
Re-thinking Scientific Realism: Structure and Beyond

Juha Tapani Saatsi

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Philosophy



September 2005

The candidate confirms that the work submitted is his own and that appropriate credit
has been given where reference has been made to the work of others

This copy has been supplied on the understanding that it is copyright material and that
no quotation from the thesis may be published without proper acknowledgement

Acknowledgements

This work owes a great deal to a number of people. First and foremost I want to express my gratitude to my supervisors, Steven French and Joseph Melia. They have had a huge influence on the ideas advanced in the following pages—although they probably agree with almost none of it—as well as on my general approach to philosophy. Steven’s initial inspiration and encouragement to pursue philosophy of science has certainly been the most significant factor to my present intellectual and academic life. His humanity and open-mindedness towards a science graduate with no background in philosophy provided an extraordinary incubator for my philosophical development. Joe’s intellectual integrity and unswerving aspiration for ultimate clarity have set a lasting standard for me to try to live up to.

I have greatly benefited from discussions with a number of philosophers, peers and friends: Otávio Bueno, Angelo Cei, Anjan Chakravartty, Jon Hodge, Dean Rickles, Chris Timpson, and others. Some of the material has been presented at the conferences of the Philosophy of Science Society and the British Society for the Philosophy of Science. The comments received in those gatherings, as well as the comments received from a referee of *Studies in History and Philosophy of Science*, are gratefully acknowledged. The School of Philosophy has provided a hospitable, relaxing and inspirational environment for my PhD studies.

Special thanks go to my ever-supporting family, and especially to my newlywed wife Sari whom I have shared all these remarkable years with. I dedicate this work to you.

I am grateful for the financial support I have enjoyed during this work, from the University of Leeds (3 years) and the Finnish Cultural Foundation (1 year).

Juha Saatsi
13th September 2005

Abstract

This thesis examines the contemporary scientific realism debate, with a special focus on the various forms of structural realism. It comprises three parts.

The first part prefaces the work with a scrutiny of the principal arguments of the classic debate. The global and local explanationist arguments are critically analysed, and juxtaposed with the so-called experimental arguments for realism. It is argued that focusing on explanatory considerations does not serve the justificatory task the realist faces, but a local analysis of ampliative reasoning can nevertheless secure a level of realist commitment to a significant class of unobservables. This part also defends the anti-realist argument of Pessimistic Induction against two challenges that take it to be fallacious.

The second part looks at the main structural realist suggestions for an image of theoretical knowledge that harmonises with our best understanding of the current and past science. It concludes that both of these—epistemic Ramsey sentence realism in the syntactic-axiomatic framework, and the ontological structural realism in the semantic framework—are incomplete and inadequate responses to the anti-realist challenge. In addition to providing a comparative review of the various structuralist motivations and intuitions, this part contributes to the literature by clarifying the often referred to, but also by and large misunderstood, problem of unintended models faced by the Ramseyfying realist ('Newman's problem').

The third part begins by looking at the notion of approximate truth in detail, in order to argue that the traditional 'standard' realist alternative to structural realism is not the only alternative. Paying due attention to the explanatory requirements for the realist argument yields an informally articulated notion of *explanatory approximate truth* which gives rise to a fresh realist perspective: *eclectic realism* is realism about success-fuelling multiply realised properties. This part concludes the thesis by developing and defending this position, by conducting a detailed case study of the historical theory change from Fresnel's ether to Maxwell's electromagnetic theory of optics.

CONTENTS

Introduction	ix
I Foundations of the Realism Debate	1
1 Explanation for What?	3
1.1 Explanationism and its challenges	4
1.2 What is IBE?	7
1.2.1 From loveliness to likeliness	8
1.2.2 Responding to the challenges	11
1.3 Forms of global explanationism	14
1.3.1 What is the explanandum?	15
1.3.2 Is the explanation good enough?	20
1.4 Is NMA just a base-rate fallacy?	24
1.5 What really is wrong with global explanationism	30
1.6 Justifying IBE	33
2 Overcoming the Empiricist Challenge: Experimental Arguments	37
2.1 The Empiricist Challenge	38
2.2 Extending our senses	42
2.3 Entities as tools	48
2.4 Cartwright on the role of causal explanation	51
2.5 Achinstein on the reality of atoms	55
Appendix: Van Fraassen's Image of Science	58
3 Overcoming the Empiricist Challenge: Ampliative Inferences undivided	63
3.1 Local vs. Global realist strategies	64

3.1.1	Graduation of the realist commitment / warrant	65
3.1.2	The unity of the scientific method	68
3.2	Justification of low-level experimental inferences	69
3.3	Kitcher, Lipton and others on (not-so-) Local Realism	72
3.3.1	Kitcher's Galilean Strategy	72
3.3.2	Lipton and McMullin.	74
3.4	How local must you go?	77
3.5	How local can you go?	78
4	Two Challenges to the Realist Image	81
4.1	The Polemic	82
4.2	Challenge from Empirical Underdetermination	83
4.3	Challenge from Historical Theory Changes	86
4.3.1	Lange's Turnover Fallacy	88
4.3.2	Lewis's False Positives Fallacy	93
4.3.3	What Pessimistic Induction is	98
II	Structural Realism and its Structure	101
5	What is Structural Realism?	103
5.1	Epistemological Motivations	104
5.2	Ontological Motivations	109
5.2.1	Metaphysical Underdetermination.	109
5.2.2	Structuralism in physics vs. epistemology.	115
5.3	Meta-Scientific Frameworks	118
6	Structural Realism and Ramseyfication	125
6.1	Defining Ramseyfication	126
6.2	Ramseyfication and Structure?	126
6.3	Model-Theoretic Arguments	132
6.3.1	Formalising the Challenge	133
6.3.2	Ramseyfication and Theoretical Predicates	135
6.3.3	Model-Theoretic Argument Finessed	138
6.4	Exegetical Commentary	141
III	Beyond Structuralism	147
7	Explanatory Approximate Truth	149
7.1	Responding to Pessimistic Induction	149

7.2	Explaining theoretical success of rejected theories	152
7.3	Worrall on Explanatory Approximate Truth	157
7.4	Psillos on Explanatory Approximate Truth	160
7.5	EAT and success-fuelling properties	161
8	Explaining the success of Fresnel's theory	165
8.1	Reflection and refraction from Fresnel to Maxwell	165
8.2	Deriving Fresnel's equations	169
8.3	Comparing Fresnel to Maxwell	175
8.4	Fresnel and explanatory approximate truth	177
9	Towards Eclectic Realism	183
9.1	From scientific explanation to scientific understanding	184
9.2	Approximate truth and reductive explanation	188
9.3	Towards a novel formulation of realism	191

Introduction

This thesis concerns the epistemological question about the nature of scientific knowledge, and represents analytic philosophy of science. The analytic style of philosophy is much about conceptual analysis. Philosophers draw distinctions that are interesting and meaningful, and hopefully helpful in explaining, clarifying or usefully categorising concepts that we employ in thinking about ourselves and the world around us. The philosophical debate about scientific knowledge, *the realism debate*, is a relatively mature one; many important conceptual distinctions have been clearly made, and various ‘-isms’ have consolidated their positions on the philosophical battlefield. This thesis, as the title suggests, broadly reconsiders the realism debate and attempts to advance it by suggesting a number of further distinctions that, I will argue, clarify the debate and may shift the balance of power on this battlefield.

Among the most central distinctions drawn in the contemporary realism debate are the following. The distinction between humanly *observable and unobservable* matters is basic to epistemology. Science is about the mind-independent world (yes, there is such a thing!) and scientific theories and propositions are either *true or false* about something they purportedly tell us about. Scientific theories that are about something unobservable to us can be true about the observable matters, or *empirically adequate*, as opposed to *true* about the unobservable matters. Finally, *anti-realism* states that we cannot have knowledge of the unobservable (physical) matters through our scientific endeavour, whilst *realism* is defined as its complement: we can have some such knowledge.

The final distinction between realism and anti-realism thus defined divides

a vast field of opinions sharply into two over-arching categories, regarding the question of whether or not we can have *any* knowledge of *anything* unobservable. Given this, it really is no wonder that there are so many different arguments for realism; surely *an* argument can be found that secures knowledge of *something* unobservable? I am a realist at heart. I, too, feel that it is simply preposterous to claim that we can have no knowledge of the things that we cannot directly perceive ‘by the naked eye’.

Of course, properly arguing for this pre-theoretical intuition is a more complex matter. But the realism debate is hugely complicated *further* by the rich spectrum of realist positions of various degrees of epistemological ambitions, or ‘realist commitments’, defended through almost equally various arguments. The sprawling literature of the complicated debate easily gives the impression that defending the realist intuition is *really* hard work. I refuse to resign my intuitions about the absurdity of not having any knowledge of things I cannot directly perceive. And I want to lessen the burden of defending my intuitions by reshaping the realism debate so as to disassociate the arguments for those degrees of realist commitment that we regard as (more) obvious, intuitively, from the arguments that pertain to higher-levels of ambition.

But, as mentioned above, this thesis looks at the whole of the modern debate. My general approach is to focus on the ‘hot’ issues in the contemporary realism debate one by one, with only a minimal underlying agenda. I am a realist, but the reader may be disappointed to learn that there is no *full-blown* novel realist position on offer here. Rather, I put forward a series of careful considerations placed in certain argumentative *contexts* which I have deemed worthy of attention (and to which I have something interesting to say!). Hence I recurrently find myself sweating over subtle distinctions and painstakingly constructed arguments, only to express discontent in the end about the basic premises on which the particular micro-debate hangs.

In the first Part of the thesis I try to make sense of the sprawling debate by advancing several distinctions that I deem useful at the general, foundational level. The distinctions drawn in the second and third Parts are concerned with more specific argumentative contexts, as circumscribed below.

Foundations. I think there is a particularly helpful meta-philosophical distinction to be had for organising the realism debate. Roughly speaking, formulating and defending a full-blown realist position requires two things. One needs to provide a description of the realist commitments regarding our theoretical knowledge, and one needs to provide an argument for taking that description to be truthful. Both of these steps are needed and neither is trivial.

More specifically, the description of realist commitments covers the grounding and the content of the realist beliefs. Scientific beliefs about the unobservable are ampliative by virtue of going beyond both what has been observed and what is observable. The realist believes that at least some of those scientific beliefs are warranted. The first descriptive task ('grounding') is to account for the inference(s) the realist makes from the nature of the scientific beliefs and of the scientific inferences leading to those beliefs, to their warranted epistemic status. Secondly, we want to know what exactly in the huge corpus of scientific beliefs the realist takes to be warranted ('content'). I call this double-task *the issue of the realist image*. The other side of the coin is *the issue of justification*: the realist needs to provide an argument to justify those ampliative scientific inferences that in her books yield warranted beliefs. Or, in other words, the realist needs to justify her realist inference(s) to her realist belief that some of the scientific beliefs are warranted.

The justificatory dimension of the realist project has close affinities with the more general problem of justifying any ampliative inferences, the problem of induction. Although the character of the particular inference is taken into account and analysed—as an inference to the best explanation, for example—the general argumentative mode is largely a priori. The descriptive dimension of the realist project, by contrast, is heavily based on our best understanding of the *actual* science, both current and past. The realist image needs to cohere with our best understanding of the actual science to be a plausible description of the scientific knowledge gained from the actual science. What gives philosophical currency to the work done in this dimension is the fact that science itself undermines, at least potentially, the plausibility of any straightforward description of the grounding and content of the realist beliefs.

Since both of these ingredients are needed for realism the distinction between the issues of image and justification may feel slightly artificial and

tenuous, but it proves to be rather useful for categorising and making sense of the various arguments in the debate. To evaluate a specific realist argument we can often situate it in the context of one of the issues and bracket the implicit concerns regarding the other. For example, at the more ambitious levels of realist commitment there has arguably been quite discernible progress with respect to the issue of image, whilst the achievements regarding the issue of justification still remain rather modest.

Although there are three semi-independent parts to this thesis, thematically it can be divided into two. The first 80 pages (chapters §1–3) deal with the issue of justification, and the rest is about different ways to understand and defend the image. I offer several helpful distinctions regarding the numerous arguments dealing with the issue of justification. First of all, I differentiate between *global* and *local* arguments, and sharpen the vague and ambiguous distinction that is implicit in the literature. Secondly, I differentiate and compare some significantly different variations of the most global justificatory argument (No-Miracles). This is an abductive, *explanationist* argument: it appeals to the central role of explanatory considerations in science. There are more local explanatory arguments as well, and these in turn can be contrasted with a class of local arguments which try to do without explicit explanatory considerations. To this end I compare several *experimental* arguments for realism, and juxtapose these with arguments from *low-level analysis of ampliative inferences*.

Moving on to the issue of the realist image (chapter §4), at the foundational level I argue for the following points. There are two well-known challenges to the realist image: the arguments from empirical underdetermination, and from historical theory changes. Although this distinction has been clearly made in the literature, it is not fully clear how these two challenges interact. Painting a realist image is a more subtle affair than is generally acknowledged in that regard. In particular, the image needs to incorporate not only predictive but also *explanatory success*. I will be mostly concerned with the challenge from the history of science—the ‘Pessimistic Meta-Induction’—and we can discern different formulations of this anti-realist argument, not all equally powerful. Disentangling these formulations requires a careful consideration of what exactly the argument is an argument against.

The second and third parts look at specific solutions for delineating a

plausible realist image vis-à-vis the challenge above.

Structuralism. During the past decade or so *structuralism*, broadly construed, has become one of the most popular perspectives on the issue of the realist image. There are numerous ideas that go under this label and these need to be properly distinguished. In chapter §5 I contrast different motivations for structural realism, and different meta-scientific frameworks, or philosophical ‘theories of theories’, for spelling out what ‘structure’ is. I will be critical of the purported motivation for the metaphysical, or ‘ontic’ conception of structural scientific realism, whilst leaving room for ontic structuralism in the philosophy of physics and metaphysics. Evaluating the competing meta-scientific frameworks allows me to differentiate between two quite different embodiments of the shared structuralist intuition. Ultimately I deem both of these conceptions of structure seriously incomplete as solutions to the issue of the realist image.

I think there is something fundamentally unappealing in the *Ramsey-sentence* approach to structural realism. This conception emerges naturally in the syntactic-axiomatic framework of theories, but the framework itself is not a natural way to understand the actual science. Hence there is some tension between this framework and the broadly naturalistic tone of the project of providing a realist image, of providing a plausible description of the truth-tracking actual science. Regardless of this, the Ramsey-sentence approach has stirred discussion in abundance in the recent literature, and again there are important distinctions to be made that clarify the heated debate (chapter §6). I will argue that the true potential of the ‘Ramseyfying’ approach for yielding an interesting notion of ‘structure’ depends on the choice made regarding the class of extra-logical predicates eliminated by the Ramsey-elimination, together with the logical framework adopted for presenting the theory. Historically speaking the adoption of Ramseyfication by the structural realists is also interesting. It turns out that the historical notion of structuralism that motivates Ramseyfication is actually sufficiently distinct from the contemporary notion to throw doubt on the exegetical connection as commonly appealed to in the literature.

Whilst I was examining the details of a case study that has been presented both for and against structural realism, I came to think about some

conceptual distinctions that (a) allowed me to draw a line between the kinds of realist images that the case study had been taken to provide an argument for, and (b) made me leave aside the worries about the appropriate meta-scientific framework for expressing the notion of structure. The distinctions that I now take to be crucial for the issue of the realist image are discussed in the third and final part.

