THE EPISTEMOLOGY OF SCIENCE: ACCEPTANCE, EXPLANATION, AND REALISM

Finnur Dellsén

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the College of Arts and Sciences.

Chapel Hill
2014

Approved by:

Marc B. Lange

John T. Roberts

Matthew Kotzen

Ram Neta

William G. Lycan

ABSTRACT

FINNUR DELLSÉN: The Epistemology of Science:
Acceptance, Explanation, and Realism.
(Under the direction of Marc B. Lange)

Natural science tells a story about what the world is like. But what kind of story is this supposed to be? On a popular (realist) view, this story is meant to provide the best possible explanations of the aspects of the world with which we are all acquainted. A realist also thinks that the story should in some sense provide explanations that are probable in light of our evidence, and that these explanations ought to fit together into a coherent whole. These requirements turn out to be surprisingly hard to satisfy given the received view of how scientific theories are evaluated. However, I argue that if scientific theories are evaluated comparatively rather than absolutely for explanatory purposes – optimifically rather than satisficingly – then we can provide a fully realist view of the connections between explanation, probability, and coherence.

It is one thing to say what science's story of the world ought ideally be like, it is another to say that the story as it is actually being told lives up to this ideal. Do we have good reasons to believe that the picture as it is currently being presented to us is true, at least for the most part? "Yes," answer realists, "as long as our theories are empirically successful." Anti-realists respond that success is a poor guide to truth, appealing to the rather depressing history of successful theories that turned out to be false. Although I count myself among realists, I argue that realists have done themselves a disservice by focusing too much on empirical success in arguing for the correctness of the current scientific worldview. Instead I argue that one of the major reasons why currently accepted theories should (typically) be taken as true concerns the fact that they enjoy a certain kind of privileged status within scientific communities.

ACKNOWLEDGMENTS

Although philosophy is for me a mostly solitary activity, there is hardly a single thought in this dissertation that hasn't been greatly improved in response to the comments and criticism of many thoughtful and knowledgable people. Most importantly, I have benefitted enormously from interactions with my three dissertation advisors: Marc Lange, John Roberts and Matt Kotzen. It was John Roberts who more than once recognized some barely viable seeds of philosophical ideas – hidden in a forest of confusion and irrelevancies – and generously helped me transform them into some of the central theses and arguments herein. To Matt I owe special thanks for both encouragement and inspiration at crucial moments, though most importantly for extremely *constructive* criticism. My greatest thanks go to Marc, however, who has repeatedly made me suspect that the platonic form of dissertation advisors is located here in the material world.

Numerous others have helped me develop the ideas herein with their support and intelligent feedback. These include nearly all of those who have passed through the philosophy department at UNC Chapel Hill since I began my graduate studies in 2009. Among the UNC faculty, Ram Neta, Bill Lycan and Keith Simmons deserve special mention for their encouragement and helpful comments both early and late in the process. A good chunk of the current version of the dissertation was written in the good company of Miroslav Losonsky, Jen Kling, Megan Mitchell, and Tigo, all but one of whom provided intelligent comments that were much appreciated. Special thanks are also due to Craig Warmke, Luke Elson, Lindsay Brainard, Kate Nolfi, Wesley Sauret, Nate Sharadin and Huginn Thorsteinsson, for very helpful discussions at various stages, and to Elanor Taylor and Katie Elliot for being my official and not-so-official graduate student mentors earlier in the dissertation writing process. Finally, I would like to thank audiences at four Work-in-Progress presentations at UNC Chapel Hill, at the North Carolina Philosophical Society annual meeting

(Spring 2014), and at the symposium on scientific realism at the University of Iceland's Hugvisindathing (Spring 2014).

On a more personal note, work on this dissertation began shortly after the birth of my dear daughter, Katrín Anna. It is my hope that Katrín will someday read this and realize how all the joy she has brought me made me capable of spending so much time on the frankly esoteric topics herein. My mother Snjólaug has assisted me in various capacities while working on this dissertation, just as she has throughout my entire life. My greatest debt, however, will always be to my beloved Erla. Any attempt to express my gratitude to her for all that she has done is sure to fail, completely and miserably. I hope I will be forgiven for not even trying.

TABLE OF CONTENTS

1 PROLOGUE

Natural science tells a story about what the world is like – perhaps the most magnificent story ever told – and in so doing, it appeals to some of the most bizarre posits ever conceived. This story promises to explain most, and perhaps ultimately all, aspects of the world with which we are acquainted. But what kind of story is this supposed to be? In particular: (i) Is it meant to be true in all respects or just some? (ii) Must the story be consistent and closed under logical operations? (ii) And should we be confident in at least some of the particular details of the story as it is being told at this moment in time? These are the three main questions addressed in this dissertation (each in a separate chapter). My answers, in brief, are: "Yes, yes, and yes!" In that sense, I am a believer in *scientific realism*, and what follows is written in its defense.

But I am no gung-ho realist. It is unreasonable, in my view, to require the scientific theories that form part of this story to be *likely* to be true. Instead I argue that accepted theories should be, and often are, *more likely* to be true than other theories of the same generality. In other words, scientific theories are (on my view) evaluated comparatively rather than absolutely – at least in so far as the question is whether they should be accepted as part of the explanatory story told by natural science. I also do not think that a given scientist is required to believe all (or even any) aspects the story, although I do think she shouldn't accept a theory as part of the story unless she has good reasons to believe it. Finally, I do not think that we should be confident in *all* accepted scientific theories – not even all of those that have proved empirically successful. Rather, I'll argue that we should be confident only in a proper subset of the theories that currently form part of the explanatory story told by natural science.

This dissertation has three main chapters (chapters 2-4), each of which is written as an independent article and should be readable on its own. (Thus this is a "three-paper dissertation".) The first of these, "Real(ist(ic)) Acceptance" concerns the debate over scientific realism as influentially defined by van Fraassen (1980). On this set-up of the debate, realism is in part the view that to accept a theory involves believing that the theory is true. The first step in entering this debate, however, is to recognize that van Fraassen's definition of realism must be clarified (or, perhaps, reinterpreted) since on the the original definition realism comes out as either trivially true or trivially false. When reformulated in a natural way, I argue that an empiricist position (in the style of van Fraassen's constructive empiricism) is not susceptible to the common objection that its central distinction between empirical adequacy and truth is somehow arbitrary or insignificant. On the contrary, I argue, empirical adequacy is a natural stopping place for those who emphasize the role of empirical predictions in natural science. At the end of the day, however, I argue that only a realist view can account for the role of scientific explanation in natural science, and thus that van Fraassen's empiricism fails.

The second of the three papers, "Explanatory Acceptability", aims to precisify the idea that appealing to a theory in explanation requires that one have good reasons to think it's true. In short, the paper argues for *comparatively* evaluating the theories that we honor as part of the explanatory story told by natural science. This is done partly by arguing against "the threshold-view", according to which theories are acceptable in this way just in case their probability exceeds some threshold. As is well-known, any threshold-view is susceptible to the infamous Preface and Lottery paradoxes, but I argue that these paradoxes are particularly damaging when it comes to the kind of acceptance of scientific theories that a realist would be interested in. By proving a few theorems concerning comparative probabilities, I then show how the alternative view that I favor – the Optimality Account – avoids these paradoxes altogether. On this view, a theory is acceptable in the relevant sense just in case it is (significantly) more probable than any other theory that purports to answer the same explanation-seeking questions.

In the last of the three papers, “Realism and the Absence of Rivals”, I turn my attention to slightly different scientific realism debate (one that was explicitly set aside by van Fraassen).[1] This debate roughly concerns whether we are epistemically justified in believing that (some specified subset of) empirically successful scientific theories are true. There is a particularly forceful anti-realist challenge in this debate, according to which even our most successful scientific theories are likely to be replaced by rival theories that we have so far failed to conceive of. I argue that a probabilistic approach to this challenge shows that a general strategy for responding is available to realists. I then argue that this reveals that realists can use the history and sociology of science to their advantage in replying to what is arguably the most serious challenge to realism today.

As this overview of the three papers should make clear, I take *acceptance* and *explanation* to play important roles in a *realist* epistemology of science. Indeed, the three terms in the subtitle of this dissertation – “acceptance”, “explanation”, and “realism” – denote the three central concepts that I am concerned with here. Another important concept, not mentioned in the subtitle, is that of “probability”. Much of what I have to say about acceptance, explanation and realism will appeal to a broadly-speaking *probabilistic* framework for epistemology. So, the methodological approach of this dissertation can be fairly characterized as approaching various realist themes by some well-known but very useful formal probabilistic tools. Not all realists may be sympathetic to this general approach, but I make no apologies for defending realism by adopting this probabilistic framework given the incredible successes of this approach in the past half-century or so.

Let me mention another methodological choice. I shall not normally be concerned with proposing elaborate thought experiments, soliciting my own (or the reader’s) “intuition” to argue for some point, nor shall I look in much detail at cases from the history of science. I do not have any objections to such approaches in general, but I have found it

[1] In chapter 2, however, I argue that once van Fraassen’s realism has been reformulated it makes clear connections with this other scientific realism debate.

more useful for current purposes to tackle these issues by abstracting away from hypothetical examples and much of scientific practice in order to get a clearer view of the larger picture. Indeed, nowhere in what follows do I claim that the picture of the epistemology of science that I present is *complete*. What I do claim, perhaps immodestly, is to have provided a partial *map* of the territory – one that is correct in so far as it purports to represent its object, albeit perhaps skewed to draw out the features of interest to the map maker.

2 REAL(IST(IC)) ACCEPTANCE

Abstract

Does accepting a scientific theory involve believing that the theory's claims about unobservable entities are true? This is one question discussed under the label "scientific realism". In the first part of the paper, I argue that the question needs to be clarified as being about the normative question of whether a theory is *acceptable* only if it is reasonable to believe that it is true. I then show how an empiricist view – according to which a theory is acceptable only if it is reasonable to believe that it is empirically adequate – is straightforwardly motivated by the relationship between acceptance and empirical prediction. However, I go on to argue against this empiricist view by appealing to a structurally identical argument that appeals to the relationship between acceptance and scientific explanations, and to analogues of Moore's paradox for scientific explanation. I conclude that there is a strong case for a realist view of acceptance.

2.1 Introduction

A popular quip has it that there are at least as many versions of scientific realism as there are scientific realists and anti-realists combined. One widely-discussed version of scientific realism, influentially discussed in Bas van Fraassen's *The Scientific Image* (1980), concerns what kind of epistemic attitude is involved in the acceptance of a scientific theory, where realists and anti-realists are seen as disagreeing about whether that attitude extends to the unobservable entities posited by scientific theories. According to the realist position defined by van Fraassen, "acceptance of a scientific theory involves the belief that it is true" (van Fraassen 1980, 12). By contrast, van Fraassen's own anti-realism, *constructive* *empiricism*, holds that "acceptance of a theory involves as belief only that it is empirically adequate" (van Fraassen 1980, 8), where a theory is "empirically adequate" roughly just in case it is correct in its claims about the observable aspects of the world.[1]

While this conception of scientific realism and anti-realism has received its fair share of attention, it has proven hard to get a grip on exactly what the realist and anti-realist are meant to be disagreeing about. Indeed, some philosophers have argued that the debate is either confused or trivially settled in favor of the realist. (Blackburn 1984, 2002; Mitchell 1988; Horwich 1991; Teller 2001) The two-fold aim of this papers is (a) to clarify the dispute so as to draw out the essential point at which realists and anti-realists part ways, and (b) argue for a realist position concerning scientific acceptance according to which accepting a theory requires that it be reasonable to believe that the theory is true.

I will proceed as follows: In section 2, I argue that the dispute between realists and antirealists needs to be clarified, and then suggest a natural and plausible way to reconstruct the debate. In section 3, I go on to explain why, on this conception of the debate, an empiricist (in the style of van Fraassen's constructive empiricism) would be inclined to put such emphasis on the distinction between observable and unobservable entities. In sections 4, 5 and 6, however, I argue that the empiricist cannot adequately account for an essential part of the scientific enterprise, viz. the practice of *explaining* natural phenomena. On this basis I present a deductively valid argument for realism, and defend each of the premises.

2.2 Realism: Reconstructing the Debate[2]

On van Fraassen's conception of the debate, the realist and the anti-realist are both making claims about the relationship between *acceptance* and *belief*. It is clear enough what the relationship is supposed to be: In saying that acceptance *involves* some belief

[1]This rough characterization of empirical adequacy will do for the purposes of this paper. For a much more precise characterization, see (van Fraassen 1980, chapter 3).

[2]The title of this section is a friendly nod to Simon Blackburn's "Realism: Deconstructing the Debate" (2002).

or other, the idea is that the belief in question is *necessary* for acceptance. So according to van Fraassen, the realist position can be characterized as holding that one accepts a scientific theory T only if one believes that T is true. An anti-realist, by contrast, denies that acceptance of a theory requires that one believe that the theory is true – although a constructive empiricist such as van Fraassen grants that a *restricted* belief is required for acceptance, namely the belief that T is *empirically adequate*.

As things stand, however, it is unclear what concept is denoted by the term "acceptance", and so it is unclear what exactly the realist and the constructive empiricist are disagreeing about. Indeed, it has been argued – by Blackburn (1984, 2002), Mitchell (1988), Horwich (1991), and Teller (2001) – that to accept a theory is conceptually identical to believing it to be true – that "acceptance" and "belief" denote the same concept. Simplifying somewhat, their argument rests on the claim that to believe a proposition just is to be disposed to behave in certain ways, and that acceptance is meant to be exactly that sort of disposition. If that's right, then constructive empiricism (and indeed any anti-realist view) is *incoherent*: One couldn't possibly accept T without believing T, and so the antirealist position would simply be confused. Moreover, scientific realism would be *analytically true*, which would surely come as a surprise to many realists.[3] No wonder Blackburn concludes that the issue of scientific realism "has not been clearly posed". (Blackburn 2002, 111)

The Blackburn-Mitchell-Horwich-Teller objection challenges anti-realists (and also realists who believe that there is a non-trivial issue concerning realism about acceptance) to find some way of defining "acceptance" such as not to make acceptance conceptually identical to belief. Unfortunately, van Fraassen himself never explicitly defined "acceptance" (or "belief"), despite its central role in his characterizations of scientific realism and constructive empiricism. However, several *other* authors have defined terms which are meant to be contrasted with "belief", and which they refer to as "acceptance". One such

[3]Realists these days typically take their thesis to be an *a posteriori* and even quasi-empirical thesis about science. (Putnam 1978; Boyd 1980, 1984; Psillos 1999)

definition stands out as particularly congenial to the debate over scientific realism.

2.2.1 Distinguishing Acceptance and Belief

In *An Essay on Belief and Acceptance*, L.J. Cohen defines "acceptance" and contrasts it with "belief", as follows:

> [...] belief that p is a disposition, when one is attending to issues raised, or items referred to, by the proposition that p, normally to feel it true that p and false that *not-*p, whether or not one is willing to act, speak, or reason accordingly. But to accept the proposition or rule of inference that p is to treat it as given that p. More precisely, to accept that p is to have or adopt a policy of deeming, positing, or postulating that p – i.e. of including that proposition or rule among one's premisses for deciding what to do or think in a particular context, whether or not one feels it to be true that p. (Cohen 1992, 4)

Cohen's definition has many interesting dimensions, but what's important for our purposes is only that to accept a proposition is on this definition to have *a policy* of *taking the proposition as given* in a particular context. This is congenial to the scientific realism debate because there is a clear sense in which scientists who use certain theories in their work as scientists have a policy of taking those theories as given in a scientific context. So, following Cohen's definition, we can say that for a scientist *qua* scientist to accept a theory T is to have or adopt a policy of taking T as given in a scientific context.[4]

Of course, this definition is not particularly informative unless we specify what is involved in taking a theory as given *in a scientific context*. We shall return to that issue below (in sections 3 and 4). For now, I want to note that if Cohen's definition is even roughly on the right track, then acceptance and belief can come apart in that one can accept something that one does not believe to be true. Indeed, one could accept something that one does not even believe to be empirically adequate. Roughly speaking, this is because accepting a theory is a matter of being *prepared to do something* with that theory, whereas

[4]For similar definitions of "acceptance", see (Alston 1996), (Lehrer 1979), (Kaplan 1981b,a, 1995), (Maher 1993), and (Lance 1995).

believing a theory is a matter of *feeling it to be the case*. Of course, it may well be true that most of what one accepts one also believes to be true or empirically adequate, but if acceptance is a matter of doing something whereas belief is a matter of feeling something, then it's surely possible to accept something one does not believe.[5]

To bring this out more vividly, consider a deeply religious evolutionary biologist – Alyssa – who uses Darwin's theory of natural selection in her practice as a scientist, e.g. by asserting the theory *ex cathedra* and using it in her scientific explanations and predictions. In other words, Alyssa has adopted a policy of using the theory of natural selection in the context of her scientific work, and thus *accepts* it. However, suppose also that because of her religious convictions, Alyssa just cannot bring herself to believe any part of Darwin's theory – she is psychologically unable to do so. Alyssa may even realize that she *ought* to believe at least some parts of the theory, perhaps because (she thinks) the evidence speaks overwhelmingly in its favor. Yet Alyssa does not believe (to any degree) that the theory is either true or empirically adequate.

If Alyssa's case is even possible, then acceptance and belief (of any kind) can come apart. And that would seem to show that a particularly extreme form of anti-realism is correct, one according to which no belief at all is necessary for acceptance. Of course, this contradicts not only the realist view defined by van Fraassen (1980), but also his own view, constructive empiricism. However, we arrived at this conclusion (I suggest) only because we set up the issue in way that avoids the central issue. That central issue is not about the metaphysical connection between acceptance and belief – i.e. whether belief that T is true as opposed to merely empirically adequate *constitutes* acceptance of T – but rather about the connection between the *normative requirements* of acceptance and belief. Let me explain.

[5] This is indeed the conclusion that Cohen (1989, 1992) himself draws concerning belief and acceptance in a scientific context.

2.2.2 Realism Redefined

Let us start not with acceptance, but with belief. Here is an utterly plausible and uncontroversial thing to say about belief:

(B) It is permissible to believe p if and only if it is reasonable to believe that p is true.

Here and throughout, I shall be using "reasonable" as a term of epistemic appraisal, much like "justified", "rational", or "warranted". So (B) posits a normative requirement on belief – a "norm of belief" – according to which one should only believe something if one has adequate reasons to think it's true. One may want to spell out (B) in various ways, e.g. by specifying whether it applies to "full beliefs" and/or "degrees of belief", and what precisely is involved in it being *reasonable* to believe something. But none of that will be important in what follows. What's important is only that according to (B), believing p normatively requires that it be reasonable to believe that p is *true*.

Now, importantly, to say that belief is governed by the requirement described in (B) is not to say that (B) cannot be overridden or outweighed in a particular case. Suppose someone threatens to murder your family unless you believe that the earth is flat. There is clearly a sense in which believing that the earth is flat is permissible in such a case, even though doing so would not be reasonable (in the epistemic sense). Nevertheless, there is also a sense in which you have failed *as a believer* if you manage to convince yourself that the earth is flat. In this respect the norms of belief are like the rules of a game, e.g. the rule in chess that says that the bishop should only be moved diagonally: The fact that one could have excellent prudential or moral reasons to move one's bishop in a different manner does not show that the rule fails to apply in a given case. Put differently, the imperative to follow the rule is overridden, not annihilated, by stronger external considerations. Similarly, (B) may be overridden or outweighed in a particular case, e.g. by moral or prudential considerations, but that does not mean that the requirement only to form reasonable beliefs does not apply.

We have discussed a normative requirement on *belief*. Now consider *acceptance*. A realist thinks that there is some intimate connection between accepting a theory and believing that the theory is true. If that connection cannot be about whether belief is necessarily involved in acceptance, as I have argued, then perhaps the connection concerns the *normative requirements* on acceptance and belief. What I am suggesting is that the realist can be understood as holding that *one should only accept a theory if it is permissible to believe it*. Now, note that given (B), this thesis is equivalent to the claim that one should only accept a theory T if it is reasonable to believe that T is true. Thus we can (re)define the realist thesis as follows:

(R) One should only accept a theory T (in a scientific context) if it is reasonable to believe that T is true.[6]

Let us say that a theory is "scientifically acceptable" just in case one may accept it (in a scientific context). Thus (R) can be restated as the view that a theory is scientifically acceptable only if it is reasonable to believe that it is true.

It's important to note that (R) is not incompatible with there being other normative requirements on scientific acceptance. Thus a proponent of (R) – a realist – may say, for example, that theories are only scientifically acceptable if they are reasonably simple and well-managed (even if she thinks these features are merely pragmatic as opposed to epistemic virtues of the theory). After all, accepting very complicated or unwieldly theories may be a bad idea from a practical standpoint, because calculations and derivations with such theories would be unnecessarily difficult. To acknowledge such "pragmatic" norms of acceptance does not make one an anti-realist on this definition as long as one also thinks that acceptance is governed by a norm requiring that it be reasonable to believe the theory to be true.

[6] A slightly weaker form of (R) replaces "true" with "*approximately* true". Nothing in what follows depends on which version of (R) one adopts, so for simplicity's sake I shall stick with (R) as formulated in the main text.

Now, why think a commitment to (R) deserves to be called a *realist* view of acceptance? There are two reasons. First, (R) captures the kernel of truth in van Fraassen's suggestion that scientific realism holds that accepting a theory involves believing it to be true, because (given plausible assumptions) van Fraassen's definition implies (R). To see this, note that if accepting a theory T did involve believing that T is true, then surely one should only accept T if believing T (which would be involved in accepting it) is permissible. Given (B), this entails that one should only accept T if it's reasonable to believe T. Thus van Fraassen's conception of realism straightforwardly *implies* (R). The implication does not go the other way, for one could commit to the claim that one should only accept what it's reasonable to believe is true, and yet deny in the same voice that acceptance *involves* any belief at all (as I have argued we should). So, (R) is a more modest conception of realism that nevertheless captures the basic idea that there is some intimate connection between the acceptance of a theory and the belief that it is true.[7]

Another reason why (R) deserves to be called a *realist* view of acceptance is that (R) is in tension with well-known anti-realist arguments. Consider, for example, the well-known *Underdetermination Argument* (UA), which concludes (roughly) that it is not reasonable to believe any scientific theories about unobservables to be true, because for any such theory there is (according to the argument) a rival theory that is at least as well supported by one's evidence. Why is this argument generally considered to be a threat to scientific realism? I suggest it is at least in part because the conclusion of UA conflicts with (R) given the claim that one shouldn't reject all scientific theories about unobservables. To see this, note that the following claims form an inconsistent triad:

(i) One should only accept a theory T if it is reasonable to believe that T is true. [(R)]

[7]Note also that those who (for whatever reason) object to my reformulation of the issue of realism about acceptance will have to agree that a defense of (R) is *necessary* (albeit not sufficient) for a defense of scientific realism on van Fraassen's conception. So while I think it would be a mistake to conceive of the issue as being about van Fraassen's conception of realism, those who do should still find something of value in the discussion of (R) in what remains of this paper.

(ii) It is not reasonable to believe that any theories about unobservables are true. [Conclusion of UA.]

(iii) It is permissible to accept some scientific theories about unobservables.

Clearly, no genuine realist would reject (iii) (and neither would most anti-realists). If realism is also committed to (i), as I'm suggesting, then it follows that realists must reject (ii), the conclusion of UA. However, if realism is *not* committed to (i), then it's not clear why realists couldn't simply embrace (ii). Of course, one might think rejecting (ii) is definitional of what it is to be a scientific realist, but given (R) we can give a principled reason why realists *must* reject UA. So, at the very least, (R) fits very well with the plausible thought that skeptical arguments like UA are distinctively *anti-realist* arguments.

