientists). This is unlikely to be attractive to many, especially as there are a number of scientists with an interest in the epistemology of science.

I have argued that evidentialist instrumentalism is daunting to implement, and that its widespread acceptance among scientists could potentially have negative consequences for scientific practice. These results caution against a rush to embrace evidentialist instrumentalism even if it is established that we do not currently have evidence that it is unlikely that there are plausible alternatives to our best theories in fundamental domains of nature. My discussion has focused on evidentialist instrumentalism as implemented by Stanford, but the caution remains in effect for realists as well. I have observed that even arch-realists might defend evidentialist instrumentalism historically, allowing that early scientists did not have the level of evidence in support of their best theories as current scientists do for their best theories. If these potential negative consequences of evidentialism are genuine, then realists too have reason to be cautious of evidentialism.

## 6. Conclusion

Evidentialism and voluntarism are different solutions to an agreed-upon problem: scientists are much better at the business of theory comparison than the business of evaluating P(Hc). Evidentialist instrumentalists seek to get us out of the latter business by defending agnosticism about our best theories. Voluntarists seek to achieve the same goal by permitting a range of permissible attitudes towards P(Hc), allowing scientists the latitude to have attitudes towards P(Hc) that are supportive of their work. I have argued that there are reasons to favor voluntarism over instrumentalism and that a full defense of instrumentalism must consider the potential harmful consequences of widespread evidentialist instrumentalism as well as the difficulty in adopting an instrumentalist attitude due to the lack of a clear distinction between instances of eliminative inference targeted by the problem of unconceived alternaives and those that are not targeted.

Among the most intriguing aspects of the problem of unconceived alternatives are the issues it raises about the connection between scientists' individual beliefs and the multi-generational collective process in which they are engaged. It is possible that scientists have successfully exhausted the space of plausible alternatives to their best theories, yet the historical evidence for this success will not be available in their lifetimes. Further research into the proper epistemic reaction to this situation would include the empirical study of the effects of agnostic attitudes on the work of scientists, and theoretical investigation into the merits of epistemic consequentialism in situations characterized by the cathedral problem.

**References**

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Darling, K. M. (2003). Motivational realism: The natural classification for Pierre Duhem. *Philosophy of Science* 70: 1125–1136.

Fahrbach, L. (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.

Fine, A. (1996). The shaky game: Einstein, realism, and the quantum theory. Second edition. Chicago: University of Chicago Press.

Hacking, I. (2000). How inevitable are the results of successful science? *Philosophy of Science* 67: S58–S71.

Ioannidis, J. (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

Popper, K. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.

Ratcliffe, M. (2011). Stance, feeling and phenomenology. *Synthese* 178: 121–130.

Shimony, A. (1970). Scientific inference. In R. G. Colodny (Ed.), *The nature and function of scientific theories: Essays in contemporary science and philosophy* (pp. 79–172). Pittsburgh: University of Pittsburgh Press.

Stanford, P. K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. Oxford: Oxford University Press.

——. (2009). Scientific realism, the atomic theory, and the catch-all hypothesis: Can we test fundamental theories against all serious alternatives? *British Journal for the Philosophy of Science* 60: 253–269.

Stringer, C. (2012). *Lone survivors: How we came to be the only humans on Earth.* New York: Henry Holt and Company.

van Fraassen, B. (1989). *Laws and symmetry*. Oxford: Oxford University Press.