... and Beyond. Different responses to the issue of the realist image, as far as the challenge from the history of science is concerned, can be based on two strategies (and their combination). These correspond to the double-task of accounting for the grounding plus the content of the realist inference(s). In principle one could be able to undermine the basis of the historical worry (or at least some of it) by explicating the grounding of the realist inference(s) in terms that are plausible by virtue of being based on our best understanding of science, and allow the prima facie force of the historical worry to be discounted. The general consensus is that as a matter of fact this strategy is not adequate all by itself, and the realist needs to complement this by delineating the content of the realist image. Without offering a justification for this consensus I will focus on different ways to implement the second strategy.

Delineating the content of the realist commitment is done by articulating a notion of *approximate truth*. The basic idea is that the realist can defuse the basis of the historical challenge by *explaining* the success of a past theory by appealing to its approximate truth in the realist sense. Although the realist's explanatory project thus understood is clearly different from the scientists project of explaining worldly phenomena by appealing to a theory, I think more attention should be paid to this distinction. I analyse different possible modes of the philosopher's explanation of the success of past science. Regarding the case study I have examined (chapter §8), more specifically, there are interesting, subtle distinctions to be made about the properties that enable a successful *derivation* of a prediction to take place from seemingly completely false premises, and from a seemingly misleading *explanatory virtues* of the theory. I conclude the thesis by putting forward some tentative remarks about the best way to delineate the content of the realist image in the light of these distinctions.

Part I

Foundations of the Realism Debate

Explanation for What?

“This intuitive ‘no miracles’ argument can be made more precise in various ways—all of them problematic and some of them more problematic than others.”

John Worrall, *Structural Realism: the best of both worlds?*

The ampliative inferences of science typically proceed through theoretical hypotheses. Coming up with theories and hypotheses and judging their plausibility is what (theoretical) scientists do, and one of the most important criteria for a good theory is its explanatory power. This criterion really counts since explaining is what science does, in addition to *describing* and *predicting* observable phenomena. Let’s call *explanationism* any attempt to take seriously enough the explanatory aspirations of science to consider explanatory success to be a mark of theoretical truth. Clearly the explanationist conclusion does not follow from the undeniable indispensability for the scientific method of entertaining and judging theoretical hypothesis. For it could be just a feature of instrumental utility, an aspect of scientific pragmatics. The explanationist-realists have put forward a number of arguments to convince us of the connection postulated to hold between the explanatory dimension of science and its ability to track the truth about the unobservable world. Spelling out and evaluating the most prominent of these arguments is at the heart of this chapter.

1.1 Explanationism and its challenges

I want to begin with a couple of brief preliminary remarks, one about the concept of explanation, and another about the connection of explanationism and realism.

The idea of counting on explanatory power to give extra-empirical evidence for a theory becomes fully intelligible only when accompanied by a bona fide theory of explanation making precise what explanatory success amounts to. And not any old theory will do: the well-known Deductive-Nomological account conceived by the logical positivists (especially Hempel, 1965) does not help the explanationist, for example. Due to the symmetry it proposes to hold between explanation and prediction it does not allow demarcation between empirically underdetermined theoretical scenarios, and hence it does not support an explanationist argument for realism about any theory for which there exists an empirically equivalent incompatible rival. The DN-account is widely discredited for its faults, but what exactly is to take its place is still a matter of considerable debate. (cf. Woodward 2004, Lipton 2004) Spelling out clearly what exactly explanatory strength amounts to is evidently the first and foremost thing on the explanationist agenda. Nevertheless, most arguments and intuitions evaluated in this chapter operate at a level of generality that allows us to consider the merits and liabilities of explanationism independently of such details.¹

Is realism just co-extensive with explanationism? Jarrett Leplin, for one, takes the connection between the two to be rather close:

Realism minimally maintains . . . that our empirical knowledge could be such as to warrant belief in theory, on the basis of its explanatory success. (Leplin, 2000: 393)

Certainly the realist should say something about explanatory power and its role in the scientific methodology, but I would not recommend taking explanationism to be a necessary or defining feature of realism. This is because I cannot see any reason to deny the possibility of states of affairs (to also indulge in loose modal talk here) in which our empirical knowledge is such as to

¹In chapters §3 and §9 I will address different aspects of the more specific question of how the lack of a unitary philosophical account of explanation influences the realist project.

warrant belief in some theoretical propositions by simple forms of induction, for example, without any recourse to explanatory considerations. Unless, of course, all inductive reasoning is either banned or foundationally reduced to inference to the best explanation. The status of such IBE foundationalism will be addressed below, as will be in due course the more general question of the relationship between explanationism and realism vis-à-vis the actual science.

But certainly the explanationist thought is a significant potential route to realism—given the major role of explanatory considerations in scientific reasoning—and we should now consider what exactly can be said for and against it. To ease ourselves into the vast morass of the sprawling debate, here are at first a couple of rather intuitive anti-realist challenges, much discussed in the literature.

There is an aspect of explanatory practice that *prima facie* may seem to immediately count against the explanationist. For there to be an objective measure of explanatory goodness it is necessary to have a level of objectivity as to what counts as an explanation in the first place. But explanations are typically highly *contextual* in character: an explanation is an answer to a question suitably presented, and nothing seems to be an explanation simpliciter. We do not explain why an ice-cube melted in water, simpliciter, but why it melted *rather than* stayed in its frozen state, or why it melted rather than exploded. In all questions there is a contrast class, often implicit, that fixes the context. Also, theories *as such* are not explanatory, although they may provide explanations given certain interests and purposes. Since the latter dimension is broadly speaking pragmatic, perhaps explanatory power only makes sense as a pragmatic (as opposed to epistemic/evidential) concept. Something like this line of thought lies at the heart of van Fraassen's general attack on explanationism and his pragmatic theory of explanation. (1980: ch. 5)

Van Fraassen's is a purely *epistemic* account of explanation, perhaps the most well developed one since Hempel's DN-model. The epistemic approach is to be contrasted with those accounts that buy into some ontological resources (causation, nomic necessity, physical mechanism. . .) in drawing the distinction between what is explanatory and what is not.² If being explanatory

²Cf. Ruben (1993: Ch. 1) for a nice explication of this contrast between epistemic and

depends on the way the world is—and not just on the epistemic relationship between our description of the world and our knowledge and interests—an empiricist (like van Fraassen) is going to have hard time accommodating the fact about scientific practice that scientists aim to provide explanations and that explanatory power is considered to be a theoretical virtue. By contrast, in van Fraassen’s account an explanation does not have to be true to be good. Rather, by construing explanation as an answer to ‘why-questions’, and as a three-term relation between theory, fact and context (rather than a relation between theory and fact alone), van Fraassen challenges the objectivity of the notion of explanatory power so central to explanationism.

But is the kind of contextuality described above actually incompatible with accounts of explanation which draw also on ontological resources? Clearly it is not. For the same ontological complex state of affairs (say a causal chain) can consistently and objectively ground various explanations by virtue of different ‘parts’ or ‘aspects’ of it, as long as the explanations are not incompatible. Lipton’s theory of contrastive explanations in the causal framework is an example of such an account. (1990, 2004) In this framework the existence of implicit contrasts is actually turned into a virtue, since it can be used to account for how only some particular bits of the (enormous) causal history of an event are explanatory for each event for each contrast. This explains why the Big Bang is rarely cited as being genuinely explanatory for anything. (2004: 33) Hence explanationism does not immediately founder upon the interest relativity of explanatory practice, despite *prima facie* appearances to the contrary. Whether there is some reason to prefer an ontological theory of explanation over the empiricist one advanced by van Fraassen needs to be decided on more general grounds. For example, the latter has been criticised on the basis that it is too permissive unless some objective (ontological) constraint is imposed on it. (Kitcher and Salmon, 1987)

But even if we accept the need for an ontological account of explanation, we are still far from securing the main explanationist thesis. Van Fraassen has opposed the spirit of explanationism not only with his epistemic pragmatist account of explanation, but he also has, and perhaps more influentially,

ontic accounts of explanation, due to Salmon (1984). See also Wright & Bechtel (forthcoming) for an insightful deconstruction of this distinction in the context of mechanical explanation.

advanced some serious criticism of the basic idea that explanatory considerations could serve as a guideline in ampliative, inductive reasoning. To begin with, if understood as a *rule* of inductive inference, then arguably *the inference to the best explanation* cannot be a rule to be trusted:

... for it is a rule that selects the best among the historically given hypotheses. We can watch no contest of the theories we have so painfully struggled to formulate, with those no one has proposed. So our selection may well be the best of a bad lot. (van Fraassen, 1989: 143)

That is, the best we pick can merely be *the best of bad lot* and it is irrational to adopt such a defective rule. And if the advocate of IBE wants to try a more sophisticated tack by incorporating explanatory factors into a probabilistic model of belief change—so that explanatory considerations are linked to graduated degrees of belief, rather than simply taking the best to be true—then she becomes incoherent by virtue of conflicting with the Bayesian probability calculus for updating belief. So in van Fraassen's books the idea that explanatory power guides our ampliative inferences is irredeemably doomed from the point of view of explanationism. (van Fraassen, 1989: ch. 6)

These points of criticism have been, in my view, successfully responded to in the modelling of IBE subsequent to van Fraassen's *Laws and Symmetry* (1989). I will next present these models at a level of sophistication required for (a) understanding how the criticism from the bad lot is defused, and (b) subsequently evaluating the various uses to which IBE has been put in the realist arguments.

1.2 What is IBE?

The deep-seated belief underlying explanationism is the idea that explanatory considerations are a guide to inference. Furthermore, this is not to be understood as a merely *descriptive* thesis about the way we can model the actual inference making or understand its psychology, but as a claim about how ampliative inferences are tracking the truth. This essentially realist dimension of explanationism makes its defence highly challenging; even characterising the position tends to get convoluted due to the way in which the issues of description are intertwined with the issues of *justification*.

The advocates of IBE have advanced varying views about its status. Some take it to be a fundamental, primitive or foundational form of inductive inference (e.g. Harman 1965, Psillos 2002, Josephson 1996, 2000), whilst others argue for a more limited role (e.g. Day & Kincaid 1994, Lipton 2004). There are, of course, other forms of inductive inference, and it does not seem reasonable to claim that explanatory considerations are invariably involved in all the actual inductive reasoning. The descriptive thesis can be interpreted at many levels, however, and there is no need to advance a very rigid claim as regards to the pervasiveness of explanatory guidelines. There is an on-going debate about the level at which the descriptive thesis is defensible.³ What comes to the justification of IBE, there is an interesting debate about which grounds which, with respect to the relative merits of enumerative and other basic forms of inference versus explanatory guidance. It is this latter debate that we will be mostly concerned with here. But let's first explicate further the nature of these two projects—descriptive and justificatory—and their relationship.

1.2.1 From loveliness to likeliness

The 'best' in 'inference to the best explanation' is naturally understood in terms of some criteria used to evaluate the virtue of inferring it. As to the virtue, we can have different reasons for inferring some propositions—different understandings of the aim of science are a case in point—but the (realist) explanationist, of course, wants her inferences to be (as) true (as possible). The criteria employed, on the other hand, ought to refer to explanatory value for the position to be worthy of its name. It is natural to take 'the best' as 'the most explanatory', a notion the analysis of which then hangs on our understanding of explanation. Peter Lipton puts all this very concisely by saying that according to IBE explanatory *loveliness* is a guide to *likeliness*: the most probable explanation is searched for by looking for the one that provides the most understanding. The word 'likeliness' here very nicely captures the essential fallibility of inductive method: a (possible) explanation estimated to be the likeliest does not have to be an actual explanation (i.e. true) for any particular scenario, as long as the method works by and large.

³For example, see the review symposium of Lipton (2004) in *Metascience* (forthcoming).

This is all very general, of course, and any detailed account of explanatory loveliness obviously requires plugging in some preferred theory of explanation. Not much work has been done at that level, with the notable exception of Lipton's efforts at defending IBE in the context of causal-contrastive theory of explanation. Since the notion of explanation has resisted a unified philosophical analysis, the notion of IBE is fragmented relative to what exactly counts as explanatory. But we can for the time being operate at the more general level to evaluate the merits of the realist appeal to IBE.⁴

First of all, if the general defence of IBE is divided into descriptive and justificatory parts, as Lipton (2004) suggests, then clearly it is only the latter that is of central interest to the realist. This is simply because the descriptive claim that inference to the best explanation describes ampliative reasoning in human subjects in various situations can be advanced and defended without those inferences tracking the truth. That is, we can follow something like IBE *as if* it was guiding us towards truth (without ever getting there). This would be inference of the *as-if-likeliest* from the loveliest explanation. Lipton mostly focuses on this descriptive task, and the critics of the first edition of his book sometimes seem to conflate the distinction between the two tasks. (Achinstein 1992, Barnes 1995) Perhaps this is quite understandable given that Lipton's initial formulation of IBE in terms of estimated likeliness has an explicitly realist flavour although, to be fair, Lipton himself is very clear about the distinction:

Taken as a principled description of an aspect of scientific practice, Inference to the Best Explanation is no argument for realism. To say that scientific inferences follow a particular pattern does not in itself tell us whether what is inferred tends to be true. And one does not need to be a realist to endorse a version of Inference to the Best Explanation, descriptively construed. (Lipton, 1993: 92).

So the descriptive part of the challenge contributes precious little to the justification of ampliative inferences. The justificatory project can be independent of the correctness of the descriptive thesis at some level of description. And

⁴The fragmented nature of IBE is taken into account in the next chapter (§3.4), and again in chapter §9

not only does the descriptive thesis not entail the justificatory thesis for realism, but neither is the former a necessary condition for the latter. If it turned out that explanatory considerations do not guide some (or even most) of the scientific ampliative inferences, there could still be room for a realist justification of those inferences. It would just have to be independent of IBE.

In addition to the distinction between the descriptive and the justificatory projects there is a line to be drawn between two levels of the descriptive project. There is the more *ambitious*, broadly speaking naturalistic project of providing a description of the *actual* inferential practices we engage in both in everyday and in scientific contexts. And there is the more *modest* project of *modelling* our inferential practices in explanatory terms, without advancing a further claim about the degree to which the model fits the actual practice over and above the model being minimally adequate in yielding the same conclusions from the same inductive premises. In particular, the modest project (that clearly is a prerequisite for the ambitious one) does not include the claim that explanatory considerations are as a matter of fact guiding our inferences in an instance of inductive reasoning in which the IBE model would, if followed, yield the same conclusion.

The default position among the friends of IBE seems to be to argue for the more ambitious claim. Indeed, one might even think that the modest claim all by itself cannot be of serious interest. But why so? Clarifying the slogan and explaining how inference to the best explanation *might* work as an underlying mechanism of the inferential black box is by no means a trivial task. Equally well, the descriptive adequacy of this model can be compared with various other formal models of induction, such as the hypothetico-deductive model, Bayesianism, et cetera. All this surely contributes to our understanding of ampliative reasoning. Nevertheless, the realist explanationists invariably argue for (or just presume) the ambitious descriptive claim in addition to arguing for the justificatory claim that the inferences thus construed are reliable in taking us towards the truth. I am personally broadly sympathetic to the (non-foundational) ambitious descriptive thesis—as expounded in Lipton (2004, esp. ch. 8), for example—but I do not wish to defend that opinion here. Rather, I want to move on to consider the question of justification, premised on the legitimacy of the descriptive thesis. First, though, we should revisit the challenge posed by van Fraassen.

1.2.2 Responding to the challenges

Lipton makes a couple of important clarifications of the above slogan: (1) The ‘best explanation’ should be read as ‘the best of *competing* explanations’, for if the hypotheses are compatible, inferring one does not preclude inferring others. (2) Inference to the best should be sanctioned only in the case that the best is *good enough*. The latter restriction is particularly important to spell out in detail, for it is the key to defusing the argument of the bad lot.