2.3 Two Anti-Realisms

The previous section argued that realism should be thought of as a commitment to (R), which requires of an acceptable theory that it be reasonable to believe that it's true. Note that to deny (R) is to say that scientific acceptability does *not* require that be is reasonable to believe that the theory in question is true, but that leaves open what, if anything, acceptability requires in terms of reasonable belief. So there will be many ways to be an anti-realist. This section examines two such anti-realist positions, arguing that only one of them can account for the role of accepted theories in empirical predictions.

2.3.1 Empiricism and Manifestationalism

According to van Fraassen's (1980) constructive empiricism, acceptance of a theory involves, as belief, only the belief that the theory is empirically adequate. This cannot be quite right, I argued, for acceptance (on a plausible definition) need not involve any belief at all. However, in much the same way as the realist can be (re)defined as committing to a normative connection between acceptance of a theory and belief that the theory is true, van Fraassen's empiricism may be (re)defined as a commitment to a *normative* connection between acceptance of a theory and belief that the theory is empirically adequate, viz. that one should only accept a theory if it is reasonable to believe that it is empirically adequate:

(E) One should only accept a theory T (in a scientific context) if it is reasonable to believe that T is empirically adequate.

In other words: A theory T is scientifically acceptable (according to (E)) only if it's reasonable to believe that T is correct in all its claims about the observable aspects of the world. Note, however, that (R) entails (E), so committing to (E) does not by itself make one a van Fraassen-style empiricist. Rather, this kind of empiricism must be seen as committing to (E) being the whole story about what acceptability requires in terms of what it's reasonable to believe. A bit more precisely, the (re)defined empiricist view – which I'll simply call *empiricism* – commits to (E) and rejects any stronger norm, such as (R), relating acceptance and reasonable belief. So, in particular, empiricism denies that it must be reasonable to believe that T is true for T to be scientifically acceptable.

A standard challenge to this kind of empiricism (one that has been voiced numerous times against van Fraassen's constructive empiricism) attacks its reliance on the distinction between observable and unobservable entities (or equivalently for our purposes, the distinction between truth and empirical adequacy). (Maxwell 1962; Churchland 1985; Kitcher 2001a) To be sure, van Fraassen himself acknowledges that the boundary between what's observable and unobservable is vague and relative to the epistemic community in which scientists are working. (van Fraassen 1980, 1985) Indeed, for van Fraassen, what's observable is itself a matter of empirical investigation, and thus there is no simple rule for telling what counts as observable or unobservable. Yet, van Fraassen argues, the fact that a distinction is vague, relative, and not yet fully specified does not mean that one cannot employ it in one's philosophical theorizing. (Muller and van Fraassen 2008; van Fraassen 2001)

Van Fraassen may very well be correct to dismiss concerns about how and where to draw and distinction between observable and unobservable entities. The deeper worry in the vicinity, however, is not that the distinction cannot be coherently drawn, but that the significance that the empiricist attributes to it is *unmotivated*. In support of this, realists often point out that it is hard to see what is in principle more problematic about forming beliefs concerning unobservable entities than forming beliefs about unobserved-but-observable entities. Why, for example, would it be more problematic for scientists to confirm that there are unobservable atoms than that there is some observable-but-as-yet-unobserved deep sea creature? More generally, it seems that if empiricists are worried about the epistemic support one could acquire for believing theories concerning unobservable entities, they ought to worry equally about theories concerning unobserved-but-observable entities. (Railton 1989; Sober 1985, 1993; Psillos 1996; Alspector-Kelly 2001; Ladyman 2007)

This suggest that any epistemic motivation for (E) in fact provides a stronger motivation for an even weaker connection between acceptance and reasonable belief:

(M) One should only accept a theory T (in a scientific context) if it is reasonable to believe that T is manifestationally adequate.

where a theory is "manifestationally adequate" just in case it is correct in all its claims about what has been observed so far. (Railton 1989) Note again that (M) does not by itself conflict with (E) or (R), so this more extreme anti-realist view – which I'll call *manifestationalism* – must be understood as claiming that (M) is the whole story about what acceptability requires in terms of reasonable belief – that there is no stronger norm, such as (E) or (R), relating acceptance and reasonable belief.

The worry this poses for the empiricist is that it is unclear why one should commit to (E) rather than its weaker counterpart (M). In what remains of this section, I show that the empiricist has a convincing answer to this worry, roughly because a theory's empirical adequacy is required for the theory to make correct empirical predictions. However (to foreshadow some of the discussion to come), I will go on to argue in the next section that the empiricist is wrong to think that scientific acceptability requires only that it be reasonable to believe that a theory is empirically adequate. Thus, my twofold conclusion concerning empiricism will be that the distinction between observables and unobservables, while not arbitrary or unmotivated, ultimately cannot carry the weight the empiricist places upon it.

2.3.2 Empirical Adequacy and Prediction

Recall that we said that the acceptance of a theory T (in a scientific context) ought to be understood in terms of the role that T plays in what a scientist *does* with T. Now, empiricists typically emphasize the use of theories in making empirical predictions[8] about the behavior of the observable world. They point out, plausibly enough, that we look to science to build airplanes, construct bridges, cure diseases, and so forth, and that requires us to have scientific theories that accurately predict the behavior of ordinary observable objects such as airplanes, bridges and organisms. This point is echoed by Nancy Cartwright in a recent sympathetic discussion of van Fraassen's position: "To accept a theory is to decide to use *it* to make all those predictions about what we might observe that will help us chart our actions." (Cartwright 2007, 40)

This suggestion has obvious repercussions for the normative connection between acceptance and belief. For if acceptance of a theory just is to adopt a policy of appealing to the theory in one's empirical predictions, as Cartwright is effectively suggesting, then clearly one shouldn't accept a theory if one shouldn't appeal to the theory in one's empirical predictions. And, of course, one should only appeal to a theory in empirical predictions if one has good reasons to think that the theory's predictions will be *correct*. But now note that to say that a theory makes correct empirical predictions is to say that the theory is empirically adequate. It follows that one shouldn't predict with a theory unless it is reasonable

[8]Here and in what follows, I will use "prediction" in a broad sense that includes predictions about the the present and the past ("retrodictions").

to believe that the theory is empirically adequate.[9]

Putting all of these points together we get the following simple argument for (E):

The Empiricist Argument

(E1) Accepting T (in a scientific context) is, at least in part, to appeal to T in empirical predictions.

(E2) If ϕ-ing is (partly or wholly) constituted by ψ-ing, then one should not ϕ if one should not ψ.

(E3) So, one should not accept T (in a scientific context) if one should not appeal to T in empirical predictions. [From (E1) and (E2).]

(E4) One should only appeal to T in empirical predictions if it is reasonable to believe that T is empirically adequate.

(E) Therefore, one should only accept T (in a scientific context) if it is reasonable to believe that T is empirically adequate. [From (E3) and (E4).]

In supporting (E), this argument shows why the manifestationalist norm (M) – according to which one shouldn't accept a theory unless it is reasonable to believe that the theory is manifestationally adequate – is not strong enough to be the norm that relates acceptability and reasonable belief. After all, a merely manifestationally (as opposed to empirically) adequate theory need not make any correct predictions at all about things we haven't yet observed. Relatedly, the argument also shows that the observable-unobservable distinction appealed to in (E) is not as irrelevant as scientific realists have often argued. The distinction is relevant not because it marks some important epistemological distinction such that all and

[9]I am simplifying significantly here (and in (E4) below). Depending one what one means by "reasonable to believe", it may very well be that one should sometimes appeal to a theory in one's predictions even if it is not fully reasonable to believe it to be empirically adequate. But all that's really at issue here is that one should only appeal to a theory in one's predictions if one has *some* reasons for believing that the theory is empirically adequate (as opposed to, say, merely manifestationally adequate), for only then does one have *some* reasons for thinking that one's empirical predictions will be correct.

only things falling on one side of the distinction are knowable, but because one should only predict with theories that one has good reasons to believe are empirically adequate.[10]

However, while this argument demonstrates the relevance of the observable-unobservable distinction and refutes manifestationalism, it does not yet support empiricism as against realism. Put differently, **the Empiricist Argument** is an argument for (E), but not an argument that acceptability does not require anything beyond what's required by (E) in terms of reasonable belief. So a realist will aim not to refute **the Empiricist Argument**, but rather to show that acceptance requires, in addition to it being reasonable to believe what the theory says about observables, that it be reasonable to believe what it says about *un*observables. The next couple of sections illustrate one way of doing precisely that.

[10] As John Roberts pointed out to me, there is an interesting intermediate position between (E) and (M):

(A) One should only accept T (in a scientific context) if it is reasonable to believe that T is *actually adequate*,

where as theory is "actually adequate" just in case it is correct about everything it says about what is, has been, or will be, observed (i.e. just in case all actual observations (past, present, and future) are in accordance with the theory).

Now, one might think that what we really care about is having theories that are correct about any *actual* observations that we make, as opposed to counterfactual and forever unrealized observations. After all, what does it matter to us whether a theory is correct in its predictions about events that will never actually come to pass? Perhaps then (E4) should be weakened to something like the following:

(A4) One should only appeal to T in empirical predictions if it is reasonable to believe that T is actually adequate.

If this is is right – if (E4) is too strong and should be replaced with (A4) – then we have only an argument for (A) and not also for (E). Hence the Empiricist Argument would not be as strong as I have indicated.

However, it seems to me that there are good reasons to resist weakening (E4) in this way. The crucial point is that one can make empirical predictions even about the outcome of non-actual observations, and these predictions might be of great importance to a scientist's work. To illustrate, consider the following example: A chemist, Beatrice, accepts a theory according to which mixing two chemical compounds will result in a massive explosion. Accordingly, Beatrice decides never to mix the compounds, and (due to her influential advice) neither does anyone else in the history of the world. Unbeknownst to her, however, mixing the two compounds results in a stable and durable substance with wonderful potential applications. Now, although the outcome of this prediction will never be actualized, it seems clear that it matters a great deal whether Beatrice accept the false theory she does in fact accept or the true theory that she could have accepted in its stead. Thus, contrary to the argument for weakening (E4) to (A4), it seems to me that it does matter greatly whether it is reasonable to believe that a theory gets things right about unactualized observations.

2.4 Accepting and Explaining

The last section provided a motivation for (E), the requirement that one should only accept a theory if it's reasonable to believe that it is empirically adequate. As I have mentioned, a realist will not disagree with (E), but rather urge that it follows from a stronger requirement (R) according to which one should only accept a theory if it's reasonable to believe it to be true. But if empirical adequacy is all one needs for making correct empirical predictions, one might wonder why anything beyond reasonable belief in a theory's empirical adequacy is involved in scientific acceptability. This section argues that the realist can exploit the fact that science provides *explanations* to argue that acceptance of a theory involves appealing to it in explanations. The next two sections explore the consequences of this for empiricism and realism.

Let me start by pointing out that the scientific enterprise uncontroversially involves (among other things) a practice of giving explanations for natural phenomena. This clearly cannot be denied without flying in the face of scientific practice. For one thing, scientific journal articles frequently reference explanations. (A casual search (performed on July 10, 2013) in the online database of *Science* reveals that the terms 'explain', 'explains' and 'explanation' occur in around 36% (6630 out of 18636) of the Original Research articles published since 1996.) Moreover, scientists themselves are quite explicit about their aim being to give explanations. Witness Steven Weinberg:

> [...] our purpose in theoretical physics is not just to describe the world as we find it, but to explain – in terms of a few fundamental principles – why the world is the way it is. (Weinberg 1995, xx)

To be sure, some early philosophers of science did appear to deny that science offers explanations. The idealist-positivist Karl Pearson wrote in the third edition of *The Grammar of Science*:

> Nobody believes now that science *explains* anything; we all look upon it as a shorthand description, as an economy of thought. (Pearson 1911, xi)

Pierre Duhem seems to have advocated a similar view:

> A physical theory is not an explanation. It is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as simply, as completely, and as exactly as possible a set of experimental laws. (Duhem 1982, 19)

As Hempel (1966) and Salmon (1989, 1992) point out, this hostility towards the idea of scientific explanations seems to have been based on the idea that to explain something would necessarily involve an appeal to theories that couldn't be confirmed by empirical methods, e.g. teleological or metaphysical principles. On that assumption, the notion of a *scientific* explanation may sound like an oxymoron, at least to positivists such as Pearson. However, we now have numerous philosophical accounts of scientific explanation according to which such explanations need not make any reference to "unscientific" principles of any sort. Indeed, the DN-model of scientific explanations given by Hempel and Oppenheim (1948) and Hempel (1965) makes empirical testability a necessary condition for something to count as a scientific explanation in the first place. So, given contemporary conceptions of scientific explanation, there is no apparent motivation anymore – not even for those with empiricist sympathies – for the once popular view that science does not explain.

Assuming then that there is a practice in science of giving explanations of natural phenomena, it is clear that this practice requires giving preference to some theories as opposed to others, in that some theories but not others are deemed worthy of being appealed to in such explanations. To illustrate, most scientists would not use the theory that the earth is flat in their proposed explanations, whereas they would happily appeal in their explanations to the theory that there are electrons (negatively charged entities with a rest mass of approximately 9.11×10^{-31} kg). So we can distinguish between theories that scientists would, and would not, be prepared to appeal in their proposed explanations.[11]

[11] Of course, some theories will be such that scientists will be prepared to appeal to them in some explanations and not in others. For example, some scientists would appeal to Newtonian Mechanics in explaining the tides, but not in explaining Mercury's perihelion. To simplify our discussion, I set this complication aside in what follows.

Now, it is natural to refer to the the attitude that we have towards the theories we are prepared to appeal to in explanations (and fail to have towards theories we are not prepared to appeal to in explanations) as "acceptance". Indeed, the definition of "acceptance" with which I have been operating, viz. that of Cohen (1992), seems to imply that accepting T involves being prepared to use T in explanations. After all, this definition has it that to accept T is to adopt a policy of taking T as given in a particular context, so when a scientist accepts a theory T *in a scientific context* then she would have to take T as given when giving scientific explanations, and thus be prepared to use T in the explanations she proposes. Thus it seems that on our definition of "acceptance", a scientist *qua* scientist only accepts a theory if she is prepared to appeal to it in scientific explanations.[12]

That being said, I think it is something of a red herring to focus on whether any particular definition of "acceptance"[13] entails that acceptance of a theory involves being prepared to appeal to it in explanations. This is because the real issue of scientific realism does not hinge on the semantic issue about what sort of attitude the term "acceptance" refers to. As long as there is an attitude involved in the scientific enterprise that can accurately be described as a disposition or willingness to appeal to certain theories in explanations – call it "explanatory willingness", if you will – then we can ask all the same questions about this attitude as we would about *acceptance*. In particular, we can ask whether one may have this attitude of *explanatory willingness* towards a theory only if it is reasonable to believe

[12] Although it is largely irrelevant to our current concerns (for reasons given below), one might wonder whether van Fraassen would agree that acceptance of a theory involves appealing to it in explanations. Some of what van Fraassen says in *The Scientific Image* seems to suggest that it does:

> [...] acceptance involves not only belief but a certain commitment. Even for those of us who are not working scientists, the acceptance involves a commitment to confront any future phenomena by means of the conceptual resources of this theory. It determines the terms in which we shall seek explanations. If the acceptance is at all strong, it is exhibited in the person's assumption of the role of the explainer, in his willingness to answer questions *ex cathedra*. (van Fraassen 1980, 12)

However, van Fraassen is not being unequivocal here, for (as John Roberts pointed out to me) he might mean that the theories we accept merely determine the *concepts* that we use in our explanations, without thereby determining which theories we appeal to in such explanations.

[13] Including van Fraassen's definition (see previous footnote).

that the theory is true. Thus, even if the anti-realist could convince us that acceptance does not involve explanatory willingness, the realist may argue that explanatory willingness – an attitude integral to the scientific enterprise – requires that it be reasonable to believe that scientific theories are true. Surely that would be only a pyrrhic victory for the anti-realist – a way of winning a semantic battle while losing the philosophical war.

Let us take stock. We saw that scientists give explanations, and that this requires a certain kind of distinction between the theories a given scientist is and is not prepared to appeal to in explanations. Therefore, I argued, acceptance (in any sense relevant to the debate over scientific realism) must be said to involve being prepared to appeal to the accepted theory in explanations. But now note that if that is so, then clearly one shouldn't accept a theory if one shouldn't appeal to the theory in one's explanations. After all, appealing to the theory in explanations would be part of what it is to accept a theory, so any normative requirement on explanation is *ipso facto* a normative requirement on acceptance. The point here is exactly analogous to the point about acceptance and empirical prediction from the previous section: Because acceptance involves appealing to theories in one's predictions, one shouldn't accept a theory if one shouldn't appeal to the theory in predictions. Similarly, because acceptance involves appealing to theories in explanations, one shouldn't accept a theory if one shouldn't appeal to the theory in one's explanations. The next section argues that this provides the resources for an argument for a modest kind of scientific realism about acceptance.

2.5 An Argument for Realism

Recall where we started our evaluations of empiricism in section 3: A common objection to empiricist positions in the realism debate is that they place an unwarranted emphasis on the supposedly arbitrary distinction between the observable and the unobservable – or equivalently for our purposes, on the distinction between empirical adequacy and truth. To this we countered that, as far as empirical predictions are concerned, the distinction between empirical adequacy and truth is not arbitrary at all, since empirical adequacy is precisely what is required for a theory to make correct predictions. However, we also saw that making empirical predictions with a theory does not exhaust what it is to accept that theory – it also involves appealing to the theory in one's explanations – and thus that one shouldn't accept a theory if one shouldn't appeal to it in explanations. So we have established a connection between acceptability and explanation, but why should this support realism and undermine empiricism? Roughly, I suggest that it does so in virtue of a normative connection between explanation and reasonable belief – one that is akin to the normative connection between prediction and reasonable belief appealed to in the argument for (E) in section 3.

Before I go any further, let me set aside a terminological issue that arises in this context. In their seminal paper on scientific explanations, Hempel and Oppenheim (1948) argued that the theories appealed to in an explanation must be true. (Hempel and Oppenheim 1948, 11) They subsequently introduced the term "potential explanation" for something that satisfies all other conditions for being an explanation. However, van Fraassen (1980) argued that we often use the term "explanation" in contexts where we are either agnostic about the truth of the proposed explanans, or even believe it to be false. We say, for example, that the phlogiston theory explained combustion, and that Bohr's model of the atom explained the hydrogen spectrum. At least for the sake of the argument, I shall grant van Fraassen that explanations need not be true. Thus, in what follows, I shall not insist on using "explanation" in a sense in which something only counts as an explanation if it is true.

Note, however, that this is not to say that there is no *normative* connection between truth and explanation – it may yet be that explanations "aim at truth" in the sense that one should only appeal to a theory in explanations if it is reasonable to believe that it is true. To draw out the difference, consider *belief*: Since one can believe false propositions, it is not a necessary condition for something to be a belief that the believed proposition is true. Yet it is uncontroversial that one should only believe propositions if one has adequate reasons to think it's true. So there is a normative connection between belief and truth, despite the fact that belief and truth can come apart. Similarly, I shall argue, there is is a normative connection between *explanation* and truth. More precisely, I shall argue that one should only appeal to T in explanations if it is reasonable to believe that T is true. Given the connection already established between acceptance and explanations, this entails the realist view that one should only accept a theory if it is reasonable to believe that it is true (as opposed to merely empirically adequate, for example).

This argument can be formalized as follows:

The Realist Argument

(R1) Accepting T (in a scientific context) is, at least in part, to appeal to T in scientific explanations.

(R2) If ϕ-ing is (partly or wholly) constituted by ψ-ing, then one should not ϕ if one should not ψ.

(R3) So, one should not accept T (in a scientific context) if one should not appeal to T in scientific explanations. [From (R1) and (R2).]

(R4) One should only appeal to T in scientific explanations if it is reasonable to believe that T is true.

(R) So, one should only accept T (in a scientific context) if it is reasonable to believe that T is true. [From (R3) and (R4).]

Note that this argument is structurally identical to **the Empiricist Argument**. Indeed, we arrive at this argument by replacing "empirical predictions" with "scientific explanations" and "empirically adequate" with "true." This is the argument that I shall be defending in the rest of this paper.[14]

[14]For full disclosure, I should say that I think (R4) to be an approximation (much like I took (E4) to be an

Is **the Realist Argument** sound? Well, it is a deductively valid argument, so let's consider the premises. (R1) was defended at length in the previous section. (R2) seems self-evident (and is at any rate required for **the Empiricist Argument** as well). This leaves us with (R4), to which we now turn.

2.6 Explanation and Reasonable Belief

The (at this point) crucial premise of **the Realist Argument**, (R4), holds that one should only appeal to a theory in one's explanations if it's reasonable to believe that the theory is true. This section argues that the paradoxical nature of certain proposition – whose closely related to Moore's paradox – support the normative constraint posited by (R4). I shall also contrast the case for (R4) with what I argue is a lack of similar considerations in favor of an empiricist alternative to (R4).

approximation in section 3 – see footnote 9). This is because "belief" is not a precise term, and consequently "reasonable to believe" is not precise either. One attitude that might be referred to by the term "belief" is a high degree of confidence in the believed proposition – one that can be represented as an assignment of a probability greater than some number $r \geq 0.5$. While I do think that the relevant kind of belief can be constructed out of degrees of confidence, I do not think that the realist thesis that one shouldn't appeal to a theory in explanations if one shouldn't believe it is plausible in that sense of "belief". One of several reasons this is not plausible is because the thesis would then fall prey to a version of the famous "Preface Paradox" (Makinson 1965): Suppose a historian has just finished writing a book on why Europeans emigrated to North America at the rate that they did. In the book, she appeals to various claims each of which she takes to be well supported by the data that she has gathered in her research. The claims in the book, *in conjunction*, constitute her explanation for the rate at which Europeans emigrated to North America. In the preface, aware of her own fallibility, she acknowledges that she has probably made at least one false claim in her book. So she believes that the conjunction of the claims made in the book – the conjunction which explains the rate of European emigration to North America – is most likely false. Unless we want to say that the historian is necessarily wrong to appeal to the conjunction she appeals to – and I see no reason to think so – we must admit that one may appeal to improbable theories in one's explanations. (See chapter 3 for a much more careful argument.)

Instead, I suggest that we conceive of the relevant kind of belief as roughly a matter of a *comparatively* high degree of confidence in the theory with which we explain, so that the requirement that one should only appeal to a theory in explanations if it is reasonable to believe that it's true – i.e. (R4) – comes to the requirement that one should only appeal to T in one's explanations if it's rational to place more confidence in T than in any explanatory rival to T, where a theory is an *explanatory rival* to another theory T just in case it is incompatible with T but answers all and only the explanation-seeking questions answered by T. On the "probabilist" assumption that degrees of confidence should be faithfully represented by some probability function $p(\cdot)$, this is equivalent to saying that one should not appeal to T in one's explanations unless T is more probable than any explanatory rival to T. Much more can be said about all of this, of course, and I do so in chapter 3.

2.6.1 Explanatory Moore-Paradoxes

I'll start by mentioning two well-known types of Moore-paradoxical sentences. Consider:

(1) p is true, but I don't believe p.[15]

There is clearly something paradoxical or incoherent about (1). But I won't be focusing on (1). Instead, let's consider another kind of Moore-paradoxical sentence – one that concerns not *belief*, *per se*, but *reasonable belief*:

(2) p is true, but it is not reasonable to believe p.[16]

Again, there is clearly something paradoxical or even incoherent about (2). But it's not that (2) is inconsistent, of course, so some broader kind of incoherence must be at work here. Without trying to pin down exactly what kind of incoherence that would be (a project that would take us too far afield), it's clear that any agent that assents to (i.e. either believes or asserts) (2) exhibits some epistemic failing purely in virtue of assenting to it. Putting the point slightly differently, assenting to (2) by itself shows the agent to have gone wrong in her epistemic housekeeping. In this respect, Moore-paradoxical sentences are akin to contradictions, for assenting to a contradiction also shows an agent to have gone wrong epistemically in a way that does not depend on any other facts about the agent, including what her evidence happens to be.