The explanations we come up with and compare in applying IBE typically form a very small subset of all the possible explanations consistent with the data. Given any old set of candidates, it is quite clear that the rule of inferring the best will spell trouble *unless* we can ensure that the (approximate) truth lies within the set. Hence the qualification: the best of the set must be good enough for the rule to be applicable. But can we ever ensure that this condition holds? To answer this we must consider the process by which the candidates are generated, and integrate that process into the IBE framework. That is, inferring the best explanation does not only refer to the selection from some set, but also to the generation of that set, and for both steps of this *two-step* inference procedure there is a role for explanatory considerations to play. (Lipton, 2004: 148–151) The explanatory considerations functioning in the first step ensure the quality of the candidates; the procedure is fallible, of course, but if it is reliable then the best is (by and large) going to be good enough.

The explanatory considerations come into play in the generation of candidate explanations (the ‘context of discovery’) through background theories: the initial candidates are generated to cohere with the background theories which are, in turn, the results of an earlier round of IBE and hence carry a mark of earlier explanatory considerations. Assuming the approximate truth of background theories we can begin to envisage how by virtue of this coherence constraint we could be able to generate a good enough set of candidates, choose the best to yield an approximately true theory, only to function as a piece of approximately true background for the next round... This picture is, of course, extremely rudimentary, and there is much filling in to be done on all counts. But there is nothing wrong with the kind of iterative depiction of IBE that gets off the ground by referring to the background. One may find

the unexplained procedure of generating quality candidates on the basis of background knowledge both surprising and unexplainable. Perhaps it is so, but maybe we just have to accept that we have a knack of coming up with quality candidates. (Lipton, 2004: 151) However surprising, it seems to be on a par with the kind of knack the realist wants to endow us with in any case.

So van Fraassen's admonitions about the status of IBE regarding the feasibility of explanationism are exaggerated in view of a more sophisticated model of IBE. The crude two-step model merely serves to indicate the consistency of the descriptive model of explanationism in the face of the bad lot criticism, and the further justificatory project of showing that the model is a good description of the actual science amounts to a full-blown realist argument itself.

Let us now consider the more specific incoherency accusation, the Dutch-book 'proof' from Bayesian probability theory. We can respond to this by pointing out that van Fraassen employs an unnecessarily stringent and unnatural model of IBE in his argument. (Okasha 2002, Lipton 2004) Without getting into much detail, van Fraassen's argument turns on the assumption that modelling IBE in the Bayesian framework requires *modifying* Bayes's rule for updating one's degrees of belief for a hypothesis, to accommodate the explanatory 'bonus points' the hypothesis receives from its explanatory virtues. The gist of the response to this is that there does not seem to be any reason why the explanatory considerations cannot be reflected in the prior probabilities and likelihoods needed to apply Bayes's theorem, *without* modifying the theorem itself. Rather than being incompatible, perhaps Bayesianism and IBE 'should be friends' (Lipton, 2004: ch. 7) as these approaches can really complement each other.

* * *

The above responses to van Fraassen's challenges partially depend on allowing the background theories to play a significant role in evaluating the loveliness of an explanation. This background also includes conceptions about explanatory loveliness itself and this renders IBE highly *contextual*: what counts as the best explanation depends on where and when the evaluation is done. Whilst the background aspect can save the coherency of descriptive

IBE from van Fraassen's criticism, I will argue that the kind of contextuality it allows works in the end against the justificatory project (§1.6 and §3.3.2). Day & Kincaid (1994) have emphasised the role of contextual factors in IBE in general, and also in regard to the scientific realism debates, which they claim 'are often at cross-purposes and flawed because they ignore the contextual factors and substantive assumptions that are essential to IBE.' (289)

Debates over realism can occur in a variety of contexts, contexts which differ substantially in what is taken as background information, what IBE is supposed to establish, and what kind of evidence is allowed. At one extreme, arguments for realism might aim to convince the traditional philosophical skeptic. . . . At the other extreme, arguments over realism can be between individuals who are willing to grant a great deal of background information—as, for example, would be the case in debates between Millikan and other early-twentieth-century physicists over the existence of electrons. (*ibid*: 290)

Whilst they argue that the realist appeal to IBE is without force at the level of the sceptic who denies the kind of ampliative reasoning employed in the first place, at the level of scientific practice the use of IBE is nevertheless fully legitimate and powerful method of making decisions. But Day & Kincaid continue by claiming that the legitimacy of IBE arguments at the scientific level does not 'provide as much support as the realists pretend.' This is because the realist arguments 'must go beyond the narrow scientific context to be very interesting.'

I am sympathetic to these worries about the explanationist arguments for realism being cast in alarmingly general terms, but the issue needs clarification. Just how 'global' a context must an explanationist argument address to be interesting? I will return to this issue in the chapter after next (§3.5). The recommendation at hand, however, simply seems to be to elevate the explanationist argument for realism to a higher, meta-scientific level by explaining some empirical facts about science by realism itself.

What then would interesting IBE arguments for and against realism look like? They would have to be compelling in the 'philosophical' context—in the context (a) where evidence from historical and social

studies of science is allowed in addition to ordinary scientific evidence, but which (b) allows assumptions that skeptics would deny. (*ibid.*)

If realism is the end-result of IBE of this *philosophical* kind—so that realism itself is the best explanation of some fact about science—then what exactly should the *explanandum* be? Let us now turn to examine possible answers to this question.

1.3 Forms of global explanationism

Day & Kincaid, having argued that IBE is only an *abstract schema* over highly context-dependent inferences, actually want to voice a caution against excessive globalism:

Convincing arguments for realism (and anti-realism) are likely to be *piecemeal*. ‘Current science’ is an enormous batch of claims, assumptions, and evidence of great diversity. Why should there be any one IBE that shows that all science should be taken realistically? ... For, in some domains or pieces thereof, sociological explanations may be empirically reasonable. In other domains, we might be able to rule out anti-realist explanations and to build convincing explanations of science by postulating its approximate truth. (1994: 292)

I am also sympathetic to this graduation of the realist commitment, and ‘going piecemeal’ in one way or another with respect to realist commitments will be a theme to be deliberated on recurrently in this thesis. Typically, however, the explanationist-realist argument is a fully global one which is taken to run over all of mature science. But at the level of detail there is a multitude of philosophical positions which occupy the logical space spanned by the two dimensions of (1) what realist position is inferred as part of the best explanans to (2) what explanandum. The purpose of this (lengthy) section is to make the best possible case for a global explanationist argument and thus evaluate its merits. Not even the fittest of these arguments will survive the scrutiny. In the next chapter I will project my hopes (and fears) onto the alternative local realist strategy.

The discussion below is organised (in a slightly non-linear fashion) around the following issues: What is the right explanandum? How good are the rival non-realist explanations? Is the realist explanation good enough? Does NMA commit a probabilistic fallacy? (§1.4) Is the argument (viciously) circular? Is the argument of the right form to be credible? (§1.5)

1.3.1 What is the explanandum?

Beginning with the explanandum question, there are more than half-a-dozen alternatives that have been entertained in connection with the No-Miracles Argument. From the literature one can extract the *universal NMA template*

“The best (and/or only) explanation for X is realism (suitably construed). *Ergo*, realism (suitably construed).”

as a function $f(X)$ with the following values for the variable X

- success of science
- success of the scientific method (E_m)
- success of the abductive method of science
- success of a (particular) theory (E_t)
- scientific progress
- existence of novel successes
- particular novel success(es)
- diachronic success(es)

Let’s focus here on the following two major alternatives, both very prominent in the literature. The first (E_t) takes the explanandum to be the success of a (particular) theory, whilst the second (E_m) focuses on the success of the scientific *methodology* on the whole. What could count in favour of taking either of these two alternatives?

There is a popular argument for taking the success of particular theories to be the sensible choice of explanandum for the realist. Arguably this follows from the explanatory insufficiency of a rival non-realist explanation of scientific success, namely the ‘evolutionary’ selectionist explanation advanced by van Fraassen as a response to the miracles argument. If one wonders why

science, the set of theories we currently possess, is so successful in making predictions and manipulating the world to our liking, the answer to ‘the scientific, Darwinian mind’ is simply to be found in the strict selection criteria we impose on theory proposals as part of the scientific method: it is no wonder that successful ones are the ones we have since we only select the successful ones, the argument goes. The realists have been quick to point out that this simple, almost tautologous, explanation cannot possibly have anything to say about what is allegedly the *real* miracle: why *a particular* theory we have selected is successful. (Leplin 1997, Musgrave 1988, Lipton 2004) Leplin, for example, asserts that

... to explain why the theories *that we select* are successful, it is appropriate to cite the stringency of our criteria for selection. But to explain why *particular theories*, those we happen to select, are successful, we must cite properties of *them* that have enabled them to satisfy our criteria. (Leplin, 1997: 9)

Thus, according to this line of thought there is an underlying property of a theory, namely, being true, which allows a deeper explanation of its success, and one which is furthermore not incompatible with the selectionist explanation. The difference between the two is sometimes likened to that between phenotypical and genotypical explanations of selected traits in evolutionary biology. Lipton puts a further, diachronic, spin on this argument by stressing that ‘the real miracle is that theories we judge to be well supported go on to make successful predictions’ (2004: 194). This would correspond to something like the ‘miracle’ that Wimbledon finalists often make it to the finals of the Australian Open, too, which can be (partially) explained by the physical dispositions and genetic endowment, but not by the initial selection mechanism (i.e. qualifying rounds) at Wimbledon.

But the obvious disanalogy between the (dispositional-causal) genotypical explanation and the realist’s (logical) truth explanation should give one pause, and perhaps lead to question the sense in which the latter is genuinely explanatory at all of the explanandum E_t . For whereas the genotypical explanation for Pete Sampras’s exceptional serve can be understood in terms of his *capacity* to develop/maintain certain phenotypical attributes under

certain conditions (psychological situation, training history, etc.), the truth explanation cannot be so understood. All explanatory value there is to the latter lies in the logical fact that valid arguments with true premises have only true conclusions. But since both unsound valid arguments and invalid arguments with true premises *can* have true conclusions as well, the truth of a theory cannot be a necessary ingredient in the explanation of its success. So clearly truth of a theory is not a necessary precondition for it being successful, nor is it a sufficient one (just consider all the boring, trivial theories which do not stick their neck out in way of any prediction).

It might be objected that the demand for necessary and sufficient conditions for explanation is too strong, that it is adequate to consider the *propensity* for producing predictive (novel) successes, given a theory's truth value. Is it not just obvious that a false theory is much *less likely* to yield a predictive (novel) success than a true one? Here we hit the muddy waters of trying to evaluate the 'intuitively obvious' probability space spanned by the logical structures of theorising—a topic we shall return to in the next section (§1.4)—but let us consider the alleged analogy between genotypical and truth explanation further.

This analogy is illuminating in another way, and construed in the following terms it speaks for the alternative explanandum E_m . To fully explain the tennis playing abilities of *The King of Swing* (a.k.a. Pete Sampras) we need to appeal not only to his genetic endowment but to various historical contingencies having to do with his training, et cetera.⁵ Similarly, the truth value of a theory can play a role in an explanation of its success, but to really explain the success, the analogy suggests, we need to provide something equivalent to the story of how the genetic disposition to become the King of Swing has been *fulfilled* by a suitable training regime, nutrition and so forth. A more complete and plausible explanation of a theory's success—of a theory *becoming* successful—would hence refer to the historical trajectory leading to the theory, and to the scientists using the theory to derive successful predictions, in addition to its probable truth. We might refer to the

⁵Of course, purely genetic explanation can furnish a contrastive explanation, for example 'How can *The King* serve consistently 10% faster than *Agassi*?'. The answer 'Well, due to his genetic constitution he's about 10% taller, giving him about that much more leverage', can be both true and very informative, but clearly only assuming that a great variety of other possible contributors are equal for the two players.

role of (true) background theories in theory production, and to the success-producing capacity of the relevant scientific ampliative method.⁶ But now we have switched to the competing explanandum E_m : to explain the success of a particular theory in these terms leaves the further, equally ‘miraculous’ explanandum, about the scientific methodology. Why is it that the scientific methodology is success producing at all?

Perhaps those who advocate the explanandum E_t think that their professed truth explanation *supplements* the selectionist explanation so as to complete the explanation in the way we have demanded. They might think that the appeal to selectionist criteria explains why we have successful theories to begin with, and the truth of a successful theory explains why *it* is successful, so together these two assumptions fully explain why a particular theory is successful. That is, the selectionist criterion really tracks the truth, not just success. But how exactly does the selectionist criterion work and what does it work with? It seems that the selectionist explanation, in all its picturesque simplicity, misrepresents the scientific method. To return to the tennis player analogy, to explain why the Wimbledon finalists are all very good players by citing the selection criteria *takes it for granted* that there are a great number of competing players (of varying abilities) queuing for the tournament, from which to select. The analogous image of science as a massive field of competing hypotheses from which the best ones are selected, possibly on explanatory grounds, is a false image. And it is an image which does not have to be part of the sophisticated No-Miracles intuition. It is this fundamental point, I believe, that really motivates some philosophers choosing the alternative explanandum E_m . As Stathis Psillos, following Boyd (1981, 1990), forthrightly asserts: ‘the explanandum of NMA is a general feature of scientific methodology—its reliability for yielding correct predictions.’ (Psillos 1999: 79)

For Boyd and Psillos this choice of explanandum is affiliated with the analysis of the cumulative nature of the scientific method:

⁶The logical explanation E_t gives the impression that predictions from theories ‘drop out’ as a matter of logical deduction. For the actual theories of real science this is, of course, far from the truth, although theories are sometimes represented as such axiomatic-deductive systems *after* they have become successful. The role of explanatory considerations in bridging the gap between theory and phenomena is analysed in chapter §9.

That the methods by which scientists derive and test theoretical predictions are theory-laden is undisputed. . . . In essence, scientific methodology is almost linearly dependent on accepted background theories: it is these theories that make scientists adopt, advance or modify their methods of interaction with the world and the procedures they use in order to make measurements and test theories. . . . These theory-laden methods lead to correct predictions and experimental success. . . . How are we to explain this? (Psillos, 1999: 78)

But perhaps stressing theory-ladenness *per se* is not the best way of capturing the point here. The emphasis should rather be on the basic explanationist idea that the theory-laden methodology works by first assuming the truth of the background theories, and *then* using the two-step ampliative reasoning to produce a *small set* of alternatives from which to select the loveliest one. Why is success to be found consistently in that small set? Why is the abductive scientific method successful?

For some reason most commentators have *not* opted for this methodology-centred explanandum. Take Lipton, for example, who has also advocated the theory-centred E_t as the right explanandum for global explanationism. This is somewhat surprising, given the details of his description of scientific IBE. Recall, in particular, how Lipton consistently emphasises the two-step character of this inferential practice: the best explanation is inferred only if the best is good enough and the initial generation yields only truly plausible potential explanations. As a matter of fact, often the initial sieve lets only one explanation through, and explanatory considerations, according to Lipton, play a big role already in the initial generation of the candidate(s). If this initial step of thus described abductive methodology is not already success-conducive, then the selection process described as the two-step strategy is not going to yield predictive successes (except as a rare lucky accident). Scientific method as a matter of fact is successful, however, and this cannot be explained by appealing to the probable truth of particular successful theories together with a selectionist story along the lines of van Fraassen. This is not the image of science projected by Lipton's depiction of explanatory virtues. Rather, the question to be posed—provided that the picture of science painted by Lipton is by and large descriptively faithful—is how this explanationist methodology can be so successful. And the answer the realist should attempt to provide as

the best explanation simply refers directly to the method itself: the abductive method of science is truth-tropic.

Focusing on the explanandum E_t eventually leads Lipton to a worry he has about the No-Miracles Argument that he deems fatal to it. It will be instructive to go through this reasoning as it provides a fresh perspective on the argument above for the supremacy of the alternative explanandum E_m . It also leads us to consider another important rival non-realist explanation for the success of science.

1.3.2 Is the explanation good enough?

Recall that for an explanation to be worthy of inference through IBE it needs to be the loveliest of a bunch of good enough candidates. Now, assuming that there are empirically equivalent incompatible theories (i.e. assuming the validity of the underdetermination problem at the first-order level), the worry is that the hypothetical truth of each of these first-order theories functions equally well as part of the explanans for the success of the theory. This, according to Lipton, makes the truth of a particular theory a very *unlovely* explanation of its success:

How lovely, then, is the truth explanation? Alas, there is a good reason for saying that it is not lovely at all. The problem is that it is too easy. For any set of observational successes, there are many incompatible theories that would have had them. This is our old friend, underdetermination. The trouble now is that the truth explanation would apply equally well to any of these theories. In each case, the theory's truth would explain its observational success, and all the explanations are equally lovely. (2004: 195)

Lipton concedes that at the first-order level the underdetermination problem is not insurmountable, for we *can have* preferences for some of the underdetermined explanations due to their different degrees of explanatory loveliness (if one buys into IBE in the first place). But the realist, Lipton insists, is not entitled to appeal to this difference at the first-order level because his argument is a *purely higher-order IBE*:

But the proponent of the miracle argument, as I have construed her position, insists that the truth explanation, applied to a particular theory, is distinct from scientific explanations that the theory provides. She is entitled to this, if she wants it. . . . But the price she pays for this separation is an exceptionally weak explanation, that does not itself show why one theory is more likely than another with the same observed consequences. (2004: 196)

The realist, according to Lipton, is whoever wants to justify the reliability of IBE as a truth-tropic inference. The underdetermination problem is fatal for this project, according to Lipton, since the reliability of IBE (employed to make the choices at the first-order level) is ‘undecided’ prior to justifying it through the (second-order) realist argument, so the realist cannot appeal to it without running desperately in circles. But again, this circularity seems to affect the realist argument only if its explanandum is taken to be the success of a single particular theory. But I have alleged that the realist argument should try to explain the success of the scientific methodology as a whole, and for this second-order explanation the first-order employment of IBE cannot be thus separated and ignored, for those first-order instances of IBE form the very explanandum!

Lipton’s underdetermination criticism of NMA is very close to a popular non-realist explanation of the success of a theory in terms of its *empirical adequacy*. (e.g. Fine 1986, 1991; Kukla, 1998) This is what the empirically equivalent theories have in common, and it is hard to deny the intuition that the explanatory advantages of truth over empirical adequacy are quite minimal (if not wholly nonexistent) relative to the risk taken in proposing the explanation.⁷ Both truth and empirical adequacy go beyond the observed and similarly entail deductively the predictions that constitute the theory’s success. Assuming the possibility of empirical underdetermination and the fallibility of the realist inference, it does not furthermore make sense (*pace* Musgrave, 1988) to wonder what explains the empirical adequacy of a theory, if not its truth. There simply are theories which are empirically adequate but not true.

⁷See Leplin (1997) for an attempt to defend truth over empirical adequacy as an explanation of a theory’s success, and Kukla & Walmsley (2004) for a successful rebuttal.

But if we concede that empirical adequacy of a theory explains its success just as well as the ‘super-erogatory’ truth explanation, do we not have an analogous less expensive non-realist explanation for the success of scientific methodology? That is, why not explain the latter by attributing to science the *capacity* of producing empirically adequate theories? According to this explanation the ampliative methodology of science is not truth-tropic, but empirical-adequacy-tropic, or just ‘instrumentally reliable’ as Fine (1991) puts it.⁸

This rival non-realist explanation is troubling and there is no way to squeeze it out of the logical space of possible explanations for E_m . But there is a significant disanalogy between this case and the apparently similar competition in the case of E_t . The point I am about to make here hangs again on the assumed descriptive adequacy of the two-step IBE model; that is, we need to assume that (a) scientific methodology is of a piece (in a suitable sense), and (b) ampliative inferences exemplifying this methodology are based on background theories which are the results of previous inferences of the same form. In light of these premises, we can see that the competing explanations here are not that scientific methodology is capable of producing (i) a true theory T_0 , versus (ii) an empirically adequate theory T_0 . Rather, the competing explanations are: scientific method is (1) truth-tropic, versus (2) capable of producing a consistent theoretical story of the world in which

⁸This is a charitable reading of Fine (1991: 83). In particular, Fine talks about instrumental reliability of background theories, not of methodology, despite first correctly identifying E_m as the explanandum of the realist argument.

What explanatory success warrants is belief in the instrumental reliability of the explanatory story. This is an explanation of outcomes by reference to inputs that have the capacity (or ‘power’) to produce such outcomes. (Fine, 1991: 83)

Psillos wonders whether there is anything to such dispositional explanation:

Is it a brute fact of nature that theories—being paradigmatic human constructions—have the disposition to be instrumentally reliable? This hardly seems credible. (Psillos, 1999: 93)

Whilst I share with Psillos this intuition about the credibility of the instrumentalist explanation, I think it is important to provide reasons for not taking it as a serious candidate explanation, but merely as a blunt assertion. (My reasons are given in the main text.) By itself the demand for an explanation of how the disposition is grounded is question begging, since the whole point of the instrumentalist manoeuvre is to pull back from the specifics of the realist explanation to something more deflationary, and logically speaking instrumental reliability of the methodology does just that.

empirically adequate theories are derived from false background assumptions, by following ampliative reasoning which remains constant in form.

Although (2) really does explain the explanandum E_m in question, it is clearly not anymore the case that this explanation is somehow entailed, or included in the realist explanation (1). Hence the anti-realist advocating (2) cannot really accuse the competition of being super-erogatory—just doing the same explanation with higher stakes. Rather, here we have two independent and incompatible explanations which are to be weighed on the merits of their plausibility. The obvious challenge for *both* sides of this micro-debate is to try to make sense of what these merits are and how the weighing is to be accomplished. The disappointing current state of play regarding this challenge will be discussed below (§1.5).

But for now, let us conclude with the following note. As far as the *prima facie* intelligibility of the scientific success goes, it seems that the realist explanation (1) gets the intuitive upper hand in this debate. For consider the additional mystery that is left behind by the rival non-realist explanation: What is it that makes the unified scientific methodology such that it can consistently build on falsehoods in a way that provides us instrumental reliability? The intuition behind this mystery should be clear: if one tells false stories about the mechanisms behind some phenomena, say, and *deploys* these falsehoods as part of the background to motivate a (small) set of further stories about some other phenomena, it seems *reasonable* to expect that the falsehoods used as a background *tend to mislead* the subsequent stories. This intuition is to be contrasted with the expectation we would have of theoretical truths to yield further truths, assuming truth-conducive character of scientific methodology. What could it be about the scientific method understood in this way that would thus *compensate* the intuitively debilitating effects of background falsehoods?⁹ Furthermore, although there are innumerable ways to argue from one set of empirically adequate theories to another set, for each theoretical circumstance, it is not at all clear that there are many such ways which keep the reasoning followed relatively constant throughout the science. But this is what the explanationist claims to find in the actual science: a

⁹At this point one might make the valid point that it is indeed hard to understand the *capacity* of the scientific method of producing empirically adequate total science, if not through it being truth-tropic.

discernible ampliative two-step methodology arguably largely based on IBE.

* * *

Assuming that we can on these intuitive grounds take the realist explanation to be better than its rivals and good enough an explanation, and assuming that we can trust IBE as a generally valid framework for making inferences in both sciences as well as in philosophy, then realism is doing well. Unfortunately, it seems that the best case we can make for the latter premise is a particularly weak one, due to the notorious circularity objection, as well as to an objection of my own device explored below (§1.5). But before examining these arguments, I want to issue yet another piece of criticism of the theory-centred construal of the realist's global explanationist argument. This serves to undermine a recently surfaced pessimistic line of thought about not only the No-Miracles Argument, but the whole realism debate. Also, it useful to delve deeper into this issue here as the crucial point at stake will recur at a later stage, in connection with our discussion of the proper understanding of the Pessimistic Meta-Induction (§4.3).

1.4 Is NMA just a base-rate fallacy?

I have argued that focusing on the theory-centred explanandum E_t leads the realist astray. Here is another unpleasant conclusion it can lead to: the whole realism debate is *ennui*, a wholly misled dialectic, motivated only by simple probabilistic fallacies we are disposed to commit as human reasoners. This is the pessimistic conclusion Magnus & Callender (2004) argue for, drawing on Lipton (2004) and Howson (2000).

With each argument [PMI and NMA] we are tricked by a base rate fallacy. If this is correct and the intuitions marshalled by [the] argument are phantoms of that fallacy, then there is much sound and fury in debates over realism that signifies nothing. (Magnus & Callender, 2004: 322)

It was stressed above that the connection between success and truth hypothesised by the explanationist is not one of entailment, in either direc-

tion.¹⁰ Surely there is no *necessary* connection between success and truth, for a true theory can be trivial and boring without any predictive success, and successes *can* be brought about by false theories, however unlikely it might feel. As a consequence of this fact the informal statements of realism typically include vague probabilistic qualifiers, such as ‘probably’, ‘typically’, ‘tends to’, ‘mostly’, etc., as in the paradigmatic expression ‘mature scientific theories with novel predictive success are typically approximately true’. It is this irreducible probabilistic character of realist statements that has led some to read the whole doctrine in statistical terms in a way that smacks of ultimately nonsensical twiddling of probabilities with undefined and indefinable base rates. This is how Magnus & Callender (2004) frame the realist argument in probabilistic terms:

Let \mathcal{H} be the set of present candidate theories. Now the no-miracles argument takes this form for all x :

$$[1] \quad Pr(Sx|x \in \mathcal{H}) \gg 0$$

$$[2] \quad Pr(Sx|Tx \ \& \ x \in \mathcal{H}) \gg 0$$

$$[3] \quad Pr(Sx|\neg Tx \ \& \ x \in \mathcal{H}) \ll 1$$

$$[4] \quad Pr(Tx|Sx \ \& \ x \in \mathcal{H}) \gg 0$$

The argument revised in this way is still valid, but its soundness should tug less at our intuitions. Premise [1] will hold only if any arbitrary member of the population is likely to be successful. On the assumption that success is a reliable indicator of truth, this is tantamount to assuming that any arbitrary member of the population is likely to be true. If $Pr(Tx|x \in \mathcal{H})$ is low (and how can we know if it is not?),

¹⁰This may seem obvious to some but it is certainly not fully acknowledged by all the commentators, on either side. For example, some realists seem to think that a single theory which is both successful and (plain) false works as a decisive ‘counter example’ to realism, and hence must be *necessarily* dealt with.

then [1] fails and the conclusion does not follow. (Magnus & Callender, 2004: 325)

In this model of NMA the explanandum is (again) taken to be the success of a particular theory, and success is construed as *a symptom* of truth. This allows a simple probabilistic gloss on the argument. Assuming that a true theory is likely to be successful [2] and that a false theory is unlikely to be successful [3], it follows that a randomly picked successful theory is likely to be true if there are not too many unsuccessful theories loitering around [1]. Represented in this way, the ‘miracle’ intuition is fully contained in [2] and [3], and the argument is reduced to the mere hope that the actual state of affairs is such that the ‘hidden premise’ [1] holds. But there is no way of knowing this, Magnus & Callender insist:

We might attempt to assess [1] by inspecting the pool of theories, \mathcal{H} . We defined \mathcal{H} as the set of candidate theories, but what theories were candidates for our present mature sciences? It is impossible to count up or even fairly sample all the theories that were considered for our mature sciences, and so it is impossible to evaluate whether [1] obtains. (*op.cit.*)

The explanationist can fault this construal of NMA on several counts. To begin with, surely ‘counting up’ the candidates is not the most appropriate idiom to use here; it is enough that we can *qualitatively* evaluate the best explanation for the success of science. Also, the use of the word ‘candidate’ is somewhat ambiguous. In view of the two-step generation-selection process of IBE presented earlier, it could be interpreted as ‘candidate for the first step’—whether ever actually thought of or not—or as ‘actual candidate in the set from which the best one is chosen’, i.e. a candidate which has already passed the initial sieve. I think the text is most naturally read in the latter way; i.e. the candidate theories are those actually entertained by the scientists. But thus interpreted we can question the leading assumption that we have *no idea* of the proportion of hypotheses entertained per each successful theory.

The point is that it seems to be implicit in the global explanationist argument that, according to our best understanding of how science works, the best explanation of scientific success is *not* that *either* a successful theory

is true, *or* it is a member of a large set of ‘attempts’ the mere cardinality of which makes it likely for one or another of its members to succeed (due to a small probability for each member).¹¹ Indeed, I earlier opposed the selectionist picture of scientific method on exactly these grounds: it should be part of the explanationist argument that we have analysed the generation of a typical scientific success to an extent that falsifies the selectionist picture of science. According to the explanationist we do have some qualitative idea of the proportion of actual theories entertained per a typical success. But this proportion does not have to be as strict as the premise [1] above portrays. The explanationist defence of realism does not hang on the assumption that *most* of the theories presently entertained are successful. All we need is that $Pr(Sx|x \in \mathcal{H})$ —in so far as we can make any sense of it—is not as low as to undermine the explanationist description of science. I take it that the recommended (sophisticated, non-dogmatic) No-Miracles Argument is continuous with our best understanding of science. There is thus more to NMA than the simple probabilistic gloss captures.¹²

Magnus & Callender draw explicitly on Lipton (2004) and Howson (2000), both of whom also see the No-Miracles Argument as committing the base-rate fallacy. But actually their understanding of the fallacy is quite different from the reasoning offered by Magnus & Callender, corresponding to the other interpretation of the word ‘candidate’ above. Lipton, for example, cites the fallacy in connection with his underdetermination problem for the loveliness of the miracle explanation.

The intuition behind the miracle argument is that it would be a miracle if a highly successful theory were false; but once we take these underdetermined competitors seriously, a miracle no longer seems required for our successful theory to be false. Quite the opposite: it would

¹¹Magnus & Callender actually see their formulation of NMA as distinct from the typical IBE expression, but I have no idea how else one can get the initial probabilities if not through some explanatory inference. Perhaps their argument could be viewed as belonging to the context of the realist image, rather than of justification, although the authors do not imply that kind of distinction either. Anyway, my criticism stands: there is no need to build one’s realist image on the naïve probabilistic representation of the No-Miracles intuition. (cf. also §4.3.2)

¹²Of course, the realist really should be non-dogmatic and allow the possibility of alternative explanations of science. Indeed, perhaps the best explanation of success gets fragmented in the way envisioned by Day & Kincaid. (cf. §1.2)

rather be a miracle if the truth did not lie instead somewhere among the innumerable underdetermined competitors. And here it is difficult not to suspect that the original plausibility of the miracle argument is just an instance of philosophers falling for the ubiquitous fallacy of ignoring the base rates... (Lipton, 2004: 196)

I have already argued that *this* argument needs to be responded to by performing an acute gestalt switch with respect to what is taken to be the ‘miracle’ explanandum. The success-tropic character of the ampliative scientific method is what we are trying to explain, and the realist appealing to global explanationism takes the truth-tropic character of this method to be the best explanation available. This explanandum cannot be divorced from the second-order explanation to create alternative equally lovely incompatible explanations because, to put it bluntly, there is only one actual scientific method and hence no underdetermination.