How is this relevant to the question of whether one should appeal to a theory in explanations only if it's reasonable to believe that the theory is true? Well, note that an

[15]This is the "omissive" form of this kind of Moore-paradoxical sentence. There is also a corresponding "commissive" form:

(1') p is true, but I believe $\neg p$.

In what follows, I focus on various Moore paradoxes of the "omissive" form, but nothing turns on this. So feel free to reconstruct each of the sentences below in a "commissive" rather than "omissive" form.

[16]The corresponding "commissive" form is:

(2') p is true, but it is reasonable to believe $\neg p$.

explanation is the sort of thing that may be expressed in an utterance of the form "T is the explanation for E" or "E because T".[17] Accordingly, we can construct Moore-paradoxical sentences of the same kind as (2) for explanations. For suppose I am trying to explain some fact E (the explanandum), and suppose that some theory T would, if true, explain E (perhaps in conjunction with auxiliary hypotheses). Now consider:

(3) T is the explanation for E, but it is not reasonable to believe that T is true.[18]

Clearly there is something wrong with (3), just as there was something wrong with (2). Consider also a somewhat more colloquial version of (3):

(4) E because T, but it is not reasonable to believe that T is true.

Again, (4) appears paradoxical in the same way as (2).

These examples are all rather abstract, of course, so let's see how these sentences strike us when we replace the variables E and T with something more concrete. In order to make contact with the issue that separates realists and empiricists, let's choose an example where the theory that is supposed to be doing the explaining posits unobservable entities. One popular example in the realism literature concerns the explanation of Brownian motion provided by the atomic theory of matter. So consider:

(5) The atomic theory is the explanation for Brownian motion, but it is not reasonable to believe that the atomic theory is true.

One might express (5) (or something very much like it) in a more natural way as follows:

(6) Brownian motion is explained by the existence of sub-microscopic particles moving around at random (i.e. atoms), but it is not reasonable to believe that there are such particles.

---

[17] Or perhaps an explanation is, at least in some cases, the act of uttering itself. (Bromberger 1965) What I have to say in what follows will be equally true either way.

[18] The corresponding "commissive" form is:

(3') T is the explanation for E, but it is reasonable to believe that T is false.

Again, it's clear that (5) and (6) are paradoxical in much the same way as (2).

Why are (3)-(6) paradoxical in this way? Well, recall that (R4) holds that one should only appeal to a theory in one's explanations if it is reasonable to believe that the theory is true. So, according to (R4), one asserts in the second part of each of (3)-(6) that one has violated a norm in the first part of the sentence. Put differently, if (R4) is correct, then each of (3)-(6) is a combination of an explanation and an admittance that one shouldn't have given that very explanation. By contrast, if one (R4) were false, then there should be nothing paradoxical about appealing to a theory in an explanation while at the same time admitting that it is not reasonable to believe it. (3)-(6) would not indicate that an agent that assents to them has gone wrong in her epistemic housekeeping, for whether it is reasonable to believe a theory to be true would be irrelevant as to whether she may appeal to it in the explanations she gives. So if (3)-(6) really are paradoxical in the same way as (2), then (R4) must be true.

2.6.2 Empiricist Alternatives

In my view the argument presented above already clinches the case for (R4), but it will be constructive to consider also an empiricist alternative to (R4). Suppose an empiricist attempts to block **the Realist Argument** by claiming that one may explain with a theory despite it not being reasonable to believe it to be true *as long as it is reasonable to believe it to be empirically adequate*. The obvious problem with this proposal is that it conflicts with the paradoxical nature of (3)-(6). Consider (6) in particular: Since the theory to which we appeal in explaining Brownian motion is about unobservable entities, there should be nothing paradoxical in appealing to it in an explanation and then immediately deny that it is reasonable to believe that it is correct in its claims about these unobservable entities.

I'll consider a different empiricist alternative to (R4) momentarily. Before I do so, let me mention another problem with the empiricist position we are now considering. Take any actual scientific theory T that posits unobservables, e.g. the atomic theory. It is easy to construct another theory by taking the conjunction of the claims that T makes about observable entities, T_O, and some set of claims about unobservable entities that is incompatible with what T says about unobservable entities. That is, for any theory that posits unobservables

$$T \equiv T_O \& T_U$$

where T_O includes all of T's claims about observable entities, we can construct an incompatible theory

$$T' \equiv T_O \& T'_U$$

where T'_U is incompatible with T_U. Thus T and T' are by construction incompatible yet empirically equivalent theories that posit unobservable entities.

Now, since T is by construction empirically adequate just in case T' is, the empiricist as we have so far construed her position is committed to claiming that if it is permissible to appeal to T one's explanations, then the same goes for the incompatible theory T'. Thus we have that on the empiricist alternative to (R4) that we are currently considering, one can be simultaneously warranted in appealing to two incompatible theories in explaining the same fact. Not only is this counterintuitive, but it also seems to conflict with scientific practice. Consider, for example, Fresnel's ether theory of light and Maxwell's electromagnetic theory. Each theory had its proponents at different times in history, but no one ever suggested that *both* theories may be appealed to simultaneously in explaining the behavior of light. Generally, if and when we have settled on an explanation for some fact, we take that to exclude any incompatible explanation for the same fact.[19] This strongly suggests that one cannot be simultaneously warranted in appealing to incompatible explanations of

[19]Except perhaps as an idealized or approximate explanation that serves some practical purpose not served by the non-idealized or precise explanation, as when we still to this day explain the behavior of the atom by appealing to the plum-pudding model. I do not take this to contradict my point here, since to my mind it merely illustrates that we are often quite happy to let non-epistemic considerations outweigh epistemic considerations.

the same facts.[20]

The empiricist may well (and fairly) respond to this argument by complaining that I have not adequately construed her position. As van Fraassen (1980) notes, anti-realists traditionally emphasize the role of pragmatic considerations in theory choice, and so an anti-realist empiricist may urge that a theory with which one explains should not only be such that it is reasonable to believe that it is empirically adequate, but it should also possess some combination of pragmatic virtues. For example, the anti-realist might insist that the atomic theory is an elegant way of conceptualizing Brownian motion and other empirical consequences of the theory – more elegant than any alternative way of conceptualizing these empirical consequences. So on this second empiricist view one may appeal to T in one's explanations only if T provides the most practical explanations among the theories that it is reasonable to believe to be empirically adequate.

The problem with reconstruing the empiricist position in this way is brought out once we return to our original concern with Moore-paradoxical sentences. For suppose as before that T and T' are empirically equivalent yet incompatible theories about unobservables. The current empiricist position predicts that the following should come out as paradoxical:

(7) T is the explanation for E, but T is not more practical than T'.

However, (7) is clearly not paradoxical in the same way as (7). Of course, we might feel *disappointed* by (7) – because we might have hoped that the (correct) explanation for E was the most practical (potential) explanation available. But we don't feel that a person uttering (7) is confused or in the wrong in the same way as the person uttering (3)-(6). We can also construct more specific versions of (7), e.g. by replacing "practical" with specific pragmatic virtues:

(8) T is the explanation for E, but T is not *simpler* than T'.

[20] Admittedly, there are many cases from the history of science in which scientists have tolerated inconsistencies between theories they accept (as when scientists accepted both classical electrodynamics and Bohr's model of the atomic). But it's one thing to tolerate inconsistencies in the set of theories one appeals to in explanations, e.g. because one fails to see how to achieve consistency, and quite another to say (as the empiricist must say) that one is under no imperative to make one's explanations consistent.

(9) T is the explanation for E, but T is not *more elegant* than T'.

(10) T is the explanation for E, but T is not *more convenient* than T'.

None of these sentences exhibit the paradoxical nature characteristic of Moore-paradoxes. But if the (second) empiricist position is correct, they should. So this empiricist position must be wrong.

Let us take stock. I have argued, first of all, that (R4) must be true given the existence of "explanatory Moore-paradoxes" such as (3)-(6). From this I conclude that (R4) is true, and thus that **the Realist Argument** is sound. In addition, in an attempt to analyze the issue further, I examined two possible empiricist alternatives to (R4). On the one hand, I argued that a *permissive* empiricist position – on which one may appeal to any theory in one's explanations as long as it is reasonable to believe that it is empirically adequate – not only fails to account for the paradoxical nature of (3)-(6), but also fails to account for the fact that scientists avoid giving incompatible explanations of the same facts. On the other hand, I argued that a *non-permissive* empiricist position – on which one should only appeal to a theory in explanations if provides the most practical explanation among those that it is reasonable to believe to be empirically adequate – neither accounts for the paradoxical nature of (3)-(6) nor for the non-paradoxical nature of (7)-(10).

2.7 Conclusion

We started with a conception of scientific realism proposed by van Fraassen (1980), one according to which a realist holds that a belief that T is true is necessary for acceptance of T, while an empiricist holds that only belief in T's empirical adequacy is necessary for acceptance of T. However, we saw that if we adopt a plausible story about how to distinguish between "acceptance" and "belief" (and we need some such story on pain of trivializing the debate, as the Blackburn-Horwich-Mitchell-Teller objection shows), then it turns out that acceptance does not require any belief at all. Accordingly, I suggested that realism and its competitors should be conceived of as theses about the *normative* connection between acceptance and belief. On this conception of the scientific realism debate, we were able to make sense of the empiricist's emphasis on the distinction between observable and unobservable entities. However, we also saw that more is involved in acceptance than making empirical predictions – to accept a theory also involves appealing to it in scientific explanations. Ultimately, this point became the bane of anti-realists, since the existence of Moore-paradoxical sentences for explanations strongly suggests that appealing to a theory in explanations requires that it be reasonable to believe what the theory says about unobservable entities.

So the two main upshots of this paper are, first, that there is non-trivial issue of realism and anti-realism about scientific acceptance, and second, that a strong case for a realist view of this issue can be made. This provides a rather more satisfying kind of realism than the kind espoused by Mitchell (1988) and Horwich (1991), for example, who argue that realism is true in virtue of being the only game in town. Indeed, on the current conception of the debate, scientific realism is a quasi-empirical thesis (as Putnam (1978) influentially argued – though in a different way than Putnam envisaged), for it could have turned out that the scientific enterprise involved nothing beyond making empirical predictions, in which case an empiricist view of acceptance would have been true. As it happens, making empirical predictions is not all there is to doing science, and this fact provides us with the resources for an argument for scientific realism.

3 EXPLANATORY ACCEPTABILITY

Abstract

Some proposed explanations are defective not because the explanans fails to stand in an explanatory relationship with the explanandum but because we lack reasons to believe the explanans. The hypotheses to which we appeal in such proposed explanations are not *explanatorily acceptable*. But what makes hypotheses explanatorily acceptable in this sense? One *prima facie* plausible answer is that a hypothesis is explanatorily acceptable just in case its probability exceeds some threshold. However, I argue that – for reasons closely related to the so-called Preface and Lottery paradoxes – exceeding a probability-threshold is neither necessary nor sufficient for being explanatorily acceptable. In an attempt to diagnose the problem further, I point out that this threshold view of acceptability is essentially a "satisficing" (as opposed to "optimizing") view of explanatory acceptability. This points towards an alternative account, one according to which an explanatorily acceptable hypothesis must be (significantly) more probable than its explanatory rivals. By proving a few theorems I show formally how such an account avoids the problems with the threshold view. Finally, based on this account, I suggest a way to further reconcile Bayesianism and Inference to the Best Explanation.

3.1 Introduction

Carl. G. Hempel and Paul Oppenheim began their seminal "Studies in the Logic of Explanation" by emphasizing that seeking explanations is an integral part of natural science:

> To explain the phenomena in the world of our experience, to answer the question "why?" rather than only the question "what?", is one of the foremost objectives of all rational inquiry; and especially, scientific research in its various branches strives to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates. (Hempel and Oppenheim 1948, 135)

Since this was written, a lot of ink has been spilled over what it takes for some hypothesis to explain something. However, not all hypotheses that would (if true) explain something provide *reasonable* explanations. For example, it would not be reasonable to appeal to the hypothesis that the moon is made of cheese to explain its color and shape, even if the former (if true) would explain the latter. So what makes some explanations reasonable and others not so much? This paper is an attempt to answer that question – what I'll call the *epistemic question of explanation*.[1]

There are at least four separate reasons for discussing the epistemic question of explanation. First, if providing explanations is indeed one of the primary goals of rational inquiry (including natural science), then the epistemic requirements on explanations are surely of interest to scholars of rational inquiry, most notably epistemologists and philosophers of science. Second, as I argue elsewhere, the epistemic question of explanation has repercussions for the kind of scientific realism famously discussed by van Fraassen (1980): Since the *acceptance* of a scientific theory involves using it in scientific explanations, a realist answer to the epistemic question of explanation leads to a realist view of scientific acceptance. (See chapter 2.) Third, once we have an answer to the epistemic question of

[1]The epistemic question of explanation is alluded to in David Lewis's classic paper "Causal Explanation", where he lists the following as one of the good-making features of an act of explaining:

> The explanatory information provided may be correct, but not thanks to the explainer. He may have said what he did not know and had no very good reason to believe. If so, the act of explaining is not fully satisfactory, even if the information provided happens to be satisfactory. (Lewis 1986, 227)

explanation on the table, we can evaluate the merits or otherwise of the controversial inferential rule known as "Inference to the Best Explanation" (IBE). (More on this in section 7 below.) Finally, I argue elsewhere that there is a kind of evidence the value of which can only be appreciated given its role in producing reasonable explanations. (Dellsén a)

The plan of the paper is as follows. In section 3, I present and argue against a *prima facie* plausible answer to the epistemic question of explanation, viz. that an explanation is reasonable just in case its probability exceeds a threshold. In section 4, I point to a very different way to explicate the relationship between reasonable explanation and probability, contrasting "satisficing" and "optimizing" views. In section 5, I then present an "optimizing" view of reasonable explanation, the Optimality Account, roughly according to which an explanation is reasonable just in case it is (significantly) more probable than its explanatory rivals. Section 6 shows how this avoids the problems of the "threshold view" in a natural and elegant way. Section 7 compares this account to an "Explanationist" (i.e. IBE-based) account, arguing that the Optimality Account provides a plausible probabilistic *reinterpretation* of IBE.

3.2 Preliminaries

Let's start by getting a better grip on the question that I'm trying to answer – the one I referred to as the "epistemic question of explanation". I said before that not all hypotheses that would (if true) explain something we take ourselves to know are such that it would be *reasonable* to appeal to them in explanations. I want to remain as neutral as possible on what it is for something to be reasonable, except that "reasonable" (in the sense in which I am interested) is a term of epistemic appraisal (much like "warranted", "rational" and "justified").[2] Note that my focus will be on *epistemic* appraisals of explanations, not e.g.

[2]One might worry that since explaining some fact is a kind of action, talk of epistemically reasonable explanations is a kind of category mistake. (An extreme version of this worry would state that only beliefs can be epistemically evaluated. More modestly, one might think that only psychological states can be epistemically evaluated.) A full reply to this worry lies outside the scope of this paper, but let me just note that it is nearly uncontroversial that one type of action can be *epistemically* evaluated, viz. *assertions*. (See e.g.

prudential or moral appraisals. This is not to deny that there is a sense of "reasonable" according to which it would be utterly reasonable to appeal to a theory one knows to be false in one's explanations, e.g. if doing so serves some pedagogical purpose (as when a grade school teacher appeals to the "plum pudding" model of the atom in explaining how atoms give and receive electrons). Rather, it is to say that I am interested only in one particular sense of "reasonable", viz. the epistemic sense.[3]

To explain is always to explain something in particular. However, I am not interested in what makes it reasonable to appeal to some hypothesis in some particular explanation. Rather, I am interested in what makes it reasonable to appeal to some hypothesis in any proposed explanation where the explanandum would indeed be explained by the hypothesis, if true. In this sense, which hypotheses it is reasonable to appeal to in one's explanations is not relative to the phenomena that are being explained at a particular time. For convenience in what follows I will adopt the following terminological convention: If it is (epistemically) reasonable to appeal to a hypothesis H in one's proposed explanations, I'll say that H is *explanatorily acceptable*, or sometimes just *acceptable*. The set of acceptable hypothesis for a given agent is thus, for the purposes of this paper, the set of hypothesis that it would be reasonable for that agent to appeal to in her explanatory endeavors.

This definition of "explanatory acceptability" is closely related to various recent definitions of "acceptance", in that a hypothesis is explanatorily acceptable roughly if it is reasonable to accept that hypothesis. Consider, first of all, the sense of "acceptance" used in the debate over scientific realism as characterized by van Fraassen (1980): Since one of the things that a scientists will do with a theory that she accepts is to explain with it, one should not accept a theory unless it is explanatorily acceptable. (See chapter 2.)

(Williamson 1996).) I take explanations to be actions of the same general category as assertions, which lends significant plausibility to explanations being epistemically evaluable as well.

[3]There is arguably both a "binary" and an "analog" concept of reasonableness – we say that something is "reasonable" and "not reasonable", indicating that reasonableness is binary, but we also say that some things are "more reasonable" than others, and that some things are "somewhat reasonable", "very reasonable", etc., indicating an analog sense. In this paper I will focus on the binary sense, although many of the lessons will carry over to an analog sense as well.

Consider also the sense of "acceptance" appealed to by some broadly-speaking Bayesian epistemologists, according to which one accepts H just in case one would assert H under some specified circumstances. (de Sousa 1971; Kaplan 1981b,a; Maher 1993; Lance 1995) Since appealing to H in an explanation involves asserting H, reasonable acceptance (in this sense) would seem to require explanatory acceptability (in my sense).[4]

Much of the discussion that follows appeals to probabilities, which raises the issue of how to interpret these probabilities. Throughout in this paper I will take the probability of H for an agent S, $p_S(H)$, to be the credence that it is rational for S to assign to H.[5] (In most of what follows, since nothing turns on whose probabilities are being referred to, I shall often write $p(H)$ instead of $p_S(H)$.) This means that I am assuming the "probabilist" thesis that rational credences must be probabilistically coherent at a time. Of course, this is an extremely demanding requirement – something that most, if not all, ordinary agents will fail to satisfy – and should be seen as a useful idealization to help us focus on the central issue at hand. Indeed, I shall throughout make a related idealization, viz. that rational agents assign probabilities to every hypothesis in logical space. I believe that this requirement can be relaxed fairly easily, but in order to keep the discussion simple and focused I shall not attempt to do so here.

3.3 The Threshold View

Let us start by examining a natural and *prima facie* plausible view of explanatory acceptability, viz., that a hypothesis is acceptable just in case its probability exceeds some threshold. A bit more precisely, the idea is that a hypothesis H is acceptable for an agent S just in case S is rational in having some credence in H that exceeds some number t. I'll

[4]Obviously, more needs to be said about the connections between these Bayesian conceptions of "acceptance" and my own definition of "explanatory acceptability" – more than I can hope to do here – but it should be clear that being reasonable in accepting a hypothesis is closely linked with explanatory acceptability.

[5]Note that, on this interpretation, the probability of H for an agent S does not represent the *actual* credence that S assigns to H. Even if 0.7, say, is the credence that it is rational for S to assign to H, S may have some credence other than 0.7 in H, or indeed no (precise) credence at all.

refer to this as *the Threshold View*. This view is a close analog of the so-called Lockean View of (full) belief, according to which (full) beliefs are credences above a certain threshold. (Foley 1992) Indeed, the Threshold View derives a certain initial plausibility from the Lockean View on the assumption that a hypothesis is acceptable just in case it is rational to have a (full) belief in the hypothesis.[6]

Although the Threshold View (as I've characterized it) is not committed to any particular value for the threshold t, it's clear that t must be lower than 1. Otherwise no hypothesis would be acceptable unless it would be rational to assign the same credence in it as one is justified in assigning to tautologies, which presumably entails that no empirical hypotheses would be acceptable. It is also natural to require the threshold not to be lower than 0.5. The motivation for this comes from the fact that when $p(H)$ goes below 0.5, the probability of the negation of H is greater than the probability of H. In what follows, however, I won't be assuming this requirement on the threshold t. So, as far as a the following discussion is concerned, we will leave open what the probability threshold t is, except only to require that $t < 1$.

So proponents of the Threshold View may disagree about the value of the relevant threshold. Indeed, they may disagree on a more fundamental point, viz. whether the threshold is always the same or whether it varies depending on context or stakes. To mark this

[6]It may initially seem plausible for someone sympathetic to the Threshold View to claim that for a hypothesis H to be acceptable, it suffices that the probability that H is *approximately* true exceed t. One problem with this version of the Threshold View is that it introduces the notoriously problematic notion of "approximate truth" (also known as "truthlikeness" or "verisimilitude"). But even if the problems with approximate truth could be solved, I think that this use of the concept of approximate truth is quite problematic. Recall that a hypothesis is (by definition) explanatory acceptable only if it is reasonable to appeal to it in *any* explanation (provided that the theory would explain the explanandum). Put differently, an explanatorily acceptable hypothesis is one that is available as a premise in any given explanation. Now, one may know that a hypothesis is *approximately* true yet also know that as part of the explanation of some explanandum the theory is widely off the mark. Suppose, for example, that Newtonian Mechanics is approximately true (as is widely claimed by proponents of approximate truth). There are many explananda, e.g. Mercury's perihelion, which it would be unreasonable in the extreme to explain by appealing to Newtonian Mechanics. This seems to me to conclusively refute the suggestion that explanatory acceptability requires only that the probability of a hypothesis being *approximately* true exceed some threshold. Hence appealing to approximate truth seems to me to not be promising in an account of explanatory acceptability.

difference, let us distinguish *rigid* and *nonrigid* views: On rigid threshold views, the threshold t is always the same, whereas on nonrigid threshold views, the threshold t varies either with context or with stakes. Although this is an important distinction that arguably makes the Threshold View more plausible, my criticism of the Threshold View will not depend on which version of the view one adopts. A proponent of the Threshold View might also insist that "acceptability" is a vague concept and thus that the threshold between acceptable and not-acceptable hypotheses is itself vague. On such a view, it is natural to think of the threshold not as a number but as an interval, I_t, where a hypothesis counts as acceptable just in case its probability exceeds the upper bound of I_t, not-acceptable just in case its probability falls below the lower bound of I_t, and indeterminate otherwise. Again, my criticism of the Threshold View will apply to both vague and precise versions of the view.

So the Threshold View (as I am presenting it) is not by itself wedded to there being any particular threshold for explanatory acceptability, nor is it wedded to the threshold being either rigid or nonrigid, vague or precise. However, I'll now argue that no matter how the Threshold View is spelled out along these dimensions, the probability of a hypothesis exceeding such a threshold is neither necessary nor sufficient for acceptability as explanation.

3.3.1 The Preface Paradox

I start by arguing that the probability of an hypothesis exceeding some threshold is not necessary for acceptability. The argument is inspired by the Preface Paradox presented by Makinson (1965), and is closely related to an argument given by Kaplan (1981b, 1996). Consider first an example similar to Makinson's original case: A historian has just finished writing a book on why Europeans emigrated to North America at the rate that they did. In the book, she appeals to various claims each of which she takes to be well supported by the data that she has gathered in her research. Thus let us suppose that each of the claims in the book is such that its probability exceeds the relevant threshold t, whatever it is. Yet we may also suppose (since it is consistent with what we have said so far) that the probability of the conjunction of the claims in the book does not exceed this threshold t.[7] Thus we have that each of the claims in the book is acceptable while their conjunction is not. Since the conjunction of any claims is a logical consequence of those claims, we have that on the Threshold View some hypotheses may be acceptable while one of their logical consequences is not.

What this shows is that the Threshold View conflicts with two plausible principles. First, it conflicts with the principle that if some hypotheses are acceptable, then their conjunction is acceptable:

(&) If $\{H_1, \ldots, H_n\}$ is a set of hypotheses each of which is acceptable, then the conjunction $(H_i \& \ldots \& H_n)$ is acceptable.

Second, and more generally, the Threshold View is also shown to conflict with the stronger principle that if some hypotheses are acceptable, then their logical consequences are acceptable:

(D*) If $\{H_1, \ldots, H_n\}$ is a set of hypotheses each of which is acceptable, then any logical consequence of $\{H_1, \ldots, H_n\}$ is acceptable.[8]

Strictly speaking, this last principle – (D*)[9] – is a bit too strong, for a logical consequence of some hypotheses may not even, if true, provide an an explanation for anything. For example, the *disjunction* of General Relativity and any other proposition follows from General Relativity, but presumably at least some such disjunctions explain nothing at all. To count

[7]To see why this is consistent with what we have said so far, note that for any set of probabilistically independent hypotheses $\{H_1, \ldots, H_n\}$, the probability of their conjunction is given by:

$$p(H_1 \& \ldots \& H_n) = p(H_1)p(H_2)\ldots p(H_n)$$

Suppose, for example, that $p(H_1) = p(H_2) = p(H_n) = 0.95$ and that $n = 100$. Then $p(H_1 \& \ldots \& H_n) \approx 0.0059$.

[8](&) follows from (D*) since any set of proposition has the conjunction of those propositions as a logical consequence. The converse does not hold since not all logical consequences of some set of hypotheses are conjunctions of those hypotheses.

[9]The asterisk is a mark of tentativeness to be removed momentarily.

such "hypotheses" as explanatorily acceptable would be at best unnatural. Fortunately, we can easily restrict principle (D*) to what I shall call *explanatorily potent* hypotheses, where a hypothesis is explanatorily potent just in case it would, if true, explain something.[10] Thus we get:

(D) If $\{H_1, \ldots, H_n\}$ is a set of hypotheses each of which is acceptable, then any explanatorily potent logical consequence of $\{H_1, \ldots, H_n\}$ is acceptable.[11]

Although (&) and (D) are intuitively plausible, one might worry that a sufficiently large number of claims would not be acceptable despite each one being so, and thus that there may be exceptions to (&) and (D). For the record, I don't think that's correct, but in any case that wouldn't save the Threshold View. For note that I did not specify the number of claims contained in the book in the above example. So consider just *two* claims, where the probability of each claim exceeds the threshold t (whatever it is). The probability of their conjunction need not exceed the threshold as well, so again we have that each claim may be acceptable while their conjunction is not. And again this entails that some hypotheses – in this case, a pair of hypotheses – may be acceptable as explanations while one of their logical consequences is not.

To really see the force of this objection, suppose H_1 is an acceptable explanation of E_1 and H_2 is an acceptable explanation of E_2. Suppose further that the conjunction of H_1 and H_2, $(H_1 \& H_2)$, provides an elegant unification of H_1 and H_2.[12] If the Threshold View is correct, each of H_1 and H_2 may be acceptable as explanations while the theory

[10]Later I will precisify this notion of explanatory potency in terms of answering some explanation-seeking question.

[11]Again this entails (&) on the assumption that any conjunction of acceptable hypotheses is explanatorily potent. To see why that assumption is true, note that a conjunction of two theories each of which explain something must at least explain the things explained by each of the conjuncts. (One might worry that this fails when the conjuncts are mutually inconsistent, since the conjunction would then be internally inconsistent and perhaps therefore explanatorily impotent. But I shall argue below that explanatorily acceptable theories cannot be mutually inconsistent, and hence that the case in question is impossible.)

[12]Of course, an elegant unification presumably would not take the form of a conjunction of two hypothesis, but it could be equivalent to such a conjunction. Since probabilities of equivalent hypotheses are equal, the the probability of such an elegant unification would be equivalent to the probability of the conjunction.

that unifies them $(H_1 \& H_2)$ is not, since the conjunction of any two hypotheses may be less probable than each of the conjuncts. But this is clearly absurd. Moreover, it goes against the well-established scientific practice of seeking more general and unified explanations of natural phenomena: If the Threshold View is correct, we could ensure that our theories are acceptable by making sure we split them into smaller and less ambitious theory-bits rather than appealing to the unified theory.

Note that the soundness of this argument does not depend on whether we take the threshold to be rigid or nonrigid. After all, we can easily fix the context and stakes by stipulating that the theories and their conjunction are all being evaluated in the same context and that the stakes are kept constant throughout. Thus the threshold t will be the same for each of the hypotheses and their conjunction. Similarly, it doesn't matter whether the threshold is precise or vague: If the threshold is an interval as opposed to a number, then we can imagine a set of claims such that the probability of each claim exceeds the upper bound of the threshold-interval I_t, yet such that their conjunction does not.

3.3.2 Improbable Explanations

Let me pause to address an objection. I've argued that exceeding a probability-threshold is not necessary for acceptability. One might push back against this conclusion by pointing out that if no particular probability is necessary for acceptability then even a very improbable hypothesis might be acceptable in some cases.[13] That, one might think, is a consequence that we cannot live with. Better then to give up on the principles that led us to such an absurd conclusion, (&) and (D). In response, I admit that it does sound somewhat strange to say that an improbable hypothesis might be acceptable, i.e. reasonably appealed to in explanations. But the reason this is so, I submit, is due to a systematic ambiguity in the term "probability". Let me explain.

So far in this paper, we have taken "probability" to refer to quantities defined by the

[13]Note though that even if no particular probability is necessary for acceptability most acceptable hypotheses may be very probable indeed.

axioms of the probability calculus, i.e. the Kolmogorov axioms. However, there are good reasons to think that the layperson's sense of "probability" is often quite radically divorced from this precise mathematical concept.[14] Consider the well-known study by Tversky and Kahneman (1984). Given a description of Linda, who has a history of left-leaning political activism, around 90% of respondents judged that it was more probable that Linda was a feminist bank teller than that she was a bank teller *tout court*. Of course, this violates the rule not to assign a higher probability to a conjunction than to one of the conjuncts – one of the most obvious theorems of Kolmogorov's axioms. Interestingly, even in a statistically sophisticated group of respondents – consisting of doctoral students in the decision science program at the Stanford Business School, who had taken several courses in probability, statistics and decision theory – 85% of the respondents gave the same answer. So if the respondents were using "probability" in the sense defined by Kolmogorov's axioms, then 85% of the most sophisticated respondents were guilty of a quite obvious error.

It is more plausible that these respondents meant something slightly different by "probability". Exactly what that is is hard to say. Fortunately, all that matters for our purposes is that their judgments do not appear to concern the quantities that are defined as conforming to Kolmogorov's axioms. Thus, whatever they meant by "probability", it may well be true that, *in that sense*, one cannot be reasonable in accepting something improbable as an explanation. So while it is true that it sounds strange to say that one improbable hypotheses can be acceptable, I submit that this should be interpreted as employing a layperson's sense of "probable" – a sense that refers to things that do not conform to the Kolmogorov axioms and so is not the sense of "probability" that I have been operating with in this paper.[15]

[14]Consider an analogy: There is a mathematically precise definition of "line", according to which lines are infinitely long. So are the folks wrong when they say that they drew a line in the sand, for example? (It wasn't infinitely long!) In my view, "probability" is much like "line", in that it is ambiguous between a precise mathematical concept and a non-precise everyday concept.

[15]That "probability" is used in more than one sense in this way is widely recognized. See, for example, (Cohen 1980) and (Lycan 2012).

There may be another reason to think that improbable hypotheses cannot be acceptable. We use hypotheses to make predictions (and retrodictions) about how the world will behave (or has behaved). If an acceptable hypothesis need not be probable, then how could we be reasonable in using accepted hypotheses for this purpose? The response to this worry is that explanatory acceptability does not concern whether one would be reasonable in using a hypothesis in some prediction (or retrodiction), but rather whether it would be reasonable to use it in an explanation. Indeed, as various Bayesian theorists have argued, it is a mistake to think that the use of theories in prediction (and, more generally, rational decision-making) requires that one assign any given probability to the theory with which one predicts. (Jeffrey 1956, 1970; Maher 1986, 1993; Kaplan 1996)

3.3.3 The Lottery Paradox

Having argued that exceeding a probability-threshold is not necessary for acceptability, I now want to argue that it is not sufficient either. My argument is closely related to the Preface Paradox's "sister paradox", the *Lottery Paradox* (first proposed by Kyburg (1961)). This paradox, appropriately modified to fit the topic of acceptability (as opposed to the related but distinct topics of knowledge and rational belief), goes as follows: Suppose you are almost certain that there was a lottery drawing yesterday, and that this lottery was a fair one with 1000 tickets. Let T_i be the proposition that ticket number i lost. Being almost certain that the lottery was fair, you assign $p(T_i) = .999$ for all $1 \leq i \leq 1000$. Yet, again because you are almost certain that the lottery was fair, you assign probability close to 1, say .999, to the proposition T_s that that *some* ticket won, i.e. that not all tickets were losing tickets. So if a hypothesis is acceptable if its probability exceeds some threshold (and if the threshold is lower than 0.999) then $T_1, T_2, \ldots, T_{1000}$, *and* T_s are all acceptable.

The problem, of course, is that these propositions are inconsistent, so the Threshold View entails that a set of propositions that are jointly inconsistent would be acceptable. This goes against the following principle:

(C) If each H_i in $\{H_1, \ldots, H_n\}$ is acceptable, then $\{H_1, \ldots, H_n\}$ is consistent.

That is bad enough, but things get worse: Given (&) (or (D)), it follows from each of the claims being acceptable that their conjunction – a contradiction – is acceptable. Furthermore, note that since a contradiction necessarily has probability 0, this conflicts with the necessity claim of the Threshold View (assuming of course that the threshold is not 0). Now, one might doubt that (&) (or (D)) hold true in all cases. Even without (&) (or (D)), however, it is quite implausible to say that a set of propositions could all be acceptable even if they are jointly inconsistent. Yet that is what the sufficiency claim of the Threshold View entails.

One might think that the Lottery Paradox can be resolved by making a minor modification to the Threshold View. This thought starts by pointing out that the probability assigned to each hypothesis T_i is based on a known objective chance (greater than zero but less then one), namely the objective chance that any particular ticket wins in a fair lottery with 1000 tickets. This suggests that perhaps both the Threshold View can be rescued from the Lottery Paradox by adding the qualification that hypotheses are not acceptable if they are based on known objective chances in this way.[16]

This modification won't do the trick, however, because there are variations of the Lottery Paradox that have nothing to do with objective chances. It's easy to see what such variations would look like: Suppose you have a set of 1000 propositions, $\{H_1, ..., H_{1000}\}$, none of which are based on known objective chances. Suppose further that you are almost certain that $\{H_1, ..., H_{1000}\}$ contains at least one false claim, and thus assign a probability close to 1, say .999, to the proposition that the conjunction $(H_1\&...\&H_{1000})$ is false. And

[16]In this vein, Nelkin (2000) suggests that propositions like T_i cannot be rationally believed because they are based on what she calls "P-inferences", where a P-inference is an inference to a proposition p from p having a high "statistical probability". Now, I argue below that this won't do as a solution to the Lottery Paradox, but some have found this requirement to be independently plausible. That is, one might think that whatever else is true of acceptability, a theory cannot be acceptable if its probability is merely "statistical" in Nelkin's sense. While I do not find this restriction plausible, I don't want to alienate the reader who does find it plausible. Fortunately, we can set this restriction aside in what follows since any account of acceptability that appeals to probabilities (including the one I present below) may simply adopt this restriction in addition to whatever else the account requires of acceptable theories.

suppose you have no reason to think that one of the claims in $\{H_1, ..., H_{1000}\}$ is more likely to be false than any other, but you do have very, very good reasons to think that each claim is true. So let's suppose that $p(H_i) = .999$ for all $1 \leq i \leq 1000$. Again we have that, if a hypothesis is acceptable just in case its probability exceeds a threshold (and if that threshold is below 0.999), an inconsistent set of hypotheses could be such that all of the hypotheses are acceptable. This violates (C) of course. Moreover, (&) (or (D)) again entails that a contradiction would be explanatorily acceptable.

In the next subsection, I will address a worry about (C) having to do with the fact that scientists in fact often accept inconsistent theories. For now, note that as with the Preface Paradox, the Lottery Paradox applies even if the threshold is nonrigid. For we can simply fix the context and stakes so that the same threshold applies to all of the relevant propositions T_1, T_2, ..., T_{1000}, and T_s). The same goes for a version of the Threshold View on which the threshold is vague: If so, i.e. if the threshold is an interval as opposed to a number, then one can still suppose that all the hypothesis involved have probabilities above the upper bound of that threshold-interval. So nonrigid and/or vague thresholds do not allow the Threshold View to avoid the Lottery Paradox.

3.3.4 Deductive Cogency

Where does this leave us? One moral of our discussion so far is that exceeding some probability threshold is neither necessary (because of the Preface Paradox) nor sufficient (because of the Lottery Paradox) for acceptability. The more general moral from these paradoxes, however, is that any account of acceptability should make acceptability accord with the principles appealed to above, (D) (and thus (&)) and (C). Putting these together, we get the following requirement on acceptability:

Deductive Cogency: If $\{H_1, \ldots, H_n\}$ is a set of hypotheses each of which is acceptable, then

(i) any explanatorily potent logical consequence of $\{H_1, \ldots, H_n\}$ is acceptable, and

(ii) $\{H_1, \ldots, H_n\}$ is consistent.

What the Preface and Lottery paradoxes in effect show is that the Threshold View violates *both* clauses of **Deductive Cogency**: The Preface shows that it violates (i), while the the Lottery shows that it violates (ii). So, in sum, the problem with the Threshold View is that it fails doubly to make acceptability deductively cogent.

Let me pause here to consider an objection to **Deductive Cogency**, in particular to clause (ii). It might seem that (ii) conflicts with the fact that there are cases of accepted scientific theories that are mutually incompatible. For example, classical electrodynamics and Bohr's model of the atom are mutually incompatible, yet they were simultaneously accepted by the scientific community at one point. Similarly, general relativity and quantum mechanics are widely taken to be mutually incompatible, yet both are currently accepted. Indeed, there may even be examples of theories that were once accepted but that are internally inconsistent, such as Dirac's original formulation of quantum mechanics, which used a mathematically inconsistent definition of the "Dirac delta function". Many other examples of this kind can be given, and so there can be no denying that scientists have sometimes accepted inconsistent theories.[17]

However, these cases, interesting as they may be, do not conflict with (ii). For recall (from section 2) that we defined explanatory acceptability in terms of whether it is *epistemically* reasonable to appeal to a theory in one's explanations. That is compatible, of course, with scientists having excellent non-epistemic reasons to accept or explain with a set of theories that are, strictly speaking, explanatorily unacceptable. For example, it would be foolish at best to reject either general relativity or quantum mechanics without viable theories to replace them with, but that does not mean that appealing to both theories in one's explanations is strictly speaking epistemically reasonable. What it shows is that given our

[17] I am grateful to John Roberts for calling my attention to these sorts of cases, and for pressing the objection on their basis.

epistemic limitations, we ought sometimes be prepared to sacrifice epistemic perfection in order to satisfy other important goals, including the goal of having informative theories.

So my claim is not that **Deductive Cogency** is an overriding or all-things-considered normative requirement on the set of hypotheses with which we should be prepared to explain. **Deductive Cogency** is best viewed as an ideal to which we should aspire – one that may very well be unreasonable to expect an agent to live up to in many circumstances, especially when due to our cognitive limitations we fail to see how to modify our theories in light of it. Kaplan put the point well in discussing an analogous requirement for (full) beliefs:

> [...] you should want to satisfy Deductive Cogency in the following sense: you should view a demonstration that a set of beliefs violates Deductive Cogency as a criticism of that set of beliefs (as, presumably, you would not view a demonstration that a set of beliefs implies that the earth is not flat) - a criticism that can only be met by revising that set of beliefs. Deductive Cogency is best understood as simply a condition that your set of beliefs must satisfy on pain of being open to criticism. (Kaplan 1995, 118)

This, *mutatis mutandi*, is how I understand the nature of **Deductive Cogency** as a normative requirement upon the set of theories we should be prepared to appeal to in explanations.

To be sure, there are still those who conclude from the fact that the Threshold View conflicts with **Deductive Cogency** that the latter, as opposed to the former, should be rejected.[18] One might be tempted to plump for such a conclusion as a last resort given the plausibility of explicating acceptability in terms of probability and the enormously fruitful probabilistic approach in epistemology. But that would be too hasty, for the Threshold View is not the only way of explicating acceptability in terms of probability. It is true that developing an probability-based account of acceptability that isn't just "rigged" to satisfy **Deductive Cogency** by *ad hoc* stipulation will not be easy. Fortunately, I will argue in what

[18]In the case of the threshold view of full belief, Kyburg (1961, 1970) and Foley (1979, 1992) both take this route.

follows that by replacing the "satisficing" approach of the Threshold View with an "optimizing" or "maximizing" approach, we open the door to a natural account of acceptability that turns out to accord with **Deductive Cogency** without any *ad hoc* maneuvers. The next section introduces the basic thought behind this idea.

3.4 Satisficing versus Optimizing Views

Here are two ways to positively evaluate something, e.g. a cup of coffee: On the one hand, we might think that the cup of coffee is *sufficiently good* for some purpose, e.g. for having something tasty, or for causing us to wake up properly. On the other hand, we might think that the coffee is a *better* in some respect than some relevant alternatives. For example, we might think that the cup of coffee in front of us is tastier, or more a more effective stimulant, than any other cup of coffee within a 10 mile radius. These are quite different ways to evaluate a cup of coffee: On the former, it does not matter how the coffee compares to something else – the coffee is evaluated positively just in case it meets some demands we make for cups of coffee. Call this a *satisficing* evaluation.[19] On the latter kind of evaluation, it matters how the coffee compares to something else – the coffee is evaluated positively just in case it is better than any relevant alternative. Call this an *optimizing* evaluation.

Note that we can have satisficing and optimizing evaluations of the same thing using the same standards of evaluation: The coffee might be sufficiently good in that it is quite *tasty*, and also better than any alternative in that it is *tastier* than any alternative. In both cases the evaluation of the coffee depends on tastiness and tastiness alone. So what separates satisficing and optimizing evaluations is not that there are different standards for evaluation. Rather, it is the *structure* of the evaluation that is different between satisficing and optimizing evaluations: Satisficing options are *good enough*, while optimizing options are *better than alternatives*. Note also that an option might be satisficing without being

[19] This term is borrowed from a decision theoretic heuristic proposed and developed by Simon (1956).

optimizing and *vice versa*, even relative to the same standards: The cup of coffee might not be the tastiest coffee available, but still be sufficiently tasty. Conversely, the coffee might indeed be the tastiest available, but still not sufficiently tasty.

What has this distinction between satisficing and optimizing evaluations got to do with explanatory acceptability? Well, notice that the Threshold View is a satisficing view of one kind of evaluation, viz. epistemic evaluation of a hypothesis for the purposes explanation. The standard of evaluation is *probability*, of course, since it is the probability of a hypothesis that must be sufficiently high in order for it to be acceptable on the Threshold View. As I've mentioned, one might be tempted to give up on probability-based views altogether given the shortcomings of the Threshold View. But we are now in a position to realize that the Threshold View is not the only possible way of explicating acceptability in terms of probability: We can instead develop an *optimizing* view of acceptability using probability as our standard for evaluation.

To see the promise of such an account, consider the following piece of scientific fiction. Suppose Darwin is trying to explain the fact that organisms evolve over time, and that there are 100 potential explanations for this fact to which Darwin assigns non-zero probabilities. H_1 is Darwin's hypothesis of natural selection, H_2 is Lamarck's hypothesis of inheritance of acquired characteristics, and H_3-H_{100} are some other hypotheses that would, if true, explain biological evolution. Now suppose that H_1 and H_2 are much more probable for Darwin than H_3-H_{100}. For definiteness, let's suppose that $p(H_1) = p(H_2) = 0.4$, whereas $p(H_3) = ... = p(H_{100}) = 0.2/98 \approx 0.00204$ (where $p(\cdot)$ is Darwin's probability function). Now, in a nearby possible world, Darwin's near-identical counterpart, Darwin*, is in almost the same situation. It's just that for Darwin*, the probabilities of the 100 hypotheses are very different (because Darwin* has access to other evidence). Let's suppose that Darwin*'s probabilities are as follows: $p^*(H_1) = .4$, $p^*(H_2) = \cdots = p^*(H_{100}) = 0.6/99 \approx 0.00606$ (where $p^*(\cdot)$ is Darwin*'s probability function).

The thing to notice about these two cases is that although the probability of H_1 is the same for Darwin and Darwin*, Darwin and Darwin* differ in that only for Darwin* is H_1 is significantly more probable than any other relevant hypothesis. Thus, a natural thought is that H_1 is acceptable for Darwin* but not for Darwin. After all, the thought goes, H_1 is 66 times as likely to be true as any other hypothesis that would explain biological evolution for Darwin, whereas for Darwin* there is an equally likely rival in Lamarck's theory of acquired characteristics H_2.[20] Now, admittedly, we should not take our intuitions about probabilities too seriously, for as we have seen the sense of "probability" we are employing here is not the layperson's sense that it's natural to think is influencing our intuitive judgments. But this suggests a promising line of thought, viz. that a hypothesis is acceptable just in case it is much more probable than any other hypothesis what would explain the same things. The next section spells out a precise account based on this idea, and section 6 shows that the account avoids the double trouble of the Preface and Lottery paradoxes in a satisfying manner.

3.5 The Optimality Account

This section develops a probability-based "optimizing" account of explanatory acceptability. The basic idea behind this account is that a hypothesis is acceptable just in case it is (significantly) more probable than any of its explanatory rivals. A bit more precisely, the idea is that a hypothesis H is acceptable for an agent S just in case the credence it is rational for S to assign to H is (significantly) higher than the credence it is rational for S to assign to H's explanatory rivals. Call this *the Optimality Account* of acceptability.[21] The main task of the current section is to spell out the idea behind the Optimality Account in three ways: First, we need to say more about what counts as an *explanatory rival* to a given

[20] One might object that, intuitively, no hypothesis can be acceptable if its probability is less than 0.5. But this, of course, lands us straight into the Preface Paradox again, so we must at least be willing to allow for the possibility of acceptable hypotheses with probability below 0.5.

[21] Part of the idea behind this account is not entirely new. It has often been suggested that scientific acceptance and/or theory-choice is comparative or contrastive in some way or other. (See, for example, Salmon (1990); Laudan (1997); Godfrey-Smith (2008).) My task is to convert this idea into a rigorous and defensible position.

hypothesis. This is the subject of subsection 5.1. Second, we need to say more about how the probability of a hypothesis is to be compared to the probabilities of its explanatory rivals. This is spelled out in subsection 5.2. Finally, we need to incorporate into the account the "global" nature of acceptability. This is done in subsection 5.3.

3.5.1 Explanatory Rivals

Recall that in a positive optimizing evaluation something is evaluated as better than some relevant alternatives. What are the relevant alternatives in the case of explanatory acceptability? A plausible thought is that the relevant alternatives are *other potential explanations*. But other potential explanations *of what*? Is a hypothesis that explains why dinosaurs became extinct a relevant alternative to a hypothesis that explains Brownian motion? Surely not. So, plausibly, the relevant alternatives are other potential explanations *of the same facts*. To a first approximation, that is what I shall refer to as an "explanatory rival".