The reasoning presented by Magnus & Callender, on the other hand, makes no mention of possible empirically equivalent rivals. Rather, they take the base rate to be ambiguous due to all the *actual* theories that have been entertained on the way to the predictively successful one that arguably grounds realism. *This* arguments needs to be responded to by simply denying that we do not have any qualitative idea of the proportion of (actual) theories that have been entertained along the way. Here, too, a gestalt switch would be in order. Focusing on our best understanding of the ampliative scientific method is exactly what the sensible global realist does; it is not just the intuitive probabilities $Pr(Sx|Tx \ \& \ x \in \mathcal{H}) \gg 0$ and $Pr(Sx|\neg Tx \ \& \ x \in \mathcal{H}) \ll 1$ that drive the argument in so much as the way *the success has been arrived at*. In particular, it seems that the selectionist image of science does not sit comfortably with the explanationist image painted by Lipton, Boyd and others. Science does not seem to advance by way of generating arbitrary candidates from which the successful ones are then picked. It simply does not make sense to model a piece of successful scientific pursuit—the quantum mechanical prediction of the magnetic moment of an electron, say—as that of ‘guessing correctly each of nine digits after the decimal point in the magnetic moment of the electron.’(Howson, 2000: 37) Proceeding in this way is bound to spell all manner of pseudo-problems, as manifested in Howson’s critique of NMA.

The statistical nature of the realist thesis is to be understood in terms of the reliability of the scientific method, rather than statistical reliability of the success of a *particular* theory as an indicator of the truth of that theory. Understanding of the success–truth connection at the level of individual theories is based on understanding the scientific methodology as being truth-tropic. Because this method is fallible we can have successful-yet-false theories without immediately undercutting realism. To spell out exactly how reliable the ampliative method is, is a very difficult challenge to which I am not planning to contribute here. And despite escaping the fallacy accusations, global explanationism faces problems elsewhere, as will be shown in the next section.

To bring this section to a close I want to briefly consider the positive lesson Magnus & Callender draw from their analysis of the realist ennui. By distinguishing overarching *wholesale* arguments from local *retail* arguments they attempt to drive a wedge between unprofitable global debates—only powered by both sides desperately clinging onto their fallacious intuitions—and profitable local debates which operate at the level of *particular* theories and resolve questions only about particular kinds or individuals.¹³

Wholesale realism debates persist not due to mere stubbornness, but because there is *no reason* for opponents to agree. The more modest reach of the narrower retail question allows for arguments that are non-statistical or for broad agreement in estimating base rates. These debates are profitable because there is reason to agree. (Magnus & Callender 2004: 336)

As it happens, I am highly sympathetic to the spirit of this conclusion. But I have a different (more traditional) set of reasons to believe that there is no reason for the opponents to agree in the debate about global explanationism—cf. Wylie (1986) and below—and I have a different understanding of what local realism amounts to. In my view Magnus & Callender have not provided any account of why there would be a reason to agree at the local ‘retail’ level and what that reason might be, and I believe providing such a philosophical

¹³The other half of their ennui claim concerns the pessimistic induction which, they claim (following Lewis, 2001), also commits the very same fallacy of ignoring base rates. I will consider this part of the argument in the chapter after next, in connection with analysing PMI itself.

account will unite the local arguments to an extent. I will return to this topic in chapter §3.

1.5 What really is wrong with global explanationism

The global explanationist argument is fundamentally circular—that much is agreed upon by virtually everyone. The question of interest is whether an argument of any considerable strength can be salvaged, and on what assumptions. I will now argue that only a very weak argument survives this long-established predicament in the foundations of global explanationism.

The section §1.3 attempted to delineate a sense in which realism (or the truth-conducive character of scientific methodology) is the best explanation of the success of science. But even if this much was fully established, we would still be far from completing the argument for realism. A case needs to be also made for taking inference to the best explanation to be a truth-tropic ampliative inference, not just in general or in some possible cases, but in this specific philosophical instance in particular. That is, a case needs to be made for taking the global realist argument—which in itself is an instance of IBE—to be an argument that delivers a true conclusion. We have already seen how some of the general criticism against explanationism can be responded to, so that there is no immediate reason to be suspicious of the explanationist dictum on the basis of IBE harbouring some fundamental inconsistency, for example. But what *positive* reason do we have for trusting the global explanationist? According to this position the only rationale for taking the loveliest potential explanations in science to be also (by and large) actual explanations is grounded on the miracle argument which is itself inferred as the loveliest explanation of the success of science. This is ‘*the realists’ Ultimate Petitio Principii*’, accuses Laudan, who states that ‘it is little short of remarkable that realists would imagine that their critics would find the argument compelling.’ (Laudan, 1981: 44)

But this is a monumental case of begging the question. The non-realist refuses to admit that a *scientific* theory can be warrantably judged to be true simply because it has some true consequences. Such non-realists are not likely to be impressed by the claim that a philosophical

theory such as realism can be warranted as true because it arguably has some true consequences. (*op.cit.*)

This venerable point against global explanationism still holds.¹⁴ The realist responses have varied: some have argued that when coupled with a naturalist-externalist epistemology the circularity, although undeniable, is not vicious (Boyd, Psillos), whilst some have tried to defend the argument alongside general inductive justification of induction (Lipton). The consensus seems to be that if there is any force in the second-order IBE of the miracle argument, it is felt only by those who already take abductive reasoning to be reliable.

Laudan himself also considers the possibility that ‘the realist is not out to convert the intransigent sceptic or the determined instrumentalist’—that the miracle argument is only one to preach to the converted with—but concludes that even if one believes the first-order theories to be true, the realist position cannot be considered to be confirmed as a scientific hypothesis on a par with our best theories. For clearly realism has not made (novel) predictions, for example, and clearly it cannot be subjected to the stringent empirical criteria which the realist himself insists on in the case of scientific theories. I think Laudan is onto something important here, although undoubtedly realism has never been put forward as a scientific hypothesis.

Where Laudan’s remark points to, and what really is a major weakness in trying to salvage a rule-circular global explanationist argument which would (justifiably) pump more faith into those who already believe, is this. The best explanation of the instrumental success of scientific methodology cannot be viewed (without a further argument) to be on a par with the first-order scientific explanations, *in the sense of* belonging to an objectively unified class of abductive arguments. As it stands, the second-order IBE attempts to gain credibility for its reliability by referring to first-order instances of IBE which are very different in kind: the second-order IBE is an instance of logical explanation, whilst the well-understood first-order instances are (arguably) mostly causal-contrastive explanations (in Lipton’s model of IBE, at least). The mere fact that all these inferences are both *ampliative* and *explanatory* (in some broad sense) is not enough to unify them as a group of inferences all of which should be reliable just because some arguably are. Lipton’s work

¹⁴The same lesson is repeated in Fine (1984) and van Fraassen (1985), for example.

on describing the scientific instances of IBE of the causal-contrastive form has yielded this much: we can agree that this kind of explanatory guidance does indeed play a big role in scientific inferences, and we can agree on the criteria of ‘loveliness’ employed, according to this description. But regarding the logical explanation of the kind that NMA manifests we do not have such ground work laid before us. How is the loveliness of logical explanation to be measured? Does this measure of loveliness correspond to (predictive?) success, as in the case of scientific inferences it (arguably) does? It seems that these questions do not yet have answers beyond mere intuitions. Indeed, it seems that the logical explanation of NMA really is idiosyncratic whilst the causal-contrastive explanation is ubiquitous.

Consider now the following defence of NMA in the framework of externalist epistemology.¹⁵ According to Psillos, successful instances of explanatory reasoning in science provide ‘the basis (and the initial rationale) for this more general abductive argument’ which attempts to ‘defend the thesis that IBE, or abduction (that is, a *type* of inferential method), is reliable’. Also, in his analysis, ‘given an externalist perspective, NMA does not have to assume *anything* about the reliability of IBE.’ (1999: 85) But this cannot be right. For clearly the characterisation of NMA above assumes that the first-order instances of IBE are *relevantly similar* to the second-order IBE of NMA, so that evidence for the reliability of IBE in the sciences doubles as evidence for the reliability of IBE *tout court*. Psillos seems to think that NMA thus establishes the reliability of IBE *simpliciter*, but surely some constraints need to be determined as to what counts as an *explanation* and what counts as a *lovely* explanation. In particular, in the case at hand we need to determine those constraints for the kind of explanatory reasoning that NMA exemplifies. Before this lacuna is filled, the argument amounts to very little over and above intuitions, even for the realist herself.¹⁶

¹⁵According to externalism one is justified using a reliable rule of inference regardless of whether one already knows or has reasons to believe that the rule is reliable. Hence one is justified using IBE as a rule of inference if it is reliable, before any positive reason is generated to believe that it really is reliable.

¹⁶It is true, of course, that if the first-order IBE is reliable, then the conclusion of NMA follows *in so far* as the loveliest (as the realist would have it) explanation of the success of scientific methodology turns out to be also the actual explanation. But it does *not* follow that IBE *simpliciter* is reliable. So even though this instance of logical IBE happens to yield the right conclusion, the argument itself does not give a good reason to believe it.

This last argument fired against global explanationism generalises to a worry about the project of justifying inference to the best explanation in general. I conclude this chapter by exploring this worry in some detail.

1.6 Justifying IBE

It was explained above (§1.2) how the challenge of making sense of IBE, as envisioned by Lipton whom I here follow, divides into two projects. It is one thing to describe the role of explanatory desire in our inferential practices, and this is a quite independent endeavour from justifying these practices as truth-conducive. It was also emphasised that IBE is a highly-generalised inferential template of two variables: an actual inference of this form is governed by what counts as explanation and what counts as loveliness of such explanation. These variables are context dependent—they are determined by ‘the background’.

The descriptive project aims to elucidate our inferential practices by providing a unified picture of it in terms of this IBE template. It is believed by the exponents of this project (the present author included) that the template covers much, albeit not all, of ampliative inference making. This unified description is afforded by the high-level generality of the template and the malleability of the two variables. But this generality which is a virtue vis-à-vis the descriptive project is in fact a bit of a liability regarding the project of justification, or so I will now argue.

A preliminary word of notice. The aims of the justificatory project are modest in that it does not pretend to be out to convert the outright Humean Sceptic. Rather, it is enough if IBE can be justified bar Hume’s problem, in the sense that all the challenges to IBE can be reduced in one way or another to general Humean scepticism. In particular, inductive justification of IBE must be allowed to be equally circular and problematic as inductive justification of enumerative induction, say. (cf. Lipton, 2004: ch. 9)

Intuitively, whatever belief we have in the reliability of some inductive method, it is based on the observed reliability of this method in the past. So presumably this is how to generate a (circular) justification of the reliability

of IBE as well.¹⁷ So the belief that ‘we live in the loveliest possible world’—the belief that our criteria for what counts as a lovely explanation in different circumstances are objectively such that they track the truth—is encouraged by all the instances of IBE (based on these criteria) that have been observably successful in guiding us. What is important to this intuition is that the criteria for loveliness stay the same, and that the relevant criteria are indeed employed at the observably confirmed inferences. This, it was argued above, is a major weakness for the second-order logical IBE of the miracle argument: the logical explanation of scientific success is not commensurable with the observable successful instances of IBE.

But the point generalises: we can see that ultimately the business of justifying IBE is completely done at the level of those constraints that determine what counts as a lovely explanation, and the mere fact that an inference appeals to explanatory reasoning has per se nothing to do with its reliability on the basis of the track record of previous explanatory inferences. So when it comes to estimating the reliability of some instance of IBE about unobservables, say, this is *not* done by acknowledging the inference being guided by explanatory considerations and thus belonging to a naturally unified class of inferences to the best explanation, many of which have been successful. Rather, what counts is the set of background beliefs that may unify a class of inferences to a greater or lesser extent. These background beliefs, wherever they come from, govern such wide-ranging explanatory virtues as unification and providing a causal mechanism which, on closer analysis, can encode some fundamental beliefs about the structure of the world.¹⁸ None of this deflates the descriptive value of the IBE project, of course. The point is simply that as far as the justification project is concerned, there is a danger of over-emphasising the template over the variables.

As far as realism is concerned, we can begin to see how the explanatory pursuit of science may not necessarily be the right focal point of its defense. Explanatory considerations undoubtedly guide us to find out about

¹⁷The miracle argument considered in the previous subsection is a more recherché attempt to justify IBE. The last problem identified with this approach crop up unsurprisingly with this more humdrum argument as well.

¹⁸In chapter §9 I will attempt to provide some descriptive unity to diverse explanatory virtues in terms of understanding.

the world—this is the message of the descriptive project of IBE—but to justify this knowledge we may have to turn to more basic elements of our analysis of ampliative reasoning. I will continue this theme in the next two chapters, in connection with assessing the numerous arguments of those realists—Lipton included—who refuse to elevate the explanationist strategy to the meta-level of the No-Miracles Argument.

* * *

We have spent a great deal of time looking at the prospects of having a general, overarching justification for the realist inference to truth-conducive character of the scientific method. The miracle arguments considered above aimed at justifying the realist’s appeal to abductive reasoning, once and for all. But although IBE is an attractive and internally coherent descriptive picture of much of inductive inference making, no acceptable non-circular argument for realism is forthcoming at the global level that simply builds on the observably successful instances of this form of inference.

I now want to turn to more *local* considerations, and focus on the nature of some particular kinds of ampliative inferences to the unobservable. Two local strategies to defend realism will be considered in the following two chapters. The first one (*Experimental arguments*) focuses on the practice of scientific experimentation and the nature of experimental instruments, whilst the second (*Ampliative Inferences undivided*) views our inferential practices as being in a significant sense of a piece. Both of these strategies yield data that *speak against* the selective scepticism that empiricist anti-realism amounts to. Although no knock-down argument against anti-realism is to be expected at the local level either, the dialectical force of these considerations is not vitiated by the kinds of problems examined towards the end of this chapter.

Overcoming the Empiricist Challenge: Experimental Arguments

“Before these pictures were taken no artist would have dared to draw a horse as a horse really is when in motion, even if it had been possible for the unaided eye to detect his real attitude. At first sight an artist will say of many of the positions that there is absolutely no ‘motion’ at all in them; yet after a little study the conventional idea gives way to truth, and every posture becomes instinct with a greater motive than the conventional figure of a trotting horse could possibly show.”

Scientific American, October 19th 1878

Epistemic anti-realism was defined in the Introduction as ‘no-knowledge-of-the-unobservable’, which in turn provides a definition of realism as its complement: realism, minimally construed, requires having some warranted knowledge of something (physical) unobservable. There is a tradition of arguing for realism thus defined by focusing on the experimental, rather than theoretical side of science. *Experimental arguments* look at how science is done in the laboratory, rather than on the pages of textbooks of theoretical accounts and explanations. The emphasis is on the acts of producing phenomena, rather

than on the acts of explaining them. These arguments are *local* in that they need to be developed in detail on case-by-case basis, rather than applying across all of successful science all at once.¹

The objective of this chapter is to provide an overview of the main experimental arguments in the literature, and to claim that these arguments indeed do furnish us with the beginnings of a promising route to a plausible, non-circular argument to a realist conclusion about (a significant class of) unobservables. Furthermore, I will insist that the strategy embodied in these arguments is *not* parasitic on IBE in the way that some advocates of the global approach have argued. But although very intuitive and compelling, the local experimental arguments mostly suffer from being enigmatic and underdeveloped. I will propose to analyse the pull of these intuitions in terms of a low-level juxtaposition of ampliative inferences which permeates the observable-unobservable barrier, in the next chapter (§3).

But let us begin by examining in detail the source of the anti-realist scepticism about theoretical knowledge—the *empiricist challenge*—which hinges on the possibility of there being many ways the unobservable world could be, compatible with our perceptual knowledge.

2.1 The Empiricist Challenge

There are many ways the world could be, for all we know. In particular, we can imagine many ways the world could be in the *future*, for all we know *now*, and we can imagine many ways the *unobservable* world could be, for all we know of the *observable* world. The empiricist challenge uses the latter fundamental, in-principle uncertainty, to challenge the idea that we have knowledge of the world unobservable to us.