But we need to be more precise, for we haven't yet said what it is for two theories to be potential explanations of the same facts. I propose that we spell this out by adopting van Fraassen's (1980) framework for thinking about scientific explanations. The idea here is to conceive of an explanation as an answer to a specific kind of question, what I shall call an "explanatory question" (roughly following Salmon (1992) who calls such questions "explanation-seeking why-questions").[22] This suggests that an explanatory rival to a hypothesis H needs to answer all and only the explanatory questions that are answered by H. Now, each hypothesis H will answer an indefinite number of explanatory questions, but only some of them will be relevant to our interests as explainers. And since it seems to me that there is no privileged set of explanatory questions that we should all be interested in, I propose that we relativize what counts as an explanatory rival to the explanatory questions that each of us is trying answer at a given time. Thus we can say (with some increased

[22]My choice of terminology is determined by a desire to allow that some explanations are answers to *how*-questions (as opposed to *why*-questions). (Cross 1991)

precision) that H_1 and H_2 are explanatory rivals relative to a set of explanatory questions Q just in case H_1 and H_2 provide different answers to the questions in Q.

This relativizes explanatory rivalry to a set of explanatory questions. If there is no single set of explanatory questions that we should all be trying to answer – as seems plausible – then there is no single objective set of explanatory rivals to a given hypothesis. Thus explanatory rivalry is to some extent a subjective matter. However, the subjectivity here might not be as widespread as one might think, for it's plausible that certain explanatory questions are imposed on us, e.g. in virtue of the roles we occupy. For example, a cosmologist working on theories about the expansion of the universe may not very well reject, say, the explanatory question that asks why the universe is expanding rather than shrinking or retaining its size. Indeed, if one thinks that it is part of the very nature of science to aim to explain certain aspects of the universe, then scientists may be required to take on certain explanatory interests in virtue of the very fact that they are scientists.[23]

So explanatory rivals are, roughly speaking, hypotheses that provide different answers to the same set of explanatory questions. Note, however, that this does not yet ensure that explanatory rivals are mutually incompatible. The intuitive rationale for requiring explanatory rivals to be mutually incompatible is clear enough: If H and its explanatory rivals are not mutually exclusive, then it's hard to see how they could be *rivals* in any meaningful sense. After all, two hypotheses that are not mutually exclusive could both be true at the same time, and so the two hypotheses need not be in any competition with each other. One way to ensure that H and its explanatory rivals are mutually incompatible is

[23] Kitcher makes a similar, if somewhat stronger, point:

> If someone asked why we want to understand [how an organism's characteristics unfold from a tiny piece of organic material] – or why we want to know why the heavenly bodies move as they do, or why we are interested in the evolution of our hominid ancestors – it would be hard to say very much. We expect other people to see the point of such questions, and we describe those who don't as lacking in "natural curiosity." Partly as the result of our having the capacities we do, partly because of the cultures in which we develop, some aspects of nature strike us as particularly salient or surprising. (Kitcher 2001b, 81)

to require that explanatory rivals give *complete* answers to explanatory questions, where a complete answer either affirms or denies any (regular) answer to the question. More precisely: Given a set of answers $\{A_1, \ldots, A_n\}$ to an explanatory question q, we define a *complete answer* as a conjunction of n propositions in which the i-th member is either A_i or $\neg A_i$. Thus, for $n = 2$ for example, the set of complete answers would be:

$$\{A_1 \& A_2, \neg A_1 \& A_2, A_1 \& \neg A_2, \neg A_1 \& \neg A_2\}$$

Using this notion, we can finally give a precise definition of explanatory rivalry as follows:

Explanatory Rivalry: A pair of hypotheses H_1 and H_2 are explanatory rivals relative to a set of explanatory questions Q iff, for some question $q \in Q$, H_1 and H_2 provide distinct complete answers to q.

In slogan form, this view has it that explanatory rivals provide distinct complete answers to explanatory questions.

3.5.2 Probabilistic Optimality

The basic idea behind the Optimality Account is that a hypothesis is explanatorily acceptable just in case it is (significantly) more probable than its explanatory rivals. Having spelled out what it means for something to be an "explanatory rival", I now want make the account more precise by also spelling out what it means for some hypothesis to be *(significantly) more probable* than its explanatory rivals.

Note that if we were to drop the parenthetical remark, "significantly", spelling this out would be straightforward: For a hypothesis H to be more probable than its explanatory rivals is for each such rival H' to be such that $p(H) > p(H')$. The parenthetical remark appears quite important, however. If H is only slightly more probable than one of its explanatory rivals H' then it is at least something of a stretch to say that H is acceptable. After all, there would be another hypothesis H' which is only slightly less probable, and which is such that, if true, H cannot be true. We might still want to say that H is *slightly* acceptable, or *not completely unacceptable*, or something like that. But there is at least a push towards requiring H not only to be more probable than its explanatory rivals, but significantly so. So let's see if there is a way to accommodate such a requirement.

One clue for how to do that is provided by the standard mathematical interpretation of the much-greater sign, "$\gg$". A claim of the form $a \gg b$ is standardly interpreted as claiming that a is larger than b by some order of magnitude (i.e. factor) r, where r is commonly (but not necessarily) a power of ten (e.g. 10, 100, 1000, etc.). So on this interpretation $a \gg b$ is equivalent to $a > r \times b$ given such a factor r. This suggests a ratio-based approach to "significantly more probable": that H is significantly more probable than H' just in case $p(H) > r \times p(H')$ for some factor $r > 1$. (In accordance with the mathematical convention, we will often write this simply as $p(H) \gg p(H')$ in what follows.) Here the factor r may be taken to vary with context or stakes, just like the threshold t might be taken to vary with context or stakes. Moreover, if one thinks "acceptability" is vague, then r may be replaced by an interval I_r so that H is neither acceptable nor not-acceptable when the most probable explanatory rival to H, H', is such that $p(H)/p(H') \in I_r$.

Before we move on, let's examine another salient possibility for interpreting "significantly more probable". Consider a difference-based approach, according to which H is significantly more probable than H' just in case $p(H) > d + p(H')$ for some $0 < d < 1$. Now, this does not accord with the standard mathematical interpretation of "$\gg$", but then again we shouldn't let mathematical practice dictate our theory of acceptability. The deeper problem with a difference-based approach is that it smuggles in a probability-threshold for acceptability. For it is not hard to see that, on this approach, the probability of any acceptable hypothesis H must exceed d (whatever it is), so d becomes an implicit probability-threshold.[24] We saw the perils of this requirement in examining the Preface Paradox, so

[24]Proof: Suppose H is acceptable just in case for some number d and any explanatory rival H' to H, $p(H) > d + p(H')$. Since any probability lies in the interval $[0, 1]$, $p(H') \geq 0$. It follows that $p(H) > d$, i.e. that the probability of H must exceed d.

difference-based approaches are non-starters.

3.5.3 From Local to Global Acceptability, and Back Again

One of the lessons of the Lottery Paradox seems to be that acceptability is determined "globally" rather than "locally": If we look at each lottery proposition individually, each one appears acceptable. After all, each lottery proposition is very, very probable, so as long as we don't consider the fact that they conflict with one another, there seems to be nothing wrong with appealing to all of them in explanations. But once we widen our perspective to notice how each proposition hangs together with other propositions, we see that they cannot all be acceptable. Thus it looks like an adequate account should take this "global" nature of acceptability into consideration.

My proposal for how to do that proceeds by focusing on what I shall call *global explanatory theories*. The idea is that acceptability of a hypothesis is determined by the extent to which the most probable global explanatory theory that entails it is more probable than the most probable global explanatory theory that entails its negation, where a "global explanatory theory" is a theory that answers *all* of our demands for explanations. This retains the idea that acceptability is determined by a probabilistic comparison between theories that would explain the same things. But it also accommodates the idea that acceptability is a global rather than local matter in that the theories being compared are the most probable *global* explanatory theories from which the hypotheses follow. The main task of the current section is to spell out this idea with more precision.

Recall that as we defined "explanatory rivals", two hypotheses are explanatory rivals relative to some set of explanatory questions Q, just in case they provide different complete answers to the questions in Q. Now consider a theory that provides complete answers to *all* explanatory questions of an agent S (including those prompted by considering the various answers to those questions). I shall refer to a theory that does that (and does nothing else) as a "global explanatory theory". A global explanatory theory will thus, by definition, give complete answers to all of S's explanation-seeking questions.[25] It's easy to see that any two global explanatory theories for an agent S are explanatory rivals to each other, relative to S's explanatory questions.

Now consider a given (complete or non-complete) answer to an explanatory question – call that answer H. Any global explanatory theory will either give that answer, and so entail H, or give an incompatible answer, and so entail H's negation, $\neg H$. To put it differently, some global explanatory theories will affirm H, and the rest will deny H – none will be neutral. A natural thought then, in line with what I suggested above, is that H is acceptable just in case the probability of the most probable global explanatory theory that entails it is (significantly) greater than the probability of the most probable global explanatory theory that entails its negation.

Let us make this precise. Let $G_S(H)$ be the set of probabilities, for an agent S, of the global explanatory theories, for S, that entail H. (I.e. let $G_S(H) = \{p_S(G_i) : G_i$ is a global explanatory theory for S that entails $H\}$.) Then the probability of the most probable global explanatory theory that entails H is given by $\max(G_S(H))$, and the probability of the most probable global explanatory theory that entails $\neg H$ is given by $\max(G_S(\neg H))$. So formally the acceptability of H is determined by a comparison of $\max(G_S(H))$ and $\max(G_S(\neg H))$. More precisely:

The Optimality Account: H is acceptable for an agent S, just in case

$$\max(G_S(H)) > r \times \max(G_S(\neg H)),$$

where $r \geq 1$.[26]

[25] Though many of these "answers" will simply reject one or more presuppositions of the questions. For example, for those of us who reject the phlogiston theory of combustion, our current global explanatory theory "answers" the question "Why is phlogiston released upon burning?" by responding that there is no such thing as phlogiston.

[26] Again r may be taken to vary with context and/or stakes, and may also be replaced by an interval I_r if one thinks "acceptability" is vague.

More succinctly, the Optimality Account holds that H is acceptable just in case $\max(G_S(H)) \gg \max(G_S(\neg H))$.[27]

We can sum up all of this more colloquially as follows: The Optimality Account holds that the acceptability of a hypothesis is determined by a probabilistic comparison between the most probable ways of answering all of one's explanatory questions that affirm and deny the hypothesis. In accordance with our conclusions in the previous subsection, this comparison is "ratio-based" in that the global explanatory theory that affirms the hypothesis must be r times as probable as the one that denies it. Now, equipped with this account, let us now return to the double trouble of the Lottery and Preface Paradoxes.

3.6 Escaping Paradox

Section 3 discussed four principles: (&), (D), (C), and **Deductive Cogency**. We saw that the Preface Paradox shows that the Threshold View violates (&), and thereby also (D) and **Deductive Cogency** (both of which entail (&)), while the Lottery Paradox shows that the Threshold View violates (C) and thereby also **Deductive Cogency** (which entails (C)). This section shows that the Optimality Account obeys these principles. Let us go through them in turn.

That the Optimality Account obeys (&) – the requirement that the conjunction of any acceptable hypotheses is acceptable – follows immediately from the following theorem (proved in appendix A.1):

[27] There is a non-obvious feature of the Optimality Account that is worth highlighting: It may be that no global explanatory theory is acceptable. Indeed, this is presumably quite common for a large enough r, since it might be that no global explanatory theory is r times as probable as all other such theories. Fortunately, it does *not* follow from this that no hypothesis would be acceptable. To see why, note that in determining whether a given hypothesis H is acceptable, we compare the probabilities of the most probable global explanatory theories that entail H, on the one hand, and those that entail its negation $\neg H$, on the other hand. If H is a global explanatory theory, then these theories will simply be the two most probable global explanatory theories. But if H is not a global explanatory theory, i.e. if H is merely entailed by a global explanatory theory (and not *vice versa*), then these two global explanatory theories may be separated by some other global explanatory theory that is less probable than one but more probable than the other. Of course, one of these theories will always be the most probable global explanatory theory, but since the other need not be the second-most probable it follows that H may be acceptable even though the most probable global explanatory theory from which it follows is not.

Theorem 1. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

then

$$\max(G_S(H_1 \& \ldots \& H_n)) > r \times \max(G_S(\neg(H_1 \& \ldots \& H_n))).$$ [28]

That the Optimality Account obeys (D) – the requirement that any explanatorily potent logical consequence of any acceptable hypotheses is acceptable – follows immediately from the following theorem (proved in appendix A.2):

Theorem 2. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

and if $\mathbf{H}$ entails some H_c (not necessarily in $\mathbf{H}$) that also answers at least one of S's explanatory questions, then

$$\max(G_S(H_c)) > r \times \max(G_S(\neg H_c)).$$

That the Optimality Account obeys (C) – the requirement that any set of acceptable hypotheses must be consistent – follows immediately from the following theorem (proved in appendix A.3):

Theorem 3. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least

[28] In this and the following theorems it is assumed that $r \geq 1$.

one of S's explanatory questions. If for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

then $\mathbf{H}$ is consistent.

So the Optimality Account obeys (&), (D), and (C). This means, of course, that the Optimality Account obeys the conjunction of these principles, **Deductive Cogency**, and thus that neither the Preface Paradox nor the Lottery Paradox applies to the Optimality Account. This is a significant result, not least because the Optimality Account was not "rigged" to avoid these paradoxes – that it does so is rather something that falls out of a fairly natural way to think about acceptability, viz. as an *optimizing* evaluation of hypotheses in terms of the probabilities of the *global* theories in which they are embedded. In this respect the Optimality Account is superior to solutions to the Preface and Lottery paradoxes that stipulate in an *ad hoc* manner some condition that is not satisfied in each of the paradoxes. Furthermore, the Optimality Account provides a *simple* and *unified* solution to the two paradoxes, unlike some other solutions that only tackles one of the paradoxes or that use entirely different machinery to solve each paradox.

3.7 Inference to the Likeliest Explanation

I now turn to an upshot of the Optimality Account, viz. that it provides for a natural probabilistic reinterpretation of "Inference to the Best Explanation" (IBE). Let us start by clarifying what IBE is supposed to be. Lycan (2012) provides a succinct and relatively uncontroversial description:

> An explanatory inference, or inference to the best explanation, proceeds from an explanandum or a set of data to a hypothesis that explains the data better than would any available competing hypothesis [...] The features that make one hypothesis a better explanation than another are the specifically explanatory

> virtues: simplicity, power (explaining more than does its competitor), testability, fecundity or fruitfulness, neatness (leaving fewer unanswered questions behind), and conservativeness (fitting with what we already reasonably believe) – in some complex combination of such factors. (Lycan 2012, 5-6)

So, in short, IBE is a rule according to which one should infer a hypothesis just in case the hypothesis has the best combination of explanatory virtues of any available hypothesis.

One of the biggest issues about IBE in the recent literature has been about whether (and, if so, how) IBE is compatible with a probabilistic or "Bayesian" approach in epistemology. Some who argue that IBE and Bayesianism are incompatible reject IBE (van Fraassen 1989), while others reject Bayesianism for this reason. (Psillos 2007a,b; Lycan 2012) A more popular approach has been to argue that IBE and Bayesianism are compatible, or even complimentary. (Okasha 2000; McGrew 2003; Lipton 2004; Weisberg 2009; Henderson 2013) Simplifying quite a bit, this approach is based on the idea that explanatory considerations may come into play in determining prior probabilities such that, at the end of the day, more explanatory hypotheses receive a higher posterior probability than less explanatory hypotheses.

However, as Psillos (2007a) points out, it's not clear that this suffices to reconcile IBE and Bayesianism:

> [...] on a strict Bayesian approach, we can never detach the probability of the conclusion of a probabilistic argument, no matter how high this probability might be. So, strictly speaking, we are never licensed to accept a hypothesis on the basis of the evidence. (Unless, of course, we institute well-known and problematic rules of acceptance.) All we are entitled to do is a) detach a conclusion about a probability, viz., assert that the posterior probability of a hypothesis is thus-and-so; and b) keep updating the posterior probability following Bayesian conditionalisation on fresh evidence. But IBE is a rule of acceptance: it authorises the acceptance of a hypothesis, on the basis that it is the best explanation of the evidence. (Psillos 2007a, 445-6)

Psillos is right, of course, that Bayesianism cannot countenance "acceptance" *unless* we adopt some rule of acceptance based on probabilities. However, in light of the above discussion, it would be premature to conclude with Psillos that Bayesianism is in conflict with IBE because "well-known" rules of acceptance are "problematic". (Presumably, Psillos is here referring to the fact that threshold views are susceptible to the Preface and Lottery paradoxes and violate **Deductive Cogency**.)

Not only is the Optimality Account not susceptible to the "well-known" problems with threshold views, but the Optimality Account moreover enables us to adopt a conciliatory approach to Bayesianism and IBE. On the Optimality Account a hypothesis H is acceptable just in case it follows from a global explanatory theory that is (significantly) more probable than any such theory that denies H. This retains the spirit of IBE in that it provides a place for *acceptance*, as Psillos rightly urges is an essential part of IBE. This approach should also be friendly to IBE in that both IBE and the Optimality Account are "optimizing" (as opposed to "satisficing") accounts of acceptability: On IBE, a hypothesis is acceptable just in case it is *better* than any competing hypothesis, not because it is *good enough*. So from the perspective of the IBE, the Optimality Account gets the *structure* of acceptability right.

Now, IBE as it is usually formulated (e.g. by Lycan (2012)) takes the *standard of evaluation* to be a combination of explanatory virtues, whereas the Optimality Account takes the standard of evaluation to be probability. But in the grand scheme of things, this can be seen as a relatively small disagreement between the two accounts. After all, on the conciliatory approach to IBE and Bayesianism advocated by Okasha, McGrew, Lipton, Weisberg and Henderson, explanatory considerations help determine posterior probabilities, so on these approaches the most explanatory theory will, at least normally, also be the most probable theory. Thus these two standards of evaluation may very well end up giving similar if not identical results regarding which theories do best on these standards.

So at the end of the day, the Optimality Account can be seen as a friendly suggestion for advocates of IBE – a way of locating both the optimizing structure and the explanatory-virtue-based content of IBE in a precise probabilistic framework. Where the Optimality Account differs from more traditional formulations of IBE is really only in urging that explanatory considerations are not "basic", but have epistemic value only if and to the extent that they make hypotheses more likely to be true.[29]

3.8 Conclusion

I have argued for a "probabilistic" account of explanatory acceptability – of what makes it reasonable to appeal to a given hypothesis in one's explanations. We used the Preface and Lottery paradoxes to show how the natural and *prima facie* plausible Threshold View of acceptability failed to obey **Deductive Cogency**. We then contrasted "satisficing" and "optimizing" views of acceptability, and constructed an account of the latter kind using probability as the standard of evaluation. Next we proved that, without introducing any *ad hoc* maneuvers, this account obeys **Deductive Cogency**, and thus is not susceptible to any version of the Preface and Lottery paradoxes. Finally, I argued that the Optimality Account allows for a further reconciliation between Bayesianism and Inference to the Best Explanation (IBE).

[29]Of course, not all proponents of IBE will welcome this suggestion, for some of them really do think that IBE is epistemically basic. (Lycan (2012) seems to argue for such a view, for example.) But my sense is that the majority of IBE's advocates would indeed see this as an improvement on more traditional formulations of IBE, not least because of the tremendous successes of the probabilistic approach to epistemology and philosophy of science.

4 REALISM AND THE ABSENCE OF RIVALS

Abstract

In one sense, *scientific realism* holds that we are epistemically justified in believing to be true at least some of the claims made by empirically successful scientific theories, including those that concern unobservable entities such as quarks, genes and electromagnetic fields. This paper is a defense of this view against one of the most influential challenges to it, namely the *underdetermination of theory by evidence*. An essential part of the defense consists in an analysis of the slogan "absence of evidence is not evidence of absence". This analysis serves as the foundation for blocking two routes to underdetermination, viz. the *New Induction* given by Kyle Stanford, and the *No-Privilege Argument* given by Bas van Fraassen. I conclude with some reflections on how social, psychological and historical studies of science may help rather than hinder the case for the trustworthiness of the scientific enterprise.

4.1 Introduction

"Scientific realism" is a label applied to a notoriously large number of views, many of which are in no conflict with one another. (In addition, there are of course many versions of realism apart from scientific realism, but we shall focus only on the latter. Thus, from this point on, "realism" denotes *scientific* realism.) The kinds of realism that will interest us in this paper are *epistemic* in that they they concern what we can justifiably believe concerning the posits of scientific theories. These epistemic realisms take for granted two different kinds of realist views, namely that there is a mind-independent world, some of which may very well be unobservable (metaphysical realism), and that the claims made by scientific theories – including the claims about unobservables – are true or false depending on the state of this mind-independent world (semantic realism).

I will also not be concerned here with versions of scientific realism that concern the aim of science or acceptance of scientific theories – the kinds of realism that are influentially discussed by van Fraassen (1980). Epistemic realisms, by contrast, concern what we can justifiably believe on the basis of scientific work – what, in other words, science teaches us about the world. More precisely, epistemic realism is the view that we are epistemically justified in believing that at least of the claims made by empirically successful scientific theories are true (or approximately true), including those that concern unobservable entities.[1] Strictly speaking, this does not entail anything about the aim of science, or about what constitutes acceptance of a scientific theory.[2]

Note that I have formulated epistemic realism in terms of epistemic justification. Many other philosophers formulate epistemic realism only in terms of *truth*, with no mention of justification (or epistemic rationality, warrant, reasonableness, etc.).[3] On such a formulation, epistemic realism asserts that empirically successful theories are true, including the theories' claims about unobservable entities. There are at least two problems with such a formulation of epistemic realism: First, it is not clear in what sense this would really be an *epistemic* position at all. After all, this position is extensionally equivalent to a conjunction of claims made by various scientific theories – and that, of course, is just yet another first-order scientific theory (albeit perhaps a very ambitious one). So unless all scientific theories are epistemic claims, *epistemic* realism cannot simply assert that certain scientific theories are true.

[1]From now on, I shall use "true" to stand for "true or approximately true". There is certainly an issue about whether the realist needs to, should, and can, appeal to the notion of approximate truth. (My view is that she need not and should not, even though she probably can.) But that is an issue for another day.

[2]That is not to say, of course, that the issues are totally unrelated. Indeed, I indicate the relationship between epistemic realism and realism about scientific acceptance in chapter 2.

[3]See, for example, (Cartwright 1983), (Hacking 1983), and (Chakravartty 2007). (To be fair, most authors are not explicit about which formulation they are using, and may very well be using the truth-formulation as a shorthand for the justification-formulation that I favor.)

Second, epistemic realism would on this view be refutable by future empirical findings, for today's realist would be committed to the scientific claims that are refuted tomorrow. This neglected point was already made by van Fraassen (1980) more than three decades ago:

> A naïve statement of [scientific realism] would be this: the picture which science gives us of the world is a true one, faithful in its details, and the entities postulated in science really exist [...] That statement is too naïve; it attributes to the scientific realist the belief that today's theories are correct. It would mean that the philosophical position of an earlier scientific realist such as C. S. Peirce had been refuted by empirical findings. (van Fraassen 1980, 6-7)

Note that my formulation of epistemic realism is not susceptible to this critique, for one may of course be epistemically justified in believing something that turns out to be false. Note also that on my formulation, epistemic realism is a *meta-theory* in that it is a theory about (the justification of) theories, as opposed to being itself a first-order scientific claim. Rather than asserting the existence of some entities or purport to explain the occurrence of some phenomenon, epistemic realism is a theory about the epistemic justification of certain scientific claims.

Which scientific claims? Well, first of all, realists tend to emphasize empirical success, urging that we should only claim to be justified in believing that *empirically successful* theories are true. In addition, most realists these days advocate "selective" as opposed to "wholesale" realisms, in that they urge that only some proper subset of claims made by empirically successful theories are epistemically justified. For example, the *entity realism* of Cartwright (1983) and Hacking (1983) focuses on the subset of claims about the causally efficacious entities posited by scientific theories. The *structural realism* of Worrall (1989, 1994) focuses on claims that concern the *structure* of the world in so far as it can be described by purely mathematical models. Finally, the *explanationist realism* of Kitcher (1993, 2001a) and Psillos (1999) focuses on the claims that are essential for explaining the theory's empirical success. This paper can be seen as proposing a different kind of selective realism, one that focuses on claims that enjoy a certain kind privileged status within scientific communities.