The empiricist challenge relies on (i) the role of perceptual experience in our epistemology, and (ii) the thesis of evidential indistinguishability. Since we can imagine that there are many ways the world could be behind the veil of experience, how could we ever come to know how the world *really* is, given that perceptual experience is the source of all our scientific knowledge, and experience does not distinguish between empirically equivalent theories? This

¹The sense of locality at play is clarified and analysed further in §3.1

challenge is a very natural one, but it leaves two potential responses open for the realist.

First of all, it is not clear that it follows from the fact that experience is the source of all our scientific knowledge that we cannot have experiences (somehow) *mediated* by instruments in a way that would give us knowledge of the unobservable. That is, we can grant that there is some (vague) observable-unobservable distinction to be drawn, but deny that this distinction matches the boundary between those things we can gain knowledge of by experience and those we cannot. Having knowledge of even observable matters of fact requires an element of interpretation and conceptualisation—sense data does not constitute knowledge in itself—and it is not clear why suitable data accompanied by interpretation cannot yield knowledge of something unobservable.

Secondly, it is not clear that the thesis of evidential indistinguishability can be tailored for the anti-realist needs in a way that avoids general Humean scepticism about all knowledge achieved through inductive generalisation. Since experience affords *some* knowledge through ampliative reasoning to the empiricist and the realist alike, perhaps the way the former escapes the limits of Humean inductive scepticism also offers a way for the latter to escape the confines of observability. Or, in other words, perhaps the kind of *selective scepticism* about our inductive powers offered by the empiricist is an unstable (or less strongly, just unappealing) position to hold.

Both of these lines of response have been well developed in the realist literature. The first intuition leads to the so-called experimental argument for realism considered in this chapter, whilst the second is part and parcel of pretty much any other realist approach. After looking at both in this chapter and the next I will conclude that the latter is more fundamental: the strength of the first response also depends on the strength of the second. This then leaves open the question of how to best spell out the latter, and in answering this I end up sharing an assumption held by the experimental realists: the case for realism is to be made without direct appeal to explanatory power.

The first experimental argument examined below (in §2.2), mostly based on Hacking's renowned analysis of microscopy, serves as the initial motivation for the general line of thought developed here. This is further strength-

ened and generalised by going through Hacking's argument from engineering (§2.3), first, and then Cartwright's more general argument from the idiosyncratic character of causal explanation (§2.4). The experimental arguments culminate in the more recent formulation given by Achinstein (§2.5), which I aim to analyse further and build upon in the next chapter (§3).

But before we examine the local realist arguments in detail, I want to make four preliminary points, the last one of which deserves an Appendix (p. 58). The points to be raised concern: (a) the status and understanding of the observable-unobservable distinction (as defined by van Fraassen); (b) the idea that underlying the empiricist challenge there is a motivation for structural realism; (c) the modest aims of the argument developed in this chapter; (d) van Fraassen's constructive empiricism.

- (a) The empiricist argument springs from the plain fact that we can directly perceive (see, smell, taste etc.) many things, but most we cannot. We cannot perceive a molecule, or the core of the sun, or the event horizon of a black hole. In fact we cannot even perceive the wings of a flying hummingbird! The observable-unobservable distinction is admittedly an anthropomorphic, fuzzy, empirical distinction defined relative to an epistemic community. (van Fraassen, 1985: 253–8) But this does nothing to detract from its reality.² The real question here concerns the epistemological significance of this boundary.
- (b) We can imagine that the unobservable world is very different from the observable one, and the modern theories of physics are indeed populated by abstract physical concepts far removed from the observations on which the theories featuring those concepts are ultimately based. According to our best theories the nature of quantum particles, fields and spacetime is *unimaginably* different from anything we ever observe, and due to the existence of alternative evidentially equivalent theories (or interpretations of mathematical formalisms) the real nature of

²There has been some polemic about the possibility to draw the distinction within the confines of constructive empiricism, and considerable care needs to be taken in handling the concept of observability from empiricist principles. (cf. Ladyman 2000, 2004; Muller 2004, 2005)

these objects could be stranger still.³ So what grounds do we have for claiming that we can know what the quantum particles are like? This characteristic of some of our best science has undoubtedly motivated the contemporary empiricist (cf. McMullin 2003). The *structuralist* rises to this challenge by recommending realist commitment only to the structure, as opposed to the nature of the unobservable world. How the nature-structure distinction is meant to be cashed out is open to question, but the prima facie motivation for the position is clear. Structural realism is the topic of the second part of this thesis.

(c) The aim of my arguments in this chapter is modest. I do not wish to argue for a particular form of realism, but only against the anti-realist complement that defines realism (cf. *Introduction*). That is, the claim here is simply that given the force of the local realist arguments, it is quite preposterous to assert that we do not have *any* knowledge of the unobservable world. Hence, I do not argue for entity realism over and above theory realism, for example. Rather, the objective is to clarify the structure of the various local realist arguments, and to evaluate their relative strengths and weaknesses. Whereas the global miracles argument for realism looks at a theory ‘as a whole’ and declares it as approximately true, the various local arguments for realism aim at something significantly less: they only concern specific existence claims, or the reliability of particular scientific inferences, which are part of the theory. Exactly how far towards the full-blown commitments of theory realism one can build bottom-up from the local level is a question beyond this work.

(d) The empiricist challenge uses the fundamental, in-principle uncertainty about the world, to challenge the idea that we have knowledge of the things unobservable to us. There is a tricky notion at play here: *knowledge*. We can adopt a working definition of knowledge as *justified belief* (the problems of which have been well-known since Gettier (1963)) for

³‘Stranger’ here is not to be read as objective. Consider underdetermined interpretations of quantum mechanics, for example: whether Bohmian theory is stranger than the Everettian one depends on various metaphysical preferences. But both are far removed from everyday physical categories, for sure.

our purposes. But if ‘justified’ is taken to be something like ‘rationally taken to be true’, then we need to know what it is for us to be rational. And it is already at this very fundamental epistemological level that van Fraassen parts company with the realist (cf. Appendix, p. 58).

2.2 Extending our senses

Ian Hacking has adduced a set of wonderful intuitions speaking for realism about entities observed through a microscope. (1981, 1983: ch. 11) His deservedly well-known manifesto is three-pronged, consisting of the sub-arguments of (a) *coincidence*, (b) *intervention*, and (c) *the grid*.⁴

Hacking places the bulk of the polemical weight on (a): the idea that in microscopy several kinds of microscopes (Hacking lists thirteen), based on different physical processes and principles, all yield essentially the same ‘image’ after we have learned to properly use the microscope. It would be, the idea goes, a coincidence of cosmic proportions if there wasn’t a real entity behind those images. Unfortunately, Hacking’s emphasis on this line of thought is misplaced. The intuition can be attacked on various grounds. One might simply raise the possibility that the instruments are just (roughly speaking) calibrated to yield the same (artefact) image. More generally, the nature of the process of discarding the differences and retaining the similarities in the images—what ‘learning to use the microscope’ in essence amounts to—can be responsible for the achieved level of invariableness, regardless of whether the similarities are due to real entities or not. (van Fraassen, 1985: 298) Also, Reiner & Pierson (1995) criticise the coincidence argument as being ultimately abductive in form:

[Hacking’s] argument, however, invokes explanatoriness as a mark of truth—in fact, it does nothing else. This, however, is just the feature of IBE that has been criticised—by Hacking (1983:52) among others.

I do not think that every inductive argument that can be given ‘the coincidence gloss’ needs to be construed as abductive. I criticised such foundationism about IBE earlier (§1.2). But as far as the present argument from the

⁴In Hacking’s polemic these three get fused together. Here I consider the intuitions involved separately, one by one.

convergence of images obtained from ‘the plethora of scopes’ is concerned, I cannot but agree with Reiner & Pierson that it is difficult to see how the argument is meant to work, if not through an abductive inference. And Hacking’s rhetoric is certainly of no help in this regard.

However, the arguments of *intervention* and *the grid*—although also dressed in the language of coincidences—are more amenable to interpretation in terms of other forms of inductive inference. Let us consider the argument from intervention first. Hacking stresses throughout that ‘you learn to see through a microscope by doing, not just by looking’ (1985: 136). The point is simple: we learn to disregard the various artefacts of the microscope observation by realising that we cannot suitably interact with them. By contrast, when we can properly interact with the entities that we think we can see in the microscopic image, then we regard these as real. But what is it to *properly* interact to this end? Hacking eloquently portrays the intuition:

The conviction that a particular part of a cell is there as imaged is, to say the least, reinforced when, using straightforward physical means, you microinject a fluid into just that part of the cell. We see the tiny glass needle—a tool that we have ourselves hand crafted under the microscope—jerk through the cell wall. We see the lipid oozing out of the end of the needle as we gently turn the micrometer screw on a large, thoroughly macroscopic, plunge. Blast! Inept as I am, I have just burst the cell wall, and must try again on another specimen. (1983: 136)

One might well capsule the polemical gist of this graphic imagery into the slogan: *If you can poke them, they are real*. This does not mean, of course, that only real things can be poked, but rather attempts to bring out the relation between microscopic and macroscopic poking that directly provides us the conviction through the observed similarities we would expect if the relevant uniformity is assumed to hold in nature. Seeing a tiny object push through an elastic membrane of some resistance bears such a great degree of similarity to our macroscopic experiences with water balloons and sewing needles (or whatever) that the distinction between the real entities and the artefacts seems more than plausible to make. The realist conclusion that the physical interactions and the properties of the microscopic objects are of the

same kind as those of the similar macroscopic entities is almost forced upon us.

But only *almost*, of course, for to say that we can observe a ‘tiny glass needle’ et cetera, is already begging the question against the empiricist! Perhaps the microscope under which the tool was crafted is simply calibrated to give us an image of a recognisable needle form, without there being anything like a small glass needle generating this image. Ditto the elastic membrane, the act of injecting, and the rest. It again appears that the only way to respond to such a possibility is to invoke the minuscule likelihood (relative to our background knowledge) of such an elaborate calibration having taken place in the construction of the microscope, that it would again require a coincidence of cosmic proportions. Barring this approach, what the realist needs is a positive argument, not appealing to explanatory factors, for the assumption that there is no partitioning in the workings of the nature matching the limitations of human perception. I will next argue that for this purpose Hacking’s third line of thought, the argument of the grid (suitably reconstructed and analysed), will do.

The argument of the grid can be construed as independent of explanatory considerations, and it provides the missing premise for the argument of intervention considered above. Hacking portrays his intuitions by describing how a set of physical processes is employed (industrially) to reduce a macroscopic grid figure, with the squares indexed by the alphabet, to a microscopic size, to be handled by a pair of tweezers but only to be seen as a grid through a microscope. The gist of the intuition is nicely presented in more general terms by Magnus (2003) who calls it *the Galilean strategy*.⁵ This is an argument for obtaining warrant, for matters observable as well as unobservable, for an experimental method *M* the workings of which can be checked independently in the observable domain. Magnus summarises the Galilean strategy as the

⁵Something like the Galilean strategy was arguably employed by Galileo in convincing his peers of the reliability of his telescopes. Magnus follows initially Kitcher’s exposition of the Galilean strategy, but ultimately produces a different realist argument. Kitcher’s is not an experimental argument per se, whereas Magnus’s more focused variant is. Also, whether Magnus’s formulation of the Galilean strategy captures exactly Hacking’s intuitions about the grid is a moot question. What matters here is that the former can do the work that the latter was meant to do, i.e. reinforce the premise that there is no discontinuity in the nature at the boundary of what is humanly observable. Kitcher’s appropriation of the Galilean strategy will be critically considered in §3.3.1

following schema:

- $\mathcal{GS}1$ M provides correct answers up to and along the vague boundary between matters we can check independently of M and ones that we cannot check.
- $\mathcal{GS}2$ Prevailing reasons for thinking that the boundary might make a difference to the reliability of M are mistaken.
- $\mathcal{GS}3$ There is some significant positive reason to think that the success of M on matters that we can check generalizes to matters that we cannot check.

So we can trust that

- \mathcal{GS} M provides the correct answers for matters that we cannot check independently of M .

This strategy is then put into action by applying it to Hacking's grids:

Grids of ordinary size are photographically reduced and metalized [sic] using techniques which operate also in the macroscopic realm. ... We can imagine making a series of grids, the largest clearly observable to the average person without any magnification and the smallest unobservable to even the keenest eyes. This series of cases would show that the microscope is reliable at and through the limits of what the average person can observe using only their unaided vision. ($\mathcal{GS}1$ is satisfied for the optical microscope.) There is no reason to believe that the operation of the microscope *changes* when we point it at things just beyond the acuity of our sharp-eyed friends. ($\mathcal{GS}2$ is satisfied.) (2003: 468–469)

Furthermore, there is also a positive reason to think that the operation of a microscope is uniform in the required way across the boundary of what is humanly observable. To satisfy $\mathcal{GS}3$, we notice that

A microscope is the same observable, material object when used to view the date on a penny and when used to look at a paramecium. ... The very material of the instrument provides continuity between the cases in which it is used to look at observables and cases in which it is used to look at unobservables. (2003: 470)

Conjointly these premises allow us to draw the Galilean inference: ‘things we see in the microscope are really there.’ (*ibid.*)

The point here is, I take it, that we can reliably latch onto low-level uniformities in the world that transcend the limits of human perception. To this end we should also emphasise the continuity in the physical processes employed for the grid-reduction: using the same (reliable) equipment in very much the same (reliable) way gives us a *prima facie* reason to believe that the end result is a grid even when we haven’t even looked at it through a microscope. The idea is that what matters for the justifiability of such beliefs is not theoretical knowledge of the behaviour of light, explaining the observed uniformity, but the combination of the arbitrariness of the limits of human perception and some basic beliefs about the behaviour of matter.

Although the character and status of these ‘basic beliefs’ remains to be analysed and argued for, I find this sort of reasoning extremely plausible to begin with. Everyday examples of this kind of inference-making abound. Consider, for example, the technology of video recording and slow motion playback, and how it has shifted the limits of observability. No human being can directly observe the flapping wings of a hummingbird: they can beat up to 55 per second. Yet this directly unobservable process is brought visible to us by slow motion video playback, the reliability of which can be checked independently (subject only to memory) by trying it out on processes we can observe. Not that there ever was an interesting philosophical debate about the way the hummingbirds fly,⁶ but this just exposes the arbitrariness of the principled distinction drawn by the constructive empiricists, as far as the epistemological questions are concerned.⁷

In the face of the strong pull of such intuitions and the fact that attractive ampliative arguments can be formulated to capture such intuitions, the

⁶There has been interesting *scientific* debates about such matters, however. Consider, for example, the invention and development of the high-speed camera by Eadweard Muybridge in the 1880s, which rendered possible the scientific determination of a horse’s motion (*Scientific American*, October 19, 1878). The apparatus was later applied to flying birds and insects, of course.

⁷It might be objected that I have shifted to talk about observable *processes*, instead of entities. But science regularly informs us about the nature of dynamical processes we cannot directly observe, and temporal dimensions of a process ought to be on a par with spatial dimensions of a spatial entity, as far as van Fraassen’s conditions for observability of a scientific posit are concerned. See also Churchland (1985).

empiricist is likely to allow some leeway with regard to her notion of what is unobservable in principle, unobservable even with instruments satisfying the premises of the Galilean strategy. Indeed, van Fraassen has made the concession that he does not really care if we reject empiricism for the optical microscope, as long as we join him as regards the electron microscope.