My main objective is to defend this version of epistemic realism against one particular challenge, namely a version of the so-called "underdetermination of theory by evidence". Let me say a little about why I consider underdetermination to be the most fundamental problem for epistemic realisms. Some would argue that the *pessimistic induction* is the more worrying problem for realists – indeed, that problem motivates both Worrall's structural realism and Kitcher's and Psillos's explanationist realisms.[4] My reason for focusing on underdetermination is that, as I explain in the next section, I consider underdetermination to encompass the most threatening form of the pessimistic induction – viz. Stanford's "New Induction" (more on which below). As we shall see, the pessimistic induction – on my way of setting things up – is but a prototype of one of many *routes* to underdetermination.[5] So if an epistemic realism avoids underdetermination, then it *ipso facto* avoids the most threatening form of the pessimistic induction as well. In that sense, it is underdetermination that is the more fundamental of the two most prominent challenges to epistemic realisms.

Here is a section-by-section breakdown of the rest of the paper. In section 2, I discuss the underdetermination of theory by evidence, arguing that the form of the problem that realists should take seriously is one that concerns undiscovered serious rival hypotheses that are at least as well supported by the available evidence. In section 3, I take a step back from the realism debate to discuss when, if ever, absence of evidence is evidence of absence. A

[4]The most prominent defender of the pessimistic induction is no doubt Laudan (1981), but the argument dates back at least to Poincaré (1952). Another well-known proponent is Putnam (1978).

[5]By calling the pessimistic induction a "prototype", I am indicating that I think that the argument is too flawed to even facilitate an interesting debate, while acknowledging that its successor – Stanford's New Induction – should be taken seriously. Indeed, this is how Stanford himself sees the issue. Stanford argues (convincingly, I think) that the pessimistic induction fails to take into account that the circumstances which prompted the acceptance of earlier theories differ markedly from the circumstances that science operates in today. There are other serious problems as well, such as the fact that the pessimistic induction commits what Lange (2002) calls "the turnover fallacy", and perhaps also the well-known "base-rate fallacy" (Lewis 2001).

probabilistic analysis reveals that two conditions must be satisfied – what I call *plausibility* and *sensitivity*. In section 4, I then apply this result to underdetermination, arguing that a theory is unlikely to be underdetermined (in the relevant sense) to the extent that *plausibility* and *sensitivity* are present to a high degree. In sections 5 and 6, I then discuss two routes to underdetermination, namely Stanford's *New Induction* and van Fraassen's *No-Privilege Argument*. I argue that neither of these arguments are successful against a suitably restricted version of epistemic realism.

4.2 Underdetermination through Underconsideration

Much of this paper will be concerned with underdetermination. However, as Kitcher (1993) points out, it is at best doubtful that there is a single problem of underdetermination, as opposed to a cluster of problems that are conceptually and/or historically related. One historically significant underdetermination thesis holds that any scientific theory that posits unobservables is underdetermined given any evidence. Call this the *Classic Underdetermination Thesis*:

CUT: For any scientific theory T that posits unobservable entities, there is a rival theory T* to T that is equally or better supported given any evidence E.[6]

In support of **CUT**, it has been argued that for any theory T, we can construct a rival T* that makes all the same empirical predictions as T, and thus no empirical evidence E could possibly distinguish between T and T*. Indeed, we can easily provide algorithms for constructing such rivals for any theory T: For example, let T* be the claim that T is correct about observable entities but incorrect about unobservable entities. Alternatively, let T* be the theory that an evil demon is systematically deceiving you into thinking that T is true. (Kukla 1993)

[6]For the purposes of this paper, a "rival" theory is simply a theory that is incompatible with the theory it rivals. Some would argue that a theory doesn't count as a *rival* theory unless it at least addresses the question answered by the theory it rivals, so that (for example) the negation of a theory doesn't normally count as a rival to the theory, despite being incompatible with it. (Leplin and Laudan 1993; Dellsén b) Though I agree, I shall not be making that assumption in this paper.

However, as Stanford (2001, 2006) points out, this argument proves too much, in that it would undermine more than just *scientific* realism. After all, we can construct algorithms of a similar kind for *any* theory – be it about unobservable entities posited by scientific theories or the ordinary everyday objects that we see and touch. For example, the theory that I have hands is supported by the the evidence of me appearing to see my hands, but so is the theory that I am a handless brain in a vat that is being deceived into thinking that I have hands. So the argument for **CUT** would seem to force us towards a more radical kind of skepticism not just about unobservable entities, but also about observable entities and ultimately about the external world itself. Perhaps the argument does indeed show that we should go in for global skepticism of that kind. However, this brings out that the argument for **CUT** wouldn't pose any *special* problem to the epistemic status of successful scientific theories and the claims they make about unobservable entities.

What would pose a special problem for scientific realism, however, is if it could be argued that successful theories that posit unobservable entities regularly have equally well or better supported *serious* rival theories – non-skeptical alternative theories that would be taken seriously by the working scientist. Accordingly, anti-realists have recently argued that our justification for theories about unobservables is threatened not just by the existence of skeptical alternatives produced by some general algorithm, but rather by serious alternative theories that propose a genuinely distinct accounts of the unobservable aspects of reality. Call this the *New Underdetermination Thesis*:

NUT: Most successful theories that posit unobservable entities are such that there exist serious rivals to those theories that are equally or better supported given the currently available evidence.[7]

[7]This is roughly what (Stanford 2001, 2006) calls "Recurrent Transient Underdetermination", although Stanford focuses on scientific claims about "fundamental constituents of the natural world" (Stanford 2006, 32) rather than those about unobservable objects. In most of what follows, this difference will not be important (though see footnote 19).

A slightly different way to state the thought behind **NUT** is as follows: If T is a successful theory that posits unobservable entities, then there is probably a serious rival $T*$ to T such that our current evidence supports $T*$ equally well or better than T. (When convenient, I shall revert to this formulation, and take it to be understood that the probability in question is to be interpreted as an actual frequency.) This is the underdetermination thesis that I shall be concerned with in the rest of this paper. Thus, from now on, I shall say that a theory T is *underdetermined* just in case our current evidence supports some serious rival $T*$ equally well or better than it supports T; and I shall refer to such a rival $T*$ as an *underdetermination rival* to T.

So how might one argue in favor of **NUT**? Well, one strategy would be to go through all successful theories that posit unobservable entities and show by example that most of them have underdetermination rivals. However, most of the time we cannot do that, because (as realists like to point out) it is often hard enough to find a single theory that fits the evidence at hand, let alone more than that.[8] So, on this strategy, the anti-realist would only undermine the justification for believing a very select group of theories, namely those for which we can find such rivals. Moreover, it's clear that not even the most devout realists would recommend that we believe theories that we know to have underdetermination rivals.[9] So, if underdetermination is to be a problem for even remotely sophisticated epistemic realisms, then there must be some way to argue that even theories for which we haven't (yet) produced underdetermination rivals are likely to be underdetermined.

That is indeed what anti-realists have tried to do in recent years. This general strategy has been labelled "the argument from underconsideration". (Lipton 1993) This label is somewhat misleading, in my view, for there are at least two sorts of arguments that aim to establish that even theories for which we haven't (yet) found underdetermination rivals are likely to have such rivals. On the one hand, investigation into the history of science suggests

[8] See, for example, (Kitcher 1993) and (Psillos 1999).

[9] See, for example, (Psillos 1999).

that at least in many cases, scientists have simply failed to conceive of underdetermination rivals to successful scientific theories, even when such rivals existed. According to proponents of this argument, this gives us (inductive, empirical) evidence for the conclusion that currently accepted theories, especially the ones that concern unobservable entities, are also likely to have rivals of this sort. In short, we would have reasons to think that presently accepted scientific theories are underdetermined by the currently available evidence. The most prominent proponent of this argument is Stanford (2001, 2006), who calls it "the New Induction over the History of Science".[10]

It has also been argued that something about our abilities as scientific theorizers makes it unlikely that scientists would come up with underdetermination rivals to a given theory if there are any. As the proponents of this argument are fond of saying, there is no reason to think that scientists are "epistemically privileged" in the sense that they are somehow prone to develop theories that are worthy of our epistemic consideration. If scientists are indeed not privileged in this way, they argue, we have good reasons to think that most theories, especially those that posit unobservable entities, have underdetermination rivals that we simply haven't considered yet. Call this *the No-Privilege Argument* for **NUT**.[11]

In the following sections, I develop a common line of response to these arguments, undermining the support for **NUT**. However, let me be upfront about the fact that I think it's clear that there is some truth to **NUT**, for sometimes we really do have good reasons to suppose that there are underdetermination rivals to our theories – even when no such rivals have been identified. Accordingly, it seems to me that the epistemic realist should look for a way of distinguishing between those theories (and indeed those parts of theories) that are more and less likely to be underdetermined. Using such a distinction, we can argue that the two arguments in favor of **NUT** – the New Induction and the No-Privilege Argument

[10]Earlier proponents of versions of this argument include Hesse (1976) and Sklar (1981), and perhaps also Laudan (1981).

[11]This is the sort of argument presented by van Fraassen (1989), Roush (2005) and Wray (2011). This idea seems to have been present in (Sklar 1981) as well, who argues that underdetermination is due to "limitations of our scientific imagination". (Sklar 1981, 18)

– do not go through for the theories that fall on the less-likely-to-be-underdetermined side of the distinction.[12] So my response to **NUT** is not to argue that it is wholly without merit, but rather to find a suitably restricted set of scientific theories for which the arguments for **NUT** simply to do apply. Of course, this response will not save *all* accepted theories from the fate of (likely) underdetermination – but then again few if any realists these days see themselves as defending epistemic realism with regards to *all* accepted theories.

Before we can do any of that, however, we need to consider an issue that may appear to be quite unrelated to underdetermination, namely the truth (or otherwise) behind the slogan "absence of evidence is not evidence of absence". I will provide a probabilistic analysis of the conditions under which the slogan (under a charitable interpretation) comes out as true. As we shall see, this will provide a basis for a response to the two arguments for **NUT** canvassed above.

4.3 Evidence and Absence

A popular slogan among scientists and statisticians has it that "absence of evidence is not evidence of absence". The slogan admits of many interpretations. On one interpretation, the slogan holds that the fact that one hasn't come across something cannot even constitute part of one's evidence for the claim that no such thing exists. Thus interpreted, the slogan appears false: Consider your belief that there are no (genuine) pink elephants. How is this belief justified? It looks like an important piece of information justifying this belief may very well be that you have never seen a (genuine) pink elephant. After all, if you had seen a pink elephant, then you'd presumably not be justified in believing that there aren't any.

As Sober (2009) points out, this intuitive judgment is borne out in a probabilistic (or Bayesian) account of evidential support. On this account, E confirms H just in case $p(H|E) > p(H)$. The right-hand side of this biconditional is equivalent to $p(E|H) >$

[12] The strategy here is similar to what Psillos (1999) refers to as the *divide et impera* move against the pessimistic induction.

$p(E|\neg H)$, so on the Bayesian account anything one learns that is more likely to be the case given the truth of some hypothesis than given its negation confirms that hypothesis. Since one can surely be more likely not to have seen a pink elephant if there are no such things than if there are, it follows that on the Bayesian account the fact that one hasn't seen a pink elephant can confirm that there are none. Thus is looks like "absence of evidence" can be "evidence of absence" on the Bayesian account of confirmation.[13]

On a more charitable interpretation of "absence of evidence is not evidence of absence", the slogan holds that not having evidence that something exists is not sufficient for being rationally confident that it doesn't exist. Here the statement "E is evidence of H" is interpreted not as claiming that E makes H more likely than it would otherwise be, but as claiming that E makes H likely to be true – sufficiently likely, perhaps, for one to believe, accept, or be reasonably confident that H is true.[14] To use the pink elephant case again to illustrate this, note that the mere fact that one hasn't come across pink elephants is presumably not by itself sufficient to justify the claim that there are no pink elephants. More generally, there are obviously many kinds of things that you shouldn't believe don't exist even though you have never come across such things. So it looks like there is something to the idea that absence of evidence is not evidence of absence, if interpreted as the claim that the fact that one hasn't come across something is not sufficient for justifying the belief that it doesn't exist.

Now, if the second version of the slogan is correct, but not the first, then absence of evidence may be part of of one's justification for believing that things of that kind (probably) don't exist, but it cannot be the only part. This raises the following question: Provided

[13]Sober (2009) also points out that there is something to the slogan that "absence of evidence is not evidence of absence" in that absence of evidence is usually not very strong evidence for absence, given a "likelihoodist" measure of strength of confirmation. As Strevens (2009) points out, however, this is hardly a completely satisfactory interpretation of the original slogan, for it contradicts the slogan's apparent meaning. I offer a different way of interpreting the slogan below – one that does *not* contradict the apparent meaning of the slogan.

[14]Salmon (1975) famously distinguished between these two interpretations, referring to the first as the "relevance concept of confirmation" and the second as the "absolute concept of confirmation".

that one knows that one has not come across something, when should one be reasonably confident that things of this kind don't exist? What other factors need to be present in order for "absence of evidence" to form part of one's justification for a reasonably high degree of confidence that the thing in question really is absent?

A natural strategy for answering this question is to look to probability theory. We start by making the modest probabilist assumption that one's degrees of confidence, when rational, are probabilistically coherent (i.e. that they can be represented by a function $p(\cdot)$ that satisfies the standard (Kolmogorov) axioms of probability). Now let E_X be the proposition that there are known instances of some particular kind X, and let A_X be the hypothesis that no things of kind X exist. Given this, we can represent the probability that there are no Xs given the fact that there are no known Xs (and also given some background knowledge B) as the conditional probability $p(A_X|\neg E_X \& B)$. The significance of this representation is revealed by Bayes's Theorem:

$$p(A_X|\neg E_X \& B) = \frac{p(\neg E_X|A_X \& B)p(A_X|B)}{p(\neg E_X|A_X \& B)p(A_X|B) + p(\neg E_X|\neg A_X \& B)p(\neg A_X|B)} \tag{4.1}$$

Before we go further, let us simplify (4.1) in two ways. First, since the probabilities of a proposition and its negation always sum to one, we can replace $p(\neg A_X|B)$ with $(1 - p(A_X|B))$, and $p(\neg E_X|\neg A_X \& B)$ with $(1 - p(E_X|\neg A_X \& B))$. Second, consider $p(\neg E_X|A_X \& B)$: This is the probability that one would not know about something given that it does not exist. But one cannot possibly know about something that doesn't exist, so $p(\neg E_X|A_X \& B) = 1$. Thus (4.1) can be reformulated as follows:

$$p(A_X|\neg E_X \& B) = \frac{p(A_X|B)}{p(A_X|B) + (1 - p(E_X|\neg A_X \& B))(1 - p(A_X|B))} \tag{4.2}$$

From (4.2) we see that the probability of A_X if $\neg E_X$ is part of your evidence (and B is your background knowledge) depends on two (and only two) factors:

(a) $p(A_X|B)$: the probability of A_X given only B,

(b) $p(E_X|\neg A_X \& B)$: the probability of E_X given B and that A_X is false.

More precisely, the probability of A_X (given $\neg E_X$ and B) is higher to the extent that (a) and (b) are higher. This will help us see when not having found something makes it likely that something doesn't exist, so let's discuss these two factors in turn.

Regarding (a): It should come as no surprise that the probability of there not being anything of a certain kind is positively dependent on how probable that hypothesis is independently of the fact that you haven't found any such thing. Returning to our previous example, the probability of there not being any pink elephants given the fact that you've never come across any such things is clearly proportional to the plausibility of there not being any pink elephants in the first place. Had you thought, for example, that there are a lot of pink mammals, or that elephants come in all sorts of colors, then you should be less confident that there are no pink elephants, quite independently of the fact that you haven't come across pink elephants. Call this factor PLAUSIBILITY.

Regarding (b): This measures the extent to which it is probable that we would have known about something of a particular kind given that there are such things. Again, it should come as no surprise that this is relevant to the probability of there being things of a given kind, for when we argue that something doesn't exist on the basis of us not having found any such thing, then we seem to be assuming that it's at least somewhat likely that we would have found such a thing if it existed. In our pink elephant case, it surely matters to the probability of pink elephants existing whether it's reasonable to assume that one would have come across pink elephants if they existed. Call this factor SENSITIVITY.

I'll say much more about SENSITIVITY in the sections that follow. In particular, one might want to know when precisely SENSITIVITY is present to a high degree, i.e. when it is probable that we would have known about something if it existed. This is clearly where we will encounter a lot of the real difficulties in determining how likely something is to exist given that it hasn't been found, and we will examine some of the factors that are relevant in due course. However, just to give you a taste of what kinds of factors are relevant here, note that to the extent that there has been a substantial (but unsuccessful) search for something one is more justified in believing that one would have known if that thing existed. If, for example, you had the tedious job of being the Official Scout of the Pink Elephant Society, you'd be even more justified than you are now in believing that there are no pink elephants.

One might think that because $p(A_X|B)$ occurs three times in the right-hand side of (4.2), whereas $p(E_X|\neg A_X\&B)$ occurs only once, PLAUSIBILITY is more important for our purposes than SENSITIVITY. That would be a mistake. One way to see this is by noting that (4.2) is equivalent to the following equation (in which each probability occurs only once):

$$p(A_X|\neg E_X\&B) = \frac{1}{1+(1-p(E_X|\neg A_X\&B))(\frac{1}{p(A_X|B)}-1)} \tag{4.3}$$

This version of the equation also illustrates that one may be justified in being quite confident that something doesn't exist even when PLAUSIBILITY is very low, provided that SENSITIVITY is high enough to compensate: If $p(A_X|B)$ is very low, then $(\frac{1}{p(A_X|B)}-1)$ will be very high, but if $p(E_X|\neg A_X\&B)$ is high enough, then $(1-p(E_X|\neg A_X\&B))(\frac{1}{P(A_X|B)}-1)$ can still be quite low, and so $p(A_X|\neg E_X\&B)$ would still be fairly close to 1. The converse is true as well, of course: One can be justified in being quite confident that something doesn't exist even when SENSITIVITY is very low, provided that PLAUSIBILITY is high enough to compensate.[15] This point is important. On the Bayesian analysis, PLAUSIBILITY and SENSITIVITY are both relevant to how confident one should be that something doesn't exist, but a strong presence of one factor can compensate for a lack of the other. This is important because it reminds us of just how misleading it can be to focus on only one of the two factors, and thus mistakenly assume that any argument is weak if that factor is not strongly present. We need to look at the two factors in concert.

[15] I leave it to the reader to verify this for herself.

Now, how is all of this relevant to epistemic realism and **NUT**? Well, consider a scientific theory T for which we haven't come across any *underdetermination rival*. Since a theory is underdetermined (in the sense we're interested in) just in case it has such an underdetermination rival, the conditions under which a theory is likely not to have any underdetermination rival are precisely the conditions under which we should conclude with a high degree of confidence that the theory is not underdetermined. Hence we can distinguish between the theories are are more and less likely to be underdetermined based on the extent to which PLAUSIBILITY and SENSITIVITY are present with regard to underdetermination rivals for such theories. This will constitute the basic idea behind my response to the two arguments for **NUT** mentioned in section 2.

4.4 Determining Underdetermination

In the previous section, I argued that as a general matter, the fact that we haven't found something of a given kind makes it likely that there are no such things just in case two factors, PLAUSIBILITY and SENSITIVITY, are present to a combined high degree. We now apply this to the question of how likely it is that a given scientific theory has underdetermination rivals, i.e. (undiscovered) rival theories that are equally or better supported given one's current evidence. Our aim will be to distinguish between theories that are more and less likely to be have such underdetermination rivals, for this (as I'll argue in the next section) will equip us with the tools to undermine the anti-realist arguments for **NUT**.

Let's start by examining PLAUSIBILITY. With regard to underdetermination rivals, PLAUSIBILITY is a measure of how probable it is that a theory T has no underdetermination rivals irrespectively of whether any such rivals have been found. Now, realists will typically urge that if a theory is very successful, we have good reasons to think there are no underdetermination rivals to that theory. In support of this, realists can point out that to be empirically successful just is to be well supported given one's current evidence, and to the extent that a theory is well supported it is unlikely that a rival theory is *better* supported than it is. In this way, realists will insist that the very fact that empirically successful theories are successful makes them unlikely to have underdetermination rivals. In effect, this is to say that an empirically successful theory will have a high degree of PLAUSIBILITY regarding the existence of underdetermination rivals.

Realists may be right about this point, but it will hardly convince anyone who has some sympathies with the anti-realist. After all, the claim that empirically successful theories are unlikely to be underdetermined is precisely the point that anti-realists are in the business of denying. They will argue that empirical success is a poor guide to PLAUSIBILITY, either by appealing to the history of science (the New Induction) or by appealing to lack of epistemic privilege (the No-Privilege Argument). We will examine these arguments in more detail below, but my point for now is that to assume that empirically successful theories have a high degree of PLAUSIBILITY is at best to beg the question against the anti-realist.

Fortunately for the realist, our analysis of when "absence of evidence" is "evidence of absence" shows that a theory may be unlikely to be underdetermined even if PLAUSIBILITY is very low, provided the SENSITIVITY is high enough to compensate. To see this, let A_U be the hypothesis that there are no underdetermination rivals to a given theory T, and let $\neg E_U$ be the claim that no underdetermination rivals to T have been discovered. Now suppose that PLAUSIBILITY is very low, but that SENSITIVITY is quite high, e.g. that

$$p(A_U|B) = 0.2$$

and

$$p(E_U|\neg A_U \& B) = 0.9.$$

Now, we know from section 3 that:

$$p(A_U|\neg E_U \& B) = \frac{p(A_U|B)}{p(A_U|B) + (1 - p(E_U|\neg A_U \& B))(1 - p(A_U|B))}$$

Plugging our hypothetical values into this equation, we get:

$$p(A_U|\neg E_U \& B) \approx 0.714.$$

This illustrates that while high PLAUSIBILITY certainly helps in making it unlikely that a theory is underdetermined, it is far from necessary. What this means is that we can grant the anti-realist the assumption that PLAUSIBILITY is quite low even for our most successful theories. Even on that assumption, we can argue that at least some successful theories are very unlikely to be underdetermined, viz. those with high SENSITIVITY. Thus we have a rationale for distinguishing between theories that are more and less likely to be underdetermined that does not beg the question against the epistemic anti-realist.

So the crucial factor in determining (in a non-question-beggin way) whether a given theory is likely to be underdetermined is whether there is a high degree of SENSITIVITY regarding underdetermination rivals. Recall that SENSITIVITY is here the extent to which it is reasonable that we would know about rivals to a given theory T if they existed. But what kind of factors determine whether that is the case for a given theory T? We shouldn't expect to find any kind of algorithm here. At best, we can try to identify some such factors given assumptions about the nature and structure of the sciences – assumptions that may very well be called into question. That said, it seems to me that it ought to be uncontroversial that the following four factors are relevant to identifying theories with a high degree of SENSITIVITY:

(i) *Incentives*. First of all, whether there has been something like a substantial search for rivals to a theory T will be relevant to whether one ought to believe that the fact that no such rival have been found tells one very much about whether any such rival exists. But since such searches are seldom systematic, and sometimes not even deliberate,[16]

[16] An interesting exception is the search for alternatives to Einstein's general theory or relativity in the twentieth century, where theoretical physicists developed a systematic map of the logical space of theories of gravity. See (Earman 1992, chapter 7).

it is hard for us to determine for a given theory whether scientists have searched for rivals. What we can do is determine how *likely* scientists are to have searched for rivals to a given theory (perhaps non-systematically, perhaps non-deliberately). We can determine this by estimating the what sort of incentives scientists have to come up with rivals to currently accepted theories.[17] This will depend not only on what scientists themselves consider to be important domains of inquiry, but also on what is the focus of external sources of fame and funding.[18]

(ii) *Time*. A second factor is simply the period of time a theory has been accepted or contemplated. A theory that has been considered for a long time, especially if was widely accepted during that time, ought to have a richer history of searches for rivals and use in various scientific and ordinary contexts. Thus, the longer a theory has been accepted or considered by a scientific community, the more confident should we be that the fact that we haven't found any underdetermination rivals to it is due to the fact that there are none to be found.

(iii) *Community*. Third, the nature and size of the scientific community working with the theory in question will be relevant for similar reasons: The more scientists that have been searching (non-deliberately, non-systematically) for underdetermination rivals to T, the more probable it is that there are no such rivals to T if they haven't yet been found. Relatedly, the better these scientists have been trained at finding plausible theories, the better justified we are in believing that there are no underdetermination rivals to T if they haven't yet been found. Of course, scientists may never trained specifically to discover plausible rivals to currently accepted theories, but the training

[17]Here the literature on reward-mechanisms in science is relevant. See (Kitcher 1990, 1993) and (Strevens 2003).

[18]Followers of Kuhn (1996, 1977) will no doubt urge that in so far as theories that are widely accepted and fundamental to a scientific discipline, scientists will have to rely heavily on them in their daily investigations, and so have less of an incentive to discover rivals to them. If so, then there are would be little incentives to discover rivals to fundamental and widely accepted theories. Though space does not permit me to argue the point here, I do not think Kuhn's thesis is correct, roughly because I do not think that relying on theories in one's investigations in any way precludes one from seeking alternatives to them.

received by some scientists better equips them for such discoveries than the training received by other scientists. Scientific theorizing (like philosophical theorizing) is a *skill* that can be learned and mastered like any other. It is also something that can be *selected for* in various selection processes (e.g. in applications for graduate school, professorships, and funding).

(iv) *Domain*. Fourth, and finally, the nature of the theoretic domain itself will be relevant to how difficult it is to discover underdetermination rivals if there are any. Some areas are such that, due to various psychological barriers, it is more difficult to construct theories in those areas. For example, if a domain of inquiry calls for very complicated theories, then it may be harder to conceptualize what alternatives to theories in that domain would look like.[19] Additionally, some domain may call for theories that are simply quite alien to our everyday ways of conceptualizing the world, in which case it may be harder to conceive of what plausible theories in that domain would be like. If it is true, for example, that our way of conceptualizing the world is basically Newtonian, then domains that call for radically non-Newtonian thinking (e.g. cosmology and subatomic particle physics) may be such that underdetermination rivals are more likely to remain undiscovered.[20]

We have identified four factors which contribute to SENSITIVITY, and thus to the probability that a given empirically successful theory for which no underdetermination rivals have been found will not have any such rivals: (i) *incentives*, (ii) *time*, (iii) *community* and (iv) *domain*. I call these "identifiable indicators" of SENSITIVITY in order to emphasize that, in contrast to SENSITIVITY itself, they are factors which can be fairly easily identified. Of course, this ease of identification comes at a price, in that these factors are less

[19]This point is neatly emphasized by Roush (2005, 211-12).

[20]Thus, assuming that unobservables are, all other things being equal, harder to conceptualize, there is some truth to the anti-realist claim that theories that concern unobservables are, all other things being equal, more likely to be underdetermined. However, as I'll emphasize below, *other things are not always, or even usually, equal*, and so theories about unobservables can very well be less likely to be underdetermined than theories about observables.

precise and more open to debate than the more precisely defined and rigorously motivated SENSITIVITY. Nevertheless, I hope that (i)-(iv) provide a rough, though certainly fallible, guide to SENSITIVITY.

At any rate, it's important to note that, as with SENSITIVITY itself, all of the identifiable indicators come in degrees. This has two important consequences: First, since each identifiable indicator is a matter of degree, no single indicator must be present to a particularly high degree in order for SENSITIVITY to be high. For example, the domain of a theory may be such that we have a particularly hard time conceptualizing alternatives, but such a theory may still have a high degree of SENSITIVITY provided that *incentives*, *time* and *community* compensate. The second consequence of the fact that all of the four factors come in degrees is that the risk of underdetermination for a given theory will also be a matter of degree. Only in very exceptional cases can we be absolutely sure either that our theory is underdetermined, or that it's not. Even when our reasons for thinking that our theories are not underdetermined are very strong, they will not eliminate the risk of underdetermination.

So the point of the identifiable indicators is not to find a way to *guarantee* that one's theories are not underdetermined, but to distinguish between theories that are more and less likely to be underdetermined. Indeed, in what follows I will not be assuming that identifiable indicators can tell us the particular degree of SENSITIVITY had by a given theory. Rather, I shall assume only that we can use the identifiable indicators as our guide in identifying some theories as having a higher degree of SENSITIVITY than other theories. In other words, I'll use the identifiable indicators only as guides to *comparative* rather than *absolute* degrees ofSENSITIVITY. As it turns out, these comparative evaluations of SENSITIVITY will suffice in replying to the antirealist arguments arguments for **NUT** discussed in section 2. It is to these arguments that we now turn.

4.5 The New Induction

The New Induction begins with the undeniable historical observation that for a great many theories that were accepted in the past, often because they were very successful, later investigation would reveal alternative theories which were equally or better supported given the available evidence. In short, these theories turned out to have underdetermination rivals, in the sense spelled out above. From this it is inferred by a simple enumerative induction that the typical situation in science is that even empirically successful theories have "unconceived" (i.e. unknown or undiscovered) underdetermination rivals. In other words, the argument concludes that empirically successful theories are typically underdetermined, which of course is what the relevant underdetermination thesis, **NUT**, claims. (Although versions of this argument have been advanced by Hesse (1976), Sklar (1981), Laudan (1981), and Roush (2005), I shall focus on the version given by Stanford (2006), since his is arguably the most developed, sophisticated and widely-discussed version.)

What makes the New Induction so challenging for realists is that it promises to undermine even empirically successful theories – on any plausible definition of "empirical success". The reason for this is that the inductive base of the New Induction consists of theories that, undeniably, enjoyed great empirical successes. For example, Fresnel's ether theory was hugely successful, even by realists' lights, yet it was superseded by what is arguably a better supported theory a few decades later. (I will return to this example shortly.) The New Induction thus concludes that empirical success is a poor indicator of whether a theory is underdetermined, and thus that we are not justified in believing even very successful scientific theories to be true, at least "in those parts of the sciences that seek to theorize about the fundamental constitutions of the various domains of the natural world and the dynamical principles at work in those domains". (Stanford 2006, 32)[21]

[21] Stanford acknowledges that his argument will not necessarily support an anti-realist attitude towards all unobservable entities, and also might support an anti-realist attitude towards some claims about observables entities. Strictly speaking, then, the kind of anti-realist supported by the New Induction cuts across "the distinction between observables and unobservables that opponents of realism have traditionally regarded as so central." (Stanford 2009, 2) This nuance will not be important in what follows, since my reply to Stanford's

As I indicated towards the end of section 2, my strategy for responding to this argument will be to identify a proper subset of empirically successful theories to which the argument does not apply. It should be clear by now what that proper subset is: It is the set of successful theories for which SENSITIVITY is significantly higher than the now discarded (but at the time successful) theories which turned out to have unconceived underdetermination rivals. Since a successful theory with higher SENSITIVITY is, as we have seen, less likely to be underdetermined than an otherwise similar theory with lower *sensitivity*, an inductive argument from a set of theories with lower *sensitivity* to those with higher *sensitivity* will clearly be a very weak argument. (Compare: Since I know that it is less likely to rain when the air pressure is higher, an argument from the fact that it rained the past three days to the conclusion that it will rain today is weak if my barometer tells me that the air pressure rose significantly this morning.)

Of course, this point will only be interesting if there are at least some theories for which SENSITIVITY is significantly higher than for the theories that form the inductive base of Stanford's New Induction. To see why there are indeed such theories, it will be useful to consider the "identifiable indicators" of SENSITIVITY for underdetermination rivals that I proposed in the previous section. In order for Stanford's argument to support the claim that a given presently accepted successful theory T is likely to be underdetermined, one would need to specify an inductive base containing theories for which *incentives*, *time*, *community*, and *domain* indicate a greater or similar level of SENSITIVITY. It seems to me that the realist can be rather optimistic that for many of our currently accepted theories there will simply not be enough now-discarded theories to figure in the skeptic's inductive base.

To see this, consider again Fresnel's ether theory of light, which was later replaced

argument will have nothing in particular to do with whether one is an anti-realist about unobservables or the "fundamental constituents" that Stanford worries about.

by Maxwell's electromagnetic theory.[22] The relevant question is: How reasonable was it to believe that we would have known about underdetermination rivals to Fresnel's theory if they existed (as we now know that they did)? (Note that the question is not whether it's reasonable to believe now that Fresnel's theory had such rivals. The question is whether it was reasonable *at the time*.) To focus the discussion, let us consider Fresnel's claim that light travels as a wave in a material medium. As is well documented, Arago showed in 1818 (to Poisson's great surprise) that Fresnel's theory's prediction that a bright spot would appear at the center of the shadow of an opaque disc was indeed correct. In 1855, however, Maxwell presented a rival theory to that of Fresnel, in which it was posited that the material in which light waves travel was a complicated system of molecular vortices (later to be replaced by the electromagnetic field, yielding Maxwell's famous electromagnetic theory of light).

Now, consider Fresnel's claim just before Maxwell's presentation of a rival theory of the medium of light waves in 1855. My claim is that Fresnel's situation was dissimilar to the situation we are in today with regard to many currently accepted theories. Consider, for example, the currently accepted theory of light – the photon theory (as I will call it)– on which light consists of collections of photons with a "dual nature" of being both electromagnetic waves and discrete particles. How does the photon theory compare with Fresnel's with regard to each of the identifiable indicators of SENSITIVITY? (i) Starting with *incentives*, since Fresnel's ether theory was always something of a fringe theory as compared to the photon theory during most of the twentieth century, the incentives for finding an underdetermination rival to the dual nature theory are surely at least as great as those for Fresnel's ether theory. (ii) Second, the photon theory has undeniably been considered for a much longer *time* than Fresnel's theory had in 1855: The photon theory was first formulated by Einstein in 1905, and so has been considered for well over a century, whereas Fresnel's theory was first formulated in 1818 and so had only been considered for less than

[22]This one of the most widely used case of an empirically successful theory for which an underdetermination rival existed. For discussions, see (Worrall 1989, 1994), (Psillos 1999), and (Chakravartty 2007).

decades. (iii) Third, and equally undeniably, the *community* of scientists that accept and work with the photon theory is, and has been, much larger than the community working on theories of light in Fresnel's days. Moreover, these scientists are arguably also better trained than they were in Fresnel's days, having a greater stock of background knowledge and better access to relevant research. (iv) Finally, the *domains* of inquiry of each theory are the same, of course, since they are theories about the very same things. So, in sum, *domain* favors neither the photon theory nor the ether theory, *incentives* favors the photon theory at least somewhat, and *community* and *time* both strongly favor the photon theory.

This is not an isolated case of a currently accepted theory being dissimilar with respect to SENSITIVITY to the theories that form the inductive base of the New Induction. In fact, the same point can be made about Stanford's central case study, viz. theories of biological heredity in the 19th century. (Stanford 2006, chapters 3-5) At least three serious theories were proposed, each one with at least some degree of empirical success: Darwin's "pangenesis" theory, which was replaced by Galton's "stirp" theory, which in turn was replaced by Weismann's "germ-plasm" theory. Now, Stanford points out that Darwin, Galton and Weismann were all apparently unable to conceive of the theory that would later replace their own. But does this support the conclusion that our currently accepted theory of heredity – roughly, a version of Mendelian genetics – will also one day be replaced by a rival that is at least as well supported by the available evidence?

Only if Darwin's, Galton's and Weismann's theories have a degree of SENSITIVITY that is comparable to that of Mendelian genetics.[23] So consider how the theories compare with regard to the four identifiable indicators of SENSITIVITY: (i) First, the *incentives* for discovering a well supported rival to Mendelian genetics are presumably at least as great as for discovering rivals to Darwin's, Galton's, and Weismann's theories. (ii) Second, there is a significant difference with regard to *time*: Mendelian genetics has been accepted, in one form or another, for at least a century, whereas all of the theories Stanford discusses

[23] I am assuming here that Mendelian genetics is at least as successful empirically as each of the other theories, and thus that it has at least as high PLAUSIBILITY.

– Darwin's, Galton's and Weismann's – were presented to, and rejected by, the scientific community within the space of around three decades (roughly from 1868, when Darwin first presented his pangenesis theory, to the turn of the century, when Mendelian genetics first became widely known). (iii) Third, the scientific community of evolutionary biologists was of course significantly smaller and less well organized at the time than it is now.[24] (iv) Finally, all of these theories are in the same *domain*. So, in sum, we again have that *domain* favors neither Mendelian genetics nor any of the other three theories, *incentives* favors the Mendelian genetics at least somewhat, and *community* and *time* both strongly favor Mendelian genetics.

The general point here, illustrated but not constituted by these examples, is that many of the theories that we accept today are quite dissimilar with regard to SENSITIVITY to those theories that are used by Stanford to infer that scientific theories are typically underdetermined. But of course, this is a distinction that really makes a difference, since SENSITIVITY is crucial in determining how likely theories are to be underdetermined. Thus it seems that Stanford's inductive argument is just too weak to threaten many currently accepted and empirically successful scientific theories. Note that this is not to deny that one can find *some* empirically successful and widely accepted theories about which the historical record suggests that we should be skeptical, e.g. theories that have recently been proposed, form part of some "fringe-science", or that concern some aspects of reality that we have reason to believe is exceptionally hard to conceptualize.[25]

It must also be admitted that there are historical cases in which a successful theory with an apparently high degree of SENSITIVITY did turn out to have an underdetermination rival. The obvious example here is Newtonian mechanics, which was replaced by Einstein's theories of special and general relativity after being accepted for a long time by a

[24]It is probably no coincidence that Galton was Darwin's half-cousin. What are the odds that the major theorists in contemporary biology would be closely related?

[25]Perhaps string theory is an example of such a theory, although I suppose it is debatable whether string theory counts as an *accepted* and *empirically successful* theory.

fairly large scientific community. But of course, we should expect *some* theories with high SENSITIVITY to be underdetermined, since a high degree of SENSITIVITY is not meant to *guarantee* that a theory is not underdetermined. What I have argued is not that we can somehow completely avoid the threat of the historical record – indeed, we should be suspicious of any realism that claims to be able to ignore history altogether. Rather, what I have shown is that Stanford's New Induction poses no *significant* threat to a modest realism on which most empirically successful theories with a relatively high degree of SENSITIVITY are taken to be epistemically justified.

4.6 The No-Privilege Argument

I have argued that the New Induction – arguably the most prominent route to anti-realism today – fails to apply to theories with a relatively high degree of SENSITIVITY, and thus does not establish a version of **NUT** that would undermine our justification for believing what many successful theories say about unobservables. However, as I mentioned in section 2, there is a slightly different sort of anti-realist argument for **NUT** to which we will need to reply as well, viz. the No-Privilege Argument. On one natural way of reconstructing the argument, it presents an explanatory challenge to the realist, and then argues that the challenge cannot be met. The challenge is to explain why it would be the case that scientists happen to consider the theories that are best supported given our evidence – what makes scientists "epistemically privileged" in this way? If the realist cannot offer an adequate answer, then the anti-realist could respond that the presumption of privilege lacks rational foundation and should be abandoned. (Although versions of this argument have been advanced by Sklar (1981), Roush (2005), and Wray (2011), I shall focus on the version given by van Fraassen (1989), since his is arguably the most developed, sophisticated and widely-discussed version.)

Now, van Fraassen (1989) does consider two possible explanations for our epistemic privilege, arguing that they are both inadequate. The first explanation is "Darwinian": On this explanation, being competent at scientific theorizing is selected for by natural selection, and thus we have a natural ability to discover the theories that are best supported given our evidence, ensuring that our theories are rarely underdetermined. Van Fraassen rejects this explanation as implausible because, he points out, it is not clear why a purely theoretical ability of this kind would have been selected for. Of course, one could reply to this by pointing out that this theoretical ability might be a *spandrel* – a byproduct of some other feature that was selected for, e.g. the ability to reason about one's environment. However, although this might explain how we *could* come to discover well-supported theoretical alternatives, van Fraassen might well reply that it would not explain why would be as reliable theorizers as the realist claims we are. That would require some sort of special faculty for scientific theorizing that would have given us no apparent evolutionary benefit when that faculty would have been selected for.

The second explanation that van Fraassen considers posits an almighty deity, which would ensure that we had a natural tendency to consider all the theories that are best supported given our evidence, thus again ensuring that our theories are not underdetermined. As van Fraassen argues, this assumes that such a deity has any interest in us having correct theoretical beliefs – which is doubtful. Besides, it would be disappointing if realism could only be defended by assuming theism (especially since many realists will take the success of science to undermine religious belief).

It is not clear that van Fraassen means for us to take these explanations seriously – they appear rather to be an attempt to draw out the absurdity of assuming that we are epistemically privileged. Clearly, the absurdity is supposed to consist in the appeal to a purported *innate* ability to hit upon the right hypotheses. As van Fraassen says, he thinks the "idea [behind epistemic privilege] is to glory in the belief that we are *by nature* [my emphasis] predisposed to hit on the right range of hypotheses." (van Fraassen 1989, 160). This does indeed appear absurd, for as van Fraassen goes a long way to show, any explanation for our innate epistemic privilege would be implausible.

Fortunately, however, epistemic privilege need not be due to any innate feature of our psychology. In particular, my discussion of SENSITIVITY suggests that epistemic privilege can be explained by the *socio-historical* nature of the scientific enterprise. Consider the "identifiable indicators" of SENSITIVITY discussed in section 4: *incentives*, *time*, *community* and *domain*. I have already argued that when favorable, these factors make us likely to have discovered an underdetermination rival if it existed. Thus when these factors are favorable, we are indeed "epistemically privileged" in the relevant sense – not individually, but as a group. And, of course, these factors are not *innate* in any sense – we do not possess them "by nature". Rather, we have *created* an enterprise, science, that puts us in a position where these factors are present for many of the theories we care most about. It is the socio-historical nature of this enterprise rather than the innate features of any of the individuals working within it that makes us epistemically privileged in the relevant sense.[26]

To see this more clearly, consider an analogy: We have no innate ability to construct the precise time-telling devices we call "clocks". Yet we do not think that the fact that we have clocks is miraculous. We have invented such devices because (i) precise time-keeping is very important to us, so there have been great *incentives* for inventing clocks, (ii) we have had plenty of *time* to invent and construct such devices, (iii) there is and has been a relatively large *community* of well-trained inventors and engineers working on clocks, and (iv) clock-construction is part of a relatively manageable *domain* of engineering. So the sorts of factors that explain our ability to construct precise time-telling devices are (at least to a significant extent) acquired characteristics that arise out of the socio-historical nature of technological advancement. There is no need to posit an innate ability for clock-construction in order to explain the existence of clocks.

Similarly, our ability to discover those rivals that are best supported given our evidence can be explained by the four factors being present in the case of many scientific theories.

[26]Indeed, I am inclined to see the most basic mistake behind the No-Privilege Argument as a part-whole fallacy: It is inferred that since any single individual scientist would not be epistemically privileged working on her own, a scientific community composed of such individuals will not be epistemically privileged either. This, as we have seen, is a fallacious inference.

In this way, we have a plausible explanation for our "epistemic privilege" – our ability to reliably discover those theories that are likeliest to be true given our evidence. Of course, no one ever said that we are *always* epistemically privileged, so we should not attempt to explain any such purported fact. But we can explain why we are privileged with respect to the theories that I recommended we take a realist attitude towards, viz. those with a relatively high SENSITIVITY: Our privilege in such cases is explained by the fact that to the extent that *incentives*, *time*, *community* and *domain* are favorable for a given theory, it is likely that someone or other in the scientific community has discovered underdetermination rivals if any such rivals existed. Contrary to the No-Privilege Argument, this does not assume the admittedly implausible claim that we have some innate ability to consider such rivals.

4.7 Conclusion

As my responses to the two routes to underdetermination make clear, the key notion in this paper is SENSITIVITY – the extent to which it's plausible that one would have found things of a certain kind given that there are such things. When applied to what I called "underdetermination rivals", i.e. theories that are equally or better supported by one's current evidence than accepted theories, I argued that SENSITIVITY varies with four "identifiable indicators", viz. the *incentives* for discovering rival theories, the nature of the *community* that potentially discovers such theories, the *time* that a theory has been accepted, and finally the psychological accessibility of the *domain* of the theory in question. Note that to a large extent these factors concern the history and sociological structure of scientific communities. Such factors have tended to be ignored or at least downplayed by scientific realists, and emphasized by what may be described as their polar opposites – so-called "social constructivists" about science.

One of the upshots of this paper is that realists have done themselves a disservice in this respect. Instead of avoiding the history and sociology of scientific communities, realists should see investigations in these fields as potential sources of considerations *in favor* of scientific realism. Relatedly, I have suggested that a plausible version of scientific realism – one that avoids Stanford's New Induction and van Fraassen's No-Privilege Argument – is a "selective" realism that selects as justified those claims made by successful scientific theories that enjoy a certain kind of privileged status within scientific communities. While I have sketched an account of what this privileged status would consist in, a fuller account would need to be based on a careful study of the actual structure and development of scientific communities. I thus conclude that defenders of scientific realism would do well to pay greater attention to historical and sociological studies of scientific communities.

5 EPILOGUE

The aim of the three papers collected in this dissertation has been to present a *realist* epistemology of science. The first paper argued that one should accept a scientific theory, for explanatory and predictive purposes, only if it's reasonable to believe that the theory is *true* (as opposed to merely consistent with all actual or possible observations). The second paper argued for a particular way of guaranteeing that the set of hypotheses to which we should be prepared to appeal in scientific explanations is both consistent and closed under logical operations. The final paper replied to a prominent challenge to realism, viz. a recent version of the so-called underdetermination of theory by evidence, arguing that the arguments for this thesis do not challenge our justification for believing that a suitably restricted set of scientific theories are true. I'd like to end by briefly mentioning some directions for future work based on two of the three papers.

The second paper, "Explanatory Acceptability" presented a view in which "probabilistic optimality" played a central role. While probabilistic optimality has already proven fruitful in this way, I believe that we have not yet seen the full extent to which this is true. That is, I believe that probabilistic optimality can form the basis of future work as well – indeed, two such projects are already well on their way.[1] On the one hand, as I hint at towards the end of chapter 3, there is room for a probabilistic reinterpretation of Inference to the Best Explanation (IBE), unifying two quite distinct approaches in epistemology and philosophy of science, viz. "Bayesianism" and "Explanationism". In particular, I argue in one paper that such a unification can remove some of the subjectivity of standard Bayesian

[1] At one point, I aimed to include the papers in which these projects are spelled out in this very dissertation. However, I slowly and hesitantly came to realize that the one and a half year that I had to write this dissertation would not suffice to present these papers in a manner that I would myself be satisfied with. Hence although these papers have already gone through several drafts, I have chosen not to include them in the dissertation.

approaches while precisifying the notoriously vague formulation of IBE. (Dellsén c) In another paper, I argue that Explanationism needs an account of roughly what I have called "explanatory rivals", and that such an account can help its proponents in deflecting some familiar worries about philosophical applications of IBE. (Dellsén b)

Another project based on probabilistic optimality examines the notions of *evidence* and *confirmation* as used by formal epistemologists and philosophers of science. It can be argued, I believe, that the Bayesian account of confirmation as probability-raising (E confirms H relative to $p(\cdot)$ just in case $p(H|E) > p(H)$) leaves out an important aspect of evidential support, viz. what I call "discriminating evidence". This sort of evidence, I argue, does not make a hypothesis more likely to be true – and thus does not count as confirming evidence on the Bayesian account – although it does make the hypothesis more acceptable for the purposes of explaining. Roughly, this is because it increases the probabilistic distance between a hypothesis and its most probable explanatory rival, thus making the hypothesis more probabilistically optimal than it would otherwise be. I believe that it can be argued that this sort of discriminating evidence is quite important in scientific practice, explaining among other things why scientists are not generally interested in falsifying or disconfirming "far out" theoretical alternatives. (Dellsén a)

A final direction for future work concerns the analysis in the third paper, "Realism and the Absence of Rivals", of when the absence of known instances of something supports the conclusion that no such things exist. There has been some discussion of such arguments in the philosophical literature, usually either under the heading of "absence of evidence" or "arguments from ignorance" (which is conceived of as a fallacy in informal logic). However, it seems to me that the *social* aspects of this epistemic phenomenon have not been fully appreciated.[2] To see this, note that the fact that I am now at least somewhat justified in believing that there are no pink elephants is due in no small part to the fact that I make

[2]With the notable exception of Goldberg (2010, chapter 6). Goldberg, however, analyzes the issue in a reliabilist framework, whereas I would be more inclined to favor a probabilistic analysis (along the lines spelled out in chapter 4).

the reasonable assumption that *others would have told me* if such things existed. In this way, our justification for believing that something is often based on a kind of *epistemic co-operation* that is worth exploring in more detail, not least because it seems to be an integral aspect of the scientific enterprise.