The point of constructive empiricism is not lost if the line is drawn in a somewhat different way from the way I draw it. The point would be lost if no such line drawing is considered relevant to our understanding of science. (2001: 163)

Although ‘the point of constructive empiricism’ is not primarily epistemological at all for van Fraassen (cf. *Appendix*), the fact remains that the ‘optical microscopes do not reveal all that much of the cosmos, no matter how veridical or accurate their images are’ (*ibid.*). So although anti-realism is strictly speaking beginning to lose the game here, the level of realism salvaged thus far is rather low. The natural question is whether the Galilean strategy can be extended to observations through instruments the reliability of which cannot be checked directly, but only by employing the overlap between the domains of applicability of instruments in iterative fashion. Magnus expresses optimism in this regard, but there are likely to be serious complications for something like the electron microscope or the scanning tunnelling microscope, due to the exponentially increasing requirement of interpretational input. (cf. Bueno, 2006) Certainly nothing like the ‘Poking Principle’ of the argument of intervention is going to apply to give meaning to our language, and the risk of running afoul of illegitimate semantic analogies grows.

The limits and inferential principles of extending the experimental argument for realism along the lines developed so far pose interesting challenges to which I am not attempting to contribute here. Rather, I want to push forward by examining experimental arguments going beyond indirect observations by virtue of marshaling some other inferential guiding lines. According to these arguments we can have good experimental grounds, independently of IBE, for realist commitment to the paradigm unobservables: *electrons* and *atoms*.

2.3 Entities as tools

Hacking has made also another frequently cited contribution to the realism debate: an argument for *entity* realism from the experimental *manipulability* of unobservable entities. (1982, 1983: ch. 16) Entity realism is to be contrasted with realism about theories. Whereas the latter is driven by successful (novel) theoretical predictions and/or accommodations of phenomena, typically complemented with broadly explanatory considerations, entity realism has more to do with the laboratory practice of creating the phenomena itself. Hacking's writing is scattered with classic slogans: 'Engineering, not theorizing, is the proof of scientific realism about entities' (1982: 86); 'If you can spray them, they exist' (1983: 23). The metaphor of 'spraying' brings out nicely the most central epistemic condition of Hacking's entity realism: some unobservable entities are used as a *tool* to *intervene* with and cause new phenomena, and if we can actually do things with them, then they better be real. But let us attempt to go beyond slogans and metaphors here, to see if we can make sense of the intuition Hacking has adopted from experimental science.⁸

There have been two kinds of critical reactions to entity realism. One line of objection denies the stability of such middle ground with respect to our best theoretical understanding of those entities that Hacking is a realist about. It is denied that on the basis of experimental practice 'one can believe in some entities without believing in any particular theory in which they are embedded' (1983: 29). For example, Psillos states that

Can we assert that electrons are real, i.e. that such entities exist as part and parcel of the furniture of the world, without also asserting that they have *some* of the properties attributed to them by our best scientific theories? I take it that the two assertions stand or fall together (1999: 256)⁹

⁸There are many facets of Hacking's argument and his overall position that I do not mean to advocate in the following. There is a metaphysical tone to his entity realism that I have no taste for, and I cannot agree with the analysis of Putnam's causal theory of reference, which plays an undeniably important part for Hacking. Rather, my purpose is to motivate the role of particular kind of ampliative inferences in the realist argument, through these intuitive examples. The *ultimate* tenability or otherwise of Hacking's (or Cartwright's) entity realism does not concern me in this section.

⁹Cf. also McMullin (1987), Resnik (1994), for example.

But whether or not *some* of the theoretical attributes of QED are taken to apply to sprayable electrons, say, is not at issue with entity realism! The move from this assertion to the assumption that QED is ‘approximately true’ is a non sequitur, however, since the realist practice of judging the approximate truth of the theory can diverge from the practice of judging the low-level phenomenological generalisations employed in engineering and in operating an electron spraying machine.¹⁰ More could be said about this, in particular in regard to the theory realist’s notion of approximate truth, but let us move on to the second line of objection which is more germane to the theme of this chapter.

I want to focus on the claim variously made about Hacking’s argument being, at the bottom, just as abductive as the realist’s No-Miracles Argument. (Resnik 1994; Reiner & Pierson 1995; Psillos 1999; Devitt 1991) It would be interesting to assess Hacking’s PEGGY II electron gun example in detail here but that would take us needlessly far afield.¹¹ Rather, I want to give an idealised, simple example of a possible unobservable posit manipulated for certain effects, to explicate Hacking’s intuition (as I read him).

Consider the following scenario. A small amount of sleep-inducing substance *S* is found in nature. The substance is so potent that a microscopic part of it, when suitably administered with water and digested, will induce an immediate dormitive effect for at least 24 hours. *Hu Jintao*, the military leader of the People’s Liberation Army (as well as the president of the country) wants to perform the feat of anaesthetising the whole personnel of PLA by just 1 gram of this amazing opiate. Using microscopic techniques the substance is broken down into a million unobservable pieces, each administered to a loyal member of PLA, who then falls asleep for at least 24 hours. Do we have a reasonable inductive warrant for the belief that we have manipulated unobservable entities of the kind *S*?

¹⁰I do not want to deny the possibility of collapsing the distinction between entity and theory realism, but only the force of the existing arguments to this end.

¹¹Peggy II is a polarising electron gun build to study parity-violation in a weak neutral current interaction. I am not sure how well this particular example (with its more complex network of causal ‘low-level’ properties) fits the moral I try to draw from my rather more simple armchair thought experiment. But what is at stake here is the principled possibility of a realist argument from intervention. We can be sceptical about the warrant achieved for electrons without taking the general argument form to be a non-starter.

One might think that the argument is (again) based on our endeavour to avoid a disproportionate conspiracy, to which IBE is offered as the most natural principle of inference: the best explanation for the success of the amazing feat—given our background knowledge—is that microscopic quantities of S really were administered. Thus, Psillos writes:

the very same process is involved in accepting the reality of [a manipulated] entity and in accepting the (approximate) correction of its theoretical description. . . . In both cases, it is a judgment based on explanatory considerations. (1999: 257)

But the realist should resist this slide to the justificatory use of abduction. Instead, more direct forms of ampliative reasoning should be marshalled to permeate the observable-unobservable barrier. The argument of the previous subsection yields warrant for indirectly observing the microscopic bits that make up the initial 1 gram of S , and (in principle) observing their spatiotemporal trajectories all the way to the mouth of a PLA volunteer. The fact that the phenomena caused is so *similar* to the phenomena caused by manipulating directly observable quantities of S —the only difference being in the induced sleeping time, itself behaving logarithmically, say—yields warrant for the belief that the microscopic bits are of the S kind. Manipulability, using an entity to cause a wanted effect, *can* provide realist belief without recourse to inference to the best explanation, provided that the assumption that *similar effects are due to similar causes* can be defended to the required degree. We will return to different ways to construe this inductive assumption below.^{12,13}

¹²Admittedly this example is artificial and simplified in various respects. One might question, for example, whether manipulability per se has provided any content over and above that provided by the microscopic observations. Or one might wonder whether the lack of theoretical content in the produced phenomena (i.e. sleep 24h) distinguishes this case from the more interesting ones discussed by Hacking, in which the aimed phenomena is interpreted theoretically (e.g. parity violation in weak neutral current interaction). I think there are interesting questions about the limits and principles of applicability of this line of argument—questions which I do not attempt solve here—but the bottom line still stands: the fact that we can *do things* with unobservable objects can yield warrant about their nature and existence directly, without recourse to IBE, by virtue of there being similar deployment of similar causal properties at the macroscopic level.

¹³Resnik (1994) also claims that Hacking is bound to appeal to IBE in his causal inferences from manipulation. But Resnik does not claim that this is the only natural way to justify causal claims; rather, he maintains that Hacking is forced to do so on pain of

Once the realist argument begins in this way to revolve more generally around causation, it is no longer clear what role manipulability *per se* has to play in the inference. If we access unobservable causal properties via ampliative inferences of this sort, why exactly do we need to employ the property carrier as a *tool*? It seems that the line of ampliative reasoning sketched here is potentially applicable to uncovering the causal processes in nature which we just *spectate*, rather than *intervene* in. This suggests a more general take on experimental entity realism, something like the argument developed by Nancy Cartwright, considered next.

2.4 Cartwright on the role of causal explanation

Nancy Cartwright has emphasised the importance of causal (as opposed to theoretical) explanations for the realist. Much like Hacking and van Fraassen, Cartwright too dismisses inference to the best explanation on the basis of the underdetermination problem. In her reading the realist's inference to the best explanation—'the argument from coincidence'—draws especially on the *unificatory* nature of a good theoretical explanation:¹⁴

The more diverse the phenomena it explains, the more likely it is to be true. It would be an absurd coincidence if a wide variety of different kinds of phenomena were all explained by a particular law, and yet not were in reality consequent from the law. (1983: 75)

High-level theoretical explanations covering many phenomena are, however, subject to the problem of underdetermination. This is why Cartwright wants to bring the realist argument down to the level of phenomenological, causal

inconsistency: Hacking's denial of the truth of fundamental high-level theoretical laws is based, according to Resnik, on non-Humean metaphysics according to which 'causes are real, regularities are not' (1994). Regardless of whether or not this correctly represents Hacking's views, the gist of the *epistemological* position argued for here—the voluntary agnosticism about the theoretical truth—does certainly not depend on such a bizarre denial of regularities. High-level theoretical laws are not just unifying regularities, and as far as I can see the position is independent of the choice between Humean vs. non-Humean metaphysics.

¹⁴According to Cartwright this pursuit for generality and unification in high-level theoretical laws is an actual feature of scientific practice.

laws: ‘there is redundancy of theoretical treatment, but not of causal account’. (*ibid.*: 76)

Prima facie, however, it is not at all clear how appealing to causal explanations could thus help to escape the underdetermination challenge, for surely we can always at least imagine alternative causal stories to explain some given phenomenon. Cartwright has argued to the contrary that it is a logically necessary precondition of explaining causally that the causes cited exist, but the arguments are not fully convincing and to my mind they are successfully rebutted by Hitchcock (1992), for example. It is simply not the case that ‘causal explanations have truth built into them’, as Cartwright has put it. (1983: 91)¹⁵

It is more interesting for us to focus on those aspects of the argument that render it an *experimental* one. The role of experiments in Cartwright’s position comes out in the distinction she draws between theoretical and causal explanation, namely the distinction that

... unlike theoretical accounts, which can be justified only by an inference to the best explanation, causal accounts have an independent test of their truth: we can perform controlled experiments to find out if our causal stories are right or wrong. (1983: 82)

Now, what could this possibly mean? How can controlled *experiments* help in facing the underdetermination worry? Is it not just question begging to refer to experiments—to more phenomena—in response to the empiricist challenge, which claims that there are many ways the world could be, for all phenomena? Let us consider in detail the argument provided to see how the nature of theoretical causal stories in an experimental setting might help the realist cause.

Cartwright (1983: 82–85) offers Wesley Salmon’s (1984: 213–27) realist study of Jean Perrin’s argument for the existence of atoms as an example suitably explicating the crucial difference between theoretical and causal explanations. Perrin’s atomist verdict was based on his meticulous experiments

¹⁵Suarez (2006) has argued against the quantum mechanical counter-example produced by Hitchcock, but he has not in my view said enough by way of a positive argument for taking causal explanation to necessarily require the existence of the entity doing the explaining.

to determine Avogadro's number N by observing the Brownian motion exhibited by microscopic particles of gamboge (a kind a gum resin) suspended in dilute emulsion, together with a number of complementary experiments also entailing essentially the same value for N .¹⁶ Referring to the role of complementary experiments in Perrin's argument, she explicates:

For many, Perrin's reasoning is a paradigm of inference to the best explanation; and it shows the soundness of that method. I think this misdiagnoses the structure of the argument. Perrin does not make an inference to the best explanation, where explanation includes anything from theoretical laws to a detailed description of how the explanandum was brought about. He makes rather a more restricted inference—an inference to the most probable cause. (1983: 83)

Cartwright claims that a well-designed experiment reveals the character of the cause from the character of its observable effects: the causal reasoning employed in designing an experiment can, if strong enough, convince us of the reality of the cause on the basis of a single successful experiment. In Perrin's case there was no single experiment that would be convincing all by itself; rather, it was the convergence of the results of thirteen experiments that ultimately assured both Perrin and many of his previously sceptical contemporaries of the legitimacy of the inference to the atom hypothesis. Yet Cartwright wants to avoid the conclusion that the appeal to this convergence, or 'coincidence', just amounts to the inference to the best explanation. She claims that the inference is instead of the more restricted form: *inference to the most probable cause* (IPC).

In each of Perrin's thirteen cases we infer a concrete cause from a concrete effect. We are entitled to do so because we assume that causes make effects occur in just the way they do, via specific, concrete causal processes. The structure of the cause physically determines the structure of the effect. Coincidence enters Perrin's argument, but not in a way that supports inference to the best explanation in general.

¹⁶'Observing' is here employed liberally—as 'observing through a microscope'—on the basis of the argument explored above (§2.2).

My most charitable reading of these lines takes the idea here to be as follows. In all thirteen experiments there were some constant basic assumptions about the causal processes involved (some characters of the ‘specific, concrete causal processes’) which effectively unite the separate experiments into one, as far as testing those basic causal assumptions is concerned. Hence the only ‘coincidence’ involved is with regard to the results of this ‘single’ experiment: if the basic causal assumptions were mistaken, how could it be that the results of this (rather inclusive) experiment were just *as if* the basic causal assumptions were correct. Given our background knowledge of different causal processes, the causal assumptions tested in this way are the most probable ones.

This attempt to draw a line between IBE and IPC has met mixed responses. Psillos (2006*a*) analyses Cartwright’s advocacy of causal explanation from the perspective of IBE, and he reaches the conclusion that the realist force of IPC is parasitic on explanatory considerations in a way that aligns it with IBE. Thus, according to Psillos, there is no stable middle ground between accepting scientific instances of IBE as truth-conducive and accepting scientific causal inferences as entity-existence confirming. The case for this conclusion is made on the grounds that (i) causal reasoning is just a species of ampliative reasoning: ‘*qua* inferential procedures, causal explanation and theoretical explanation are on a par’, and hence the same justificatory arguments act for and against them both; (ii) on a closer analysis IPC gets its impetus from the *explanatory function* of the concrete causal processes appealed to in determining the most probable cause. Hence, if IPC is justified at all it is justified *qua* a species of IBE, and once this much has been established, it is only a short step for Psillos to argue that his more liberal use of IBE gets justified just the same. That is, allegedly Cartwright’s experimental argument for realism presupposes a more general defence of (causal) IBE, which she has not provided. And once the required justification for thus appreciating explanatory considerations is obtained through the No-Miracles Argument, say, the experimental argument is both valid and sound, but of no significant additional value to NMA itself.¹⁷

¹⁷Clarke (2001) attempts to drive a wedge between IPC and full-blown IBE by arguing that no explanatory considerations are needed in judging the most probable cause. But unfortunately he comes short of providing *any* account for how the most probable cause is

Psillos may be justified in being critical of Cartwright's rhetoric and exposition of IPC, but I want to dispute his negative verdict regarding the claims (i) and (ii) above in general. It seems that Psillos is painting with rather broad strokes here. By taking a closer look at Perrin's argument for the existence of atoms we can discern a mode of causal reasoning acting without the explanatory dimension of IBE. Peter Achinstein has provided the most helpful analysis of Perrin's theorising for this purpose. It is to this analysis that we now turn.