A APPENDIX: PROOFS OF THEOREMS

A.1 Proof of Theorem 1

I will prove Theorem 1 by proving three lemmas:

Lemma 1. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If H_i entails H_j, then

$$\max(G_S(H_i)) \leq \max(G_S(H_j))$$

Proof of Lemma 1: If H_i entails H_j, any global explanatory theory that entails H_i entails H_j (by the transitivity of entailment). Thus the/any[1] most probable global explanatory theory that entails H_i also entails H_j. Thus the/any most probable global explanatory theory that entails H_i cannot be more probable than the/any most probable theory that entails H_j.

Lemma 2. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If for some $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

(where $r \geq 1$) then for *any* $H_j \in \mathbf{H}$,

$$\max(G_S(H_i)) \geq \max(G_S(H_j))$$

[1] "The" if there is a single most probable such theory, "any" if there is more than one.

Proof of Lemma 2: Since any $H_i \in \mathbf{H}$ answers at least one of S's explanatory questions, any global explanatory theory either entails H_i or entails $\neg H_i$. Thus we can write the the probability of the/any most probable global explanatory theory there is as

$$\max\{\max(G_S(H_i)), \max(G_S(\neg H_i))\}.$$

Assuming that

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i)) \tag{A.1}$$

(where $r \geq 1$) we have

$$\max\{\max(G_S(H_i)), \max(G_S(\neg H_i))\} = \max(G_S(H_i)). \tag{A.2}$$

So $\max(G_S(H_i))$ is the probability of the/any most probable global explanatory theory there is. However, for any $H_j \in \mathbf{H}$, the global explanatory theories that entail H_j are but a subset of all global explanatory theories. Thus the probability of the/any most probable such theory, $\max(G_S(H_j))$, cannot be greater than the probability of the/any most probable global explanatory theory there is. Hence we have

$$\max(G_S(H_i)) \geq \max(G_S(H_j)), \tag{A.3}$$

as desired.

Lemma 3. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. For any $H_i, H_j \in \mathbf{H}$,

$$\max(G_S(H_i)) = \max(G_S(H_i \& H_j)),$$

$$or$$

$$\max(G_S(H_i)) = \max(G_S(H_i \& \neg H_j)).$$

Proof of Lemma 3: By definition, global explanatory theories provide *complete answers* to any explanatory question of S's, so any global explanatory theory that entails H_i either entails H_j or entails $\neg H_j$. The/any most probable global explanatory theory that entails H_i is therefore equal in probability to either the/any most probable global explanatory theory that entails both H_i and H_j, or the/any most probable global explanatory theory that entails both H_i and $\neg H_j$.

Theorem 1. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

then

$$\max(G_S(H_1 \& \ldots \& H_n)) > r \times \max(G_S(\neg(H_1 \& \ldots \& H_n))).$$

Proof: We prove that the theorem holds for $n = 2$, and note that the theorem clearly generalizes to any $n \in \mathbb{N}$. We start by assuming that

$$\max(G_S(H_1)) > r \times \max(G_S(\neg H_1)) \tag{A.4}$$

and

$$\max(G_S(H_2)) > r \times \max(G_S(\neg H_2)) \tag{A.5}$$

By Lemma 2, this entails:

$$\max(G_S(H_1)) \geq \max(G_S(H_2)) \tag{A.6}$$

and

$$\max(G_S(H_2)) \geq \max(G_S(H_1)) \tag{A.7}$$

so

$$\max(G_S(H_1)) = \max(G_S(H_2)). \tag{A.8}$$

Now, by Lemma 3, we have:

$$\begin{gathered} \max(G_S(H_1)) = \max(G_S(H_1 \& H_2)), \\ or \\ \max(G_S(H_1)) = \max(G_S(H_1 \& \neg H_2)) \end{gathered} \tag{A.9}$$

By (A.8), this is equivalent to:

$$\begin{gathered} \max(G_S(H_2)) = \max(G_S(H_1 \& H_2)), \\ or \\ \max(G_S(H_2)) = \max(G_S(H_1 \& \neg H_2)) \end{gathered} \tag{A.10}$$

Since $H_1 \& \neg H_2$ entails $\neg H_2$, we have by Lemma 1:

$$\max(G_S(H_1 \& \neg H_2)) \leq \max(G_S(\neg H_2)). \tag{A.11}$$

By (A.5), this entails

$$\max(G_S(H_2)) > r \times \max(G_S(H_1 \& \neg H_2)) \tag{A.12}$$

Since $r \geq 1$, this entails that

$$\max(G_S(H_2)) \neq \max(G_S(H_1 \& \neg H_2)) \tag{A.13}$$

So, by (A.10), we have:

$$\max(G_S(H_2)) = \max(G_S(H_1 \& H_2)) \tag{A.14}$$

And by (A.8), we also have:

$$\max(G_S(H_1)) = \max(G_S(H_1 \& H_2)) \tag{A.15}$$

Using (A.14) and (A.15) to substitute into (A.4) and (A.5) respectively, we get:

$$\max(G_S(H_1 \& H_2)) > r \times \max(G_S(\neg H_1)) \tag{A.16}$$

$$\max(G_S(H_1 \& H_2)) > r \times \max(G_S(\neg H_2)) \tag{A.17}$$

Now consider $\max(G_S(\neg(H_1 \& H_2)))$. Since both H_1 and H_2 answer some of S's explanatory questions, any global explanatory theory either entails H_1 or entails $\neg H_1$, and also either entails H_2 or entails $\neg H_2$. So we can divide all global explanatory theories into four (mutually exclusive and collectively exhaustive) categories: (i) those that entail $\neg H_1 \& H_2$, (ii) those that entail $H_1 \& \neg H_2$, (iii) those that entail $\neg H_1 \& \neg H_2$, and (iv) those that entail $H_1 \& H_2$. Only the global explanatory theories that fall into categories (i)-(iii) entail $\neg(H_1 \& H_2)$. Thus we have:

$$\begin{aligned}\max(G_S(\neg(H_1 \& H_2))) = \max\{&\max(G_S(\neg H_1 \& H_2)), \\ &\max(G_S(H_1 \& \neg H_2)), \max(G_S(\neg H_1 \& \neg H_2))\}\end{aligned} \tag{A.18}$$

By Lemma 1, since $\neg H_1 \& H_2$ entails $\neg H_1$, $H_1 \& \neg H_2$ entails $\neg H_2$, and $\neg H_1 \& \neg H_2$ entails (say) $\neg H_1$, we have:

$$\max(G_S(\neg H_1 \& H_2)) \leq \max(G_S(\neg H_1)), \tag{A.19}$$

$$\max(G_S(H_1 \& \neg H_2)) \leq \max(G_S(\neg H_2)), \tag{A.20}$$

and

$$\max(G_S(\neg H_1 \& \neg H_2)) \leq \max(G_S(\neg H_1)). \tag{A.21}$$

From (A.18) (A.19), (A.20), and (A.21), we now get:

$$\max(G_S(\neg(H_1 \& H_2)) \leq \max\{\max(G_S(\neg H_1)), \max(G_S(\neg H_2))\} \tag{A.22}$$

We now consider two cases, proving the result for each case. For the first case, suppose $\max(G_S(\neg H_1)) \geq \max(G_S(\neg H_2))$. Then:

$$\max(G_S(\neg(H_1 \& H_2)) \leq \max(G_S(\neg H_1)) \tag{A.23}$$

Multiplying both sides by the positive number r, we get:

$$r \times \max(G_S(\neg(H_1 \& H_2)) \leq r \times \max(G_S(\neg H_1)) \tag{A.24}$$

This together with (A.16) gets us the desired result:

$$\max(G_S(H_1 \& H_2)) > r \times \max(G_S(\neg(H_1 \& H_2)). \tag{A.25}$$

For the second case, suppose $\max(G_S(\neg H_1)) < \max(G_S(\neg H_2))$. Then:

$$\max(G_S(\neg(H_1 \& H_2)) \leq \max(G_S(\neg H_2) \tag{A.26}$$

Multiplying both sides by the positive number r, we get:

$$r \times \max(G_S(\neg(H_1 \& H_2)) \leq r \times \max(G_S(\neg H_2)) \tag{A.27}$$

This together with (A.17) gets us the desired result again:

$$\max(G_S(H_1 \& H_2)) > r \times \max(G_S(\neg(H_1 \& H_2)). \tag{A.28}$$

A.2 Proof of Theorem 2

For the proof of Theorem 2, we start by proving a final lemma:

Lemma 4. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. For any $H_i, H_j \in \mathbf{H}$ such that H_i entails H_j: If

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

then

$$\max(G_S(H_j)) > r \times \max(G_S(\neg H_j)).$$

Proof: We assume the antecedent, i.e. that

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i)) \tag{A.29}$$

where H_i deductively entails H_j. By Lemma 1,

$$\max(G_S(H_i)) \leq \max(G_S(H_j)). \tag{A.30}$$

Since H_i entails H_j, $\neg H_j$ entails $\neg H_i$. Thus again by Lemma 1 we have:

$$\max(G_S(\neg H_j)) \leq \max(G_S(\neg H_i)). \tag{A.31}$$

(A.29), (A.30) and (A.31) entail

$$\max(G_S(H_j)) > r \times \max(G_S(\neg H_j)), \tag{A.32}$$

as desired.

Theorem 2. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

and if $\mathbf{H}$ entails some H_c (not necessarily in $\mathbf{H}$) that also answers at least one of S's explanatory questions, then

$$\max(G_S(H_c)) > r \times \max(G_S(\neg H_c)).$$

Proof: Assume that for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i)). \tag{A.33}$$

Then by Theorem 1,

$$\max(G_S(H_1 \& \ldots \& H_n)) > r \times \max(G_S(\neg(H_1 \& \ldots \& H_n))). \tag{A.34}$$

Now, if $\mathbf{H}$ entails H_c, then $(H_1 \& \ldots \& H_n)$ entails H_c. So by Lemma 4, (A.34) entails

$$\max(G_S(H_c)) > r \times \max(G_S(\neg H_c)), \tag{A.35}$$

as desired.

A.3 Proof of Theorem 3

Theorem 3. Let $\mathbf{H} = \{H_1, \ldots, H_n\}$ be a set of hypotheses each of which answers at least one of S's explanatory questions. If for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i))$$

then $\mathbf{H}$ is consistent.

Proof: Assume the antecedent, i.e. that for all $H_i \in \mathbf{H}$,

$$\max(G_S(H_i)) > r \times \max(G_S(\neg H_i)). \tag{A.36}$$

Assume for *reductio* that $\mathbf{H}$ is *not* consistent. Then $(H_1 \& \ldots \& H_n)$ entails a contradiction, $\perp$. So by Theorem 2, (A.36) entails:

$$\max(G_S(\perp)) > r \times \max(G_S(\neg\perp)). \tag{A.37}$$

Since the probability of any proposition that entails a contradiction is 0, this entails:

$$0 > r \times \max(G_S(\neg\perp)). \tag{A.38}$$

Since $r \geq 1$, this entails that

$$\max(G_S(\neg\perp)) < 0. \tag{A.39}$$

However, since $\max(G_S(\neg\perp))$ is a probability (and thus cannot be less than 0), (A.39) cannot be true. So $\mathbf{H}$ must be consistent.

REFERENCES

Alspector-Kelly, M. (2001). Should the Empiricism be a Constructive Empiricist? *Philosophy of Science*, 68:413–433.

Alston, W. (1996). Belief, Acceptance, and Religious Faith. In Jordan, J. and Howard-Snyder, D., editors, *Faith, Freedom, and Rationality*. Rowman & Littlefield, Lanham, MD.

Blackburn, S. (1984). *Spreading The Word*. Clarendon, Oxford.

Blackburn, S. (2002). Realism: Deconstructing the Debate. *Ratio (new series)*, 15:111–133.

Boyd, R. (1980). Scientific Realism and Naturalistic Epistemology. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 613–662.

Boyd, R. (1984). On the Current Status of the Realism Debate. In Leplin, J., editor, *Scientific Realism*. University of California Press, Berkeley, CA.

Bromberger, S. (1965). An Approach to Explanation. In Butler, R. J., editor, *Analytical Philosophy, Second Series*, pages 72–105. Blackwell, Oxford.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press, Oxford.

Cartwright, N. (2007). Why Be Hanged for Even a Lamb? In Monton, B., editor, *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. van Fraassen*, pages 32–45. Oxford University Press, Oxford.

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge University Press, Cambridge.

Churchland, P. M. (1985). The Ontological Status of Observables: In Praise of the Superempirical Virtues. In Churchland, P. M. and Hooker, C. A., editors, *Images of Science*, pages 35–48. University of Chicago Press.

Cohen, L. J. (1980). Some Historical Remarks on the Baconian Conception of Probability. *Journal of the History of Ideas*, 41:219–231.

Cohen, L. J. (1989). Belief and Acceptance. *Mind*, 93:367–389.

Cohen, L. J. (1992). *An Essay on Belief and Acceptance*. Clarendon Press, Oxford.

Cross, C. B. (1991). Explanation and the Theory of Questions. *Erkenntnis*, 34:237–260.

de Sousa, R. B. (1971). How to Give a Piece of Your Mind: Or, the Logic of Belief and Assent. *The Review of Metaphysics*, 25:52–79.

Dellsén, F. Discriminating Evidence and Acceptance. Unpublished manuscript.

Dellsén, F. Explanatory Rivals. Unpublished manuscript.

Dellsén, F. Inference to the Likeliest Explanation. Unpublished manuscript.

Duhem, P. (1954/1982). *The Aim and Structure of Physical Theory*. Princeton University Press, Princeton, NJ.

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

Foley, R. (1979). Justified Inconsistent Beliefs. *American Philosophical Quarterly*, 16:247–257.

Foley, R. (1992). The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly*, 29:111–124.

Godfrey-Smith, P. (2008). Recurrent Transient Underdetermination and the Glass Half Full. *Philosophical Studies*, 137:141–148.

Goldberg, S. (2010). *Relying on Others*. Oxford University Press, Oxford.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, Cambridge.

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays*, chapter Aspects of Scientific Explanation. Free Press, New York, NY.

Hempel, C. G. (1966). *Philosophy of Natural Science*. Prentice-Hall, Englewood Cliffs, NJ.

Hempel, C. G. and Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15:135–147.

Henderson, L. (2013). Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science*.

Hesse, M. (1976). Truth and the Growth of Scientific Knowledge. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 261–280.

Horwich, P. (1991). On the Nature and Norms of Teoretical Commitment. *Philosophy of Science*, 58:1–14.

Jeffrey, R. C. (1956). Valuation and Acceptance of Scientific Hypotheses. *Philosophy of Science*, 23:237–246.

Jeffrey, R. C. (1970). Dracula Meets Wolfman: Acceptance vs. Partial Belief. In Swain, M., editor, *Induction, Acceptance, and Rational Belief*. D. Reidel.

Kaplan, M. (1981a). A Bayesian Theory of Rational Acceptance. *Journal of Philosophy*, 78:305–330.

Kaplan, M. (1981b). Rational Acceptance. *Philosophical Studies*, 40:129–145.

Kaplan, M. (1995). Believing the Improbable. *Philosophical Studies*, 77:117–146.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge University Press, Cambridge.

Kitcher, P. (1990). The Division of Cognitive Labor. *Journal of Philosophy*, 87:5–21.

Kitcher, P. (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press, New York.

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

Kitcher, P. (2001b). *Science, Truth, and Democracy*. Oxford University Press, Oxford.

Kuhn, T. S. (1962/1996). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, 3rd edition.

Kuhn, T. S. (1977). Objectivity, Value Judgments, and Theory Choice. In *The Essential Tension*. University of Chicago Press, Chicago, IL.

Kukla, A. (1993). Laudan, Leplin, Empirical Equivalence and Underdetermination. *Analysis*, 53:1–7.

Kyburg, H. E. (1961). *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT.

Kyburg, H. E. (1970). Conjunctivitis. In Swain, M., editor, *Induction, Acceptance, and Rational Belief*, pages 55–82. D. Reidel, Dordrecht.

Ladyman, J. (2007). The Epistemology of Constructive Empiricism. In Monton, B., editor, *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. van Fraassen*, pages 46–61. Oxford University Press, Oxford.

Lance, M. N. (1995). Subjective Probability and Acceptance. *Philosophical Studies*.

Lange, M. (2002). Baseball, Pessimistic Inductions and the Turnover Fallacy. *Analysis*, 62:281–285.

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

Laudan, L. (1997). How about Bust? Factoring Explanatory Power Back into Theory Evaluation. *Philosophy of Science*, 67:306–316.

Lehrer, K. (1979). The Gettier Problem and the Analysis of Knowledge. In Pappas, G. S., editor, *Justification and Knowledge*, pages 65–78. D. Reidel Publishing Company.

Leplin, J. and Laudan, L. (1993). Underdetermination Underdeterred: Reply to Kukla. *Analysis*, 53:8–16.

Lewis, D. (1986). Causal Explanation. In *Philosophical Papers, volume II*, pages 214–240. Oxford University Press, Oxford.

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

Lipton, P. (1993). Is the Best Good Enough? *Proceedings of the Aristotelian Society*, 93:89–104.

Lipton, P. (2004). *Inference to the Best Explanation*. Routledge, London, 2nd edition.

Lycan, W. G. (2012). Explanationist Rebuttals (Coherentism Defended Again). *The Southern Journal of Philosophy*, 50:5–20.

Maher, P. (1986). The Irrelevance of Belief to Rational Action. *Erkenntnis*, 24:363–384.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press, Cambridge.

Makinson, D. C. (1965). The Paradox of the Preface. *Analysis*, 25:205–207.

Maxwell, G. (1962). The Ontological Status of Theoretical Entities. In Feigl, H. and Maxwell, G., editors, *Minnesota Studies in the Philosophy of Science*, volume 3, pages 3–15. Minnesota University Press.

McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science*, 54:553–567.

Mitchell, S. (1988). Constructive Empiricism and Anti-Realism. In *PSA 1988: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1, pages 174–180.

Muller, F. and van Fraassen, B. C. (2008). How to Talk about Unobservables. *Analysis*, 68:197–205.

Nelkin, D. K. (2000). The lottery paradox, knowledge, and rationality. *The Philosophical Review*, 109:373–409.

Okasha, S. (2000). Van Fraassen's Critique of Inference to the Best Explanation. *Studies in the History and Philosophy of Science*, 31:691–710.

Pearson, K. (1911). *The Grammar of Science, 3rd edition*. Macmillan, New York, NY.

Poincaré, H. (1952). *Science and Hypothesis*. Dover, New York. Republication of the first English translation, published by Walter Scott Publishing, London, 1905.

Psillos, S. (1996). On Van Fraassen's Critique of Abductive Reasoning. *The Philosophical Quarterly*, 46:31–47.

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

Psillos, S. (2007a). The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research*, 74:441–448.

Psillos, S. (2007b). Putting a Bridle on Irrationality: An Appraisal of van Fraassen's New Epistemology. In Monton, B., editor, *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. van Fraassen*, pages 134–164. Oxford University Press, Oxford.

Putnam, H. (1978). *Meaning and the Moral Sciences*. Routledge and Kegan Paul, Boston.

Railton, P. (1989). Explanation and Metaphysical Controversy. In Kitcher, P. and Salmon, W. C., editors, *Minnesota Studies in the Philosophy of Science, vol. 13*, pages 220–252. University of Minnesota Press.

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

Salmon, W. C. (1975). Confirmation and relevance. In Maxwell, G. and Anderson, R., editors, *Minnesota Studies in the Philosophy of Science, Vol. 6, Induction, Probability, and Confirmation*, pages 3–36. University of Minnesota, Minneapolis, MT.

Salmon, W. C. (1989). Four Decades of Scientific Explanation. In Kitcher, P. and Salmon, W. C., editors, *Scientific Explantion. Minnesota Studies in the Philosophy of Science*, pages 3–219. University of Minnesota Press, Minneapolis, MT.

Salmon, W. C. (1990). Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes. In Savage, W., editor, *Minnesota Studies in the Philosophy of Science, Vol. 14*, pages 175–204. Minnesota University Press.

Salmon, W. C. (1992). Scientific Explanation. In et al., M. H. S., editor, *Introduction to the Philosophy of Science*, pages 7–41. Prentice-Hall, Englewood Cliffs, NJ.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63:129–138.

Sklar, L. (1981). Do Unborn Hypotheses Have Rights? *Pacific Philosophical Quarterly*, 62:17–29.

Sober, E. (1985). Constructive Empiricism and the Problem of Aboutness. *British Journal for the Philosophy of Science*, 36:11–18.

Sober, E. (1993). Epistemology for Empiricists. *Midwest Studies in Philosophy*, 18:39–61.

Sober, E. (2009). Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection with Fossils, Fishing, Fine-tuning, and Firing Squads. *Philosophical Studies*, 143:63–90.

Stanford, P. K. (2001). Refusing the Devil's Bargain: What Kind of Underdetermination Should We Take Seriously? *Philosophy of Science (Proceedings Supplement)*, 68:S1–S12.

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

Stanford, P. K. (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.

Strevens, M. (2003). The Role of the Priority Rule in Science. *Journal of Philosophy*, 100:55–79.

Strevens, M. (2009). Objective Evidence and Absence: Comment on Sober. *Philosophical Studies*, 143:91–100.

Teller, P. (2001). Whither Constructive Empiricism. *Philosophical Studies*, 106:123–150.

Tversky, A. and Kahneman, D. (1984). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement. *Psychological Review*, 91:293–315.

van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon, Oxford.

van Fraassen, B. C. (1985). Empiricism in the Philosophy of Science. In Churchland, P. M. and Hooker, C. A., editors, *Images of Science*. Chicago University Press, Chicago, IL.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Clarendon Press, Oxford.

van Fraassen, B. C. (2001). Constructive Empiricism Now. *Philosophical Studies*, 106:151–170.

Weinberg, S. (1995). *The Quantum Theory of Fields*. Cambridge University Press, Cambridge.

Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese*, 167:125–143.

Williamson, T. (1996). Knowing and Asserting. *Philosophical Review*, 105:489–523.

Worrall, J. (1989). Structural Realism: The Best of Both Worlds. *Dialectica*, 43:99–124.

Worrall, J. (1994). How to Remain (Reasonably) Optimistic: Scientific Realism and the "Luminiferous Ether". In Richard M. Burian, D. H. and Forbes, M., editors, *PSA 1994: Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association*, volume 1, pages 334–342.

Wray, K. B. (2011). Epistemic Privilege and the Success of Science. *Nous*, 46:375–385.