2.5 Achinstein on the reality of atoms

Jean Perrin's reasoning to the existence of atoms from his experiments on Brownian motion has been analysed as *eliminative-causal reasoning* by Achinstein (2001: ch. 12, 2002). Whereas Cartwright (1983) and Salmon (1984) emphasise the *convergence* of thirteen different kinds of experiments for determining Avogadro's number, Achinstein focuses on the conditions on which the kind of eliminative-causal reasoning that Perrin employs for the Brownian motion itself, is justified.¹⁸ The initial argument that Perrin provides prior to invoking his quantitative experimental data has the causal-eliminative form (2002: 474):

1. Given what is known, the possible causes of effect E (for example, Brownian motion) are C, C_1, \dots, C_n (for example, the motion of molecules, external vibrations, heat convection currents).
2. C_1, \dots, C_n do not cause E (since E continues when these factors are absent or altered).

So probably

3. C causes E .

then to be judged!

¹⁸Salmon (unlike Cartwright) closely analyses the 'argument of coincidence' in the framework of common cause inferences/explanations (1984: ch. 8). Unfortunately, the common cause argument can only yield justification for inference to *a* common cause (or common element in the experimental conditions), not to atoms as *the* common cause. Cf. van Fraassen (1980: 123) and Achinstein (2001).

We can observe the microscopic particles dancing around, continually accelerating and decelerating in a way that indicates the existence of internal forces responsible for such behaviour, assuming that no plausible external cause can be found. And the meticulous experiments performed by Guoy did indeed allow Perrin to eliminate the plausible external candidate causes C_1, \dots, C_n . The various experiments performed by himself and others then allow Perrin to claim quantitative evidence for his initial conclusion and for the numerical value of Avogadro's constant.¹⁹

There are two obvious anti-realist worries about the initial causal-eliminative argument of Perrin's. First of all, there is the possibility of the hypothesis of internal molecular forces being singled out by the eliminative reasoning merely as the best of a bad lot. How do we know that all the possible alternative causes of the phenomenon have been cited and eliminated by the experiments? Achinstein's response is to insist that the realm of possibility here is restricted by our background knowledge.

One can be justified in employing an eliminative-causal argument if, *given one's background information*, one has considered and eliminated all but one of the possible causes, or at least, all but one of the causes that (on the basis of the background information) have any significant probability of causing the phenomenon in question. The claim that the possible causes cited probably include the actual one can be defended by appeal to the fact that the phenomenon in question is of a certain type that, experience has shown, in other cases is caused by one or the other of the causes cited. (2002: 478)

The second worry has the form of the empiricist challenge, and it questions the last sentence of the above quote: how can we justify inferences to the unobservable on the basis of our experiences? For example, in Perrin's argument we need to justify the enumerative inference from 'All observed accelerating bodies in contact with other bodies exert forces on them' to 'All accelerating bodies, including molecules (if any exist), in contact with other bodies exert forces on them' (*ibid.*: 481). And empiricists like van Fraassen,

¹⁹The notion of evidence involved in the argument is carefully analysed in Achinstein (2001).

of course, take such inductive inferences to the unobservable to be unjustified and unjustifiable.

Achinstein responds to this second worry by adopting a line of thought very close to the Galilean strategy as employed by Magnus in connection with Hacking's argument of the grid.²⁰ The way Achinstein puts it, the realist can provide a positive empirical reason for taking 'observability' *not* to be a biasing condition for an inductive generalisation from a sample.

One can vary conditions or properties in virtue of which something is observable (or unobservable). For example, items can be observable (or unobservable) in virtue of their size, their distance from us in space or time, their duration, their interactions (or lack of them) with other items, and so on. ... If we vary the conditions in virtue of which bodies are observable and find no differences in whether bodies have mass, and if we have no contrary empirical information, then we have offered an empirical argument to support the claim that the fact that all observed bodies are observable does not bias the observed sample with respect to the property of having mass. (*ibid*: 484–485)

Hence, the kind of selective scepticism that the anti-realist advocates about ampliative inferences should feel some tension here. In particular, the kind of variation in the conditions and properties that the realist here appeals to do count when making legitimate ampliative inferences about unobserved observables. So whence the difference? After all, the logical possibility of *observability* being a biasing condition is on a par with the logical possibility of *observed* being a biasing condition.

Achinstein's construal of Perrin's argument for the existence of atoms is independent of the kinds of explanatory considerations that Psillos takes to be foundational for the realist project. In Cartwright's analysis the most probable cause is tracked down by higher-level causal reasoning, to be then corroborated by the experiment(s). This leaves the door open for Psillos to insist that unless more is said about the details of that kind of reasoning, we cannot properly distinguish between it and the kind of ampliative reasoning

²⁰Achinstein takes his position to resemble Kitcher's (2001) application of the Galilean strategy but, as will be discussed below in (3.3.1), these two are fundamentally different.

that is fuelled by explanatory considerations. By analysing the causal reasoning involved in the justificatory argument as *eliminative*, Achinstein brings out the missing detail: low-level reasoning about the possible causes yields a set from which one emerges as ‘the most probable cause’ through experimental elimination of the alternatives. The justificatory argument boils down to (a) the principle asserting that a phenomenon of a certain type is most probably caused by *a* type of cause that is related to it in our experience of the observable world, and (b) the applicability of the Galilean strategy. It remains to be discussed how to best understand these low-level inductive assumptions doing the justificatory work here.

* * *

There is a noticeable convergence in the above arguments for realism about various experimentally fathomable entities. We have surveyed a family of variegated but loosely connected arguments for observing and interacting with, and more generally, causally inferring the existence of, directly unobservable entities. It seems that the best way to construe these arguments by and large comes down to a juxtaposition of some rather basic beliefs about the uniformity of the world in (a) the ‘microscopic-macroscopic’ dimension, against (b) some other dimensions (e.g. ‘near-far’ in space/time). The anti-realist is happy only with (b) and the realist attempts to show that (a) and (b) stand or fall together. In the next chapter I will consider this argumentative strategy in the abstract, and different ways of expressing the kind of juxtaposition the strategy boils down to.

Appendix: Van Fraassen’s Image of Science

In this appendix I want to examine a popular line of attack on van Fraassen’s critique of abductive reasoning that simultaneously aims to gain some realist momentum. I believe that there is some useful insight to these arguments, but unfortunately they fail when targeted expressly against van Fraassen’s image of science. Luckily, my epistemic realist agenda in this thesis is largely independent of the realism/anti-realism issue from the perspective of constructive empiricism.

Psillos (1996*a*) accuses van Fraassen of adopting a selective attitude against inference to the best explanation.

Clearly, van Fraassen sustains a selective attitude towards IBE. The latter *is* a means of going beyond the realms of what has been actually observed and forming warranted beliefs about *unobserved* things and processes. Yet IBE is *not* a means of forming warranted beliefs about the realm of *unobservable* things or processes. (Psillos, 1996*a*: 34)

Thus, according to Psillos, van Fraassen is happy to countenance *horizontal* IBE to hypotheses about unobserved but observable entities, but at the same time denies the legitimacy of *vertical* IBE that involves hypotheses about unobservables. This division is strained, he then accuses, when it comes to the argument from the bad lot. The realist requires a level of *privilege* to ensure that the candidate explanations in the lot are good enough—a move we examined briefly earlier (§1.2)—but so allegedly does the constructive empiricist:

So in order to claim that the best currently available theory is empirically adequate, an ampliative claim is needed, asserting that scientists have already hit upon an empirically adequate theory. In particular, it would have to be claimed that it is unlikely that a theory which squares with observations up to now will cease to do so in the future, or in not yet tried space-time regions. . . . In all this, constructive empiricism would appeal to a background knowledge privilege, of the kind denied to realism. (1996*a*: 41–2)

McMullin (2003) tries out a somewhat similar line of attack on van Fraassen, albeit now in terms of van Fraassen's selective realism, rather than selective scepticism. He points out that there are many theoretical entities which van Fraassen is happily a realist about: the neutron stars or the dinosaurs, or the asteroid impacts. These are all *observable* in van Fraassen's sense but so far removed from us in time and/or place that they cannot be observed. What does it take for van Fraassen to be a realist about the *O-theories* postulating such theoretical objects, McMullin asks?

Van Fraassen's understanding of [constructive empiricism] commits him to holding that in the case of O-theory (but not U-theory), proven

empirical adequacy is sufficient to establish the realist credentials of the theory. That this makes him a realist of some sort in the locality of O-theory seems a fair conclusion. What kind of realist? That will depend on what kind of argument he can put forward for going beyond the safe haven of merely claiming to save phenomena at hand to make the more hazardous ampliative claim of empirical adequacy. (McMullin, 2003: 466)

The answer, as McMullin reads van Fraassen (esp. 1985: 266), is to be found in the measure of independent support a theory can enjoy over and above the data fit, gained from the predictions of novel phenomena and the unification the theory provides. But these are also the very virtues guiding the realist *retroduction* in the case of unobservable theories.²¹

To the extent that he can commit himself to asserting that a particular O-theory is empirically adequate and that thus its theoretical entities are real, it is to the diachronic virtues, in his version to those virtues that afford independent support, that he would need to turn. . . . But if this be so, why should not this form of argument be open in the case of U-theory also? (McMullin, 2003: 474)

These arguments against constructive empiricism nicely exemplify a general strategy to pursue against epistemological anti-realism—try to find a ‘natural’ way to juxtapose those ampliative inferences the anti-realist is happy with, to those she isn’t happy with—but against van Fraassen’s image of science this strategy fails. The reason is that the principal position of constructive empiricism is not ultimately an epistemological one at all! The constructive empiricist responses to Psillos (1996*a*) and McMullin (2003) are illuminating, and I will now recapitulate the reaction these pieces of criticisms have received. (Ladyman et al., 1996; van Fraassen, 2003)

Van Fraassen’s constructive empiricism is not an all encompassing philosophical position (although in van Fraassen’s writings it does get meshed with his more general epistemological views on empiricism, rationality etc.), but

²¹McMullin uses the term ‘retroduction’—going back to Peirce—to emphasise his perhaps somewhat idiosyncratic understanding of abductive inference. The details of this understanding do not matter to us here.

‘merely’ an answer to the question: What is science? In the framework of van Fraassen’s preferred (‘semantic’, or ‘model-theoretic’) view of theories, it purports to account for all of scientific activity in terms of *empirical adequacy*—a theory is empirically adequate if it has a model fitting the observable part of the world (past, present, future) described by the theory. The constructive empiricist *accepts* a theory as empirically adequate. Acceptance involves cognitive and pragmatic components: accepting a theory amounts to the belief that the theory is empirically adequate, *plus* a pragmatic commitment to employ the theory as a guide to further research, a commitment to the research program in question, a commitment to account for all the relevant phenomena without having to give up the theory. (van Fraassen, 1980: 88) Van Fraassen *defines* scientific anti-realism as the belief that we can understand in these terms *what science is*. By contrast, scientific realism (as defined by van Fraassen) is the view that in answering this question we need to appeal, not just to empirical adequacy, but to truth simpliciter.

I do not advocate constructive empiricism on the basis of the epistemic inaccessibility of the unobservable. I do not see the controversy between empiricist and realist in the philosophy of science in the first instance as a dispute over how much to believe. (van Fraassen, 2003: 490)

The crux of the matter is that the (anti-)realism debate thus understood is logically independent of the epistemological question of whether or not we can have warrant to believe in the scientific claims about the unobservable.

I see core realist and anti-realist views of science as answers to ‘What *is* science?’ which are logically independent of any epistemology. In this sense one could have an anti-realist view of science while believing in the complete literal truth of all currently accepted science. (van Fraassen, 2003: 481)

The significance of this separation of issues becomes clear in connection with the criticism that Psillos, McMullin and others (cf. Churchland & Hooker, 1985) have mounted against constructive empiricism. These arguments would be compelling against a purely epistemological divider, at least *prima facie*: if empirical adequacy were *just* a line drawn to guide what to believe in, it would

in light of these arguments appear quite unconvincing and unnatural. (But see the section §3.3, below, for my concerns about these particular ‘global’ ways to implement the general argumentative strategy.) But if we can understand what science is—its aims and its success in reaching those aims—without appealing to theoretical truth about the unobservable, as van Fraassen claims to have shown, then belief in truth really becomes superfluous *vis-à-vis* this particular philosophical project.

For van Fraassen the epistemological question of what to believe in on pain of irrationality is an additional one, and his empiricist answer to it is closely tied with his idiosyncratic and permissive views on rationality. The analysis of this ‘new epistemology’ is beyond the present work, but its potential repercussions should nevertheless be acknowledged: van Fraassen seeks to replace the whole traditional issue of whether we should be either agnostics or realists about theoretical entities, by the issue of whether we are rationally required to choose at all.²² Also, it becomes clear from the response to Psillos (Ladyman et al., 1996) that van Fraassen does *not* construe IBE as a warranted means to infer to the unobserved-but-observable. Rather, he is an atheist about IBE through and through.²³

But in a way this separation of issues is good news for epistemological realism (or *gnosticism*, as van Fraassen now calls it, following Forrest (1994)). Since the epistemic dimension is divorced from constructive empiricism, we can argue for (epistemic) realism without confronting van Fraassen on the issue of whether *understanding* science and its aims can proceed purely in terms of empirical adequacy and without recourse to IBE (at any level).

²²For an interesting assessment of van Fraassen’s epistemology see Psillos (2005a).

²³Perhaps there is a slight worry whether this is compatible with defending constructive empiricism, as defined above. I am inclined to agree with McMullin (2003: 465) here, in wondering what sense it makes to talk about empirical adequacy as the aim of science, unless that aim is in principle achievable. And if (i) empirical adequacy can be achieved, and (ii) if it is achieved through an IBE-infested methodology, then does it not follow that IBE is a warranted means to empirical adequacy, i.e. to truth about observables? Van Fraassen himself does, of course, deny the descriptive premise (ii) regarding IBE. (cf. van Fraassen, 2005)

Overcoming the Empiricist Challenge: Ampliative Inferences undivided

“If this criticism of the standard arguments for realism is right, a valid argument will not be all-embracing. It will not describe a general pattern of success characteristic of many sciences, and show that reason dictates acceptance of a corresponding realm of unobservables, wherever this pattern is found.”

Richard Miller, *Fact and Method*

Many realists have responded to the empiricist challenge of underdetermination by directly analysing the nature of ampliative reasoning, without either taking the debate down to the level of particular kinds of scientific experiments and instruments, or ascending to the global explanationist meta-level of the No-Miracles Argument. The general strategy of these realists is to argue that as a result of their analysis we can see how ampliative reasoning is in a significant sense of a piece, and respecting that integrity warrants realism about (some) unobservables. The thoroughgoing selective scepticism of the anti-realist, by contrast, disrespects the nature of our inductive practices in a way that makes it unnatural, or ad hoc.

This basic idea has been put to work in numerous ways. To place the various arguments in order a more fine-grained, graduated distinction between

global and local realism is first proposed (§3.1). I will then (in §3.2) claim that the experimental realist arguments of the previous chapter should be construed as a local realist analysis of the relevant instances of ampliative reasoning. Understanding the local-global distinction in the way I propose makes it in an interesting and useful sense a matter of degree. There are ‘local’ realist strategies that actually turn out to be rather global despite not adhering to the miracles intuition, and the reasoning of the first chapter can be extended to these arguments: there are reasons for being wary of the realist strategies that rely on a unitary high-level description achieved by analysing some feature of our inductive practices. Some of these ‘less local’ arguments by Kitcher, Lipton and others are reviewed and criticised (§3.3), before considering how local the realist must/can really go (§3.4 and §3.5).

3.1 Local vs. Global realist strategies

In the first chapter we considered the No-Miracles Argument, an attempt to appeal to the Inference to the Best Explanation at the meta-level in a rule-circular fashion, to yield a justification for the realist inference globally, across the board of mature, successful science. We found this global explanationist argument deeply problematic. For many realists this has signalled the need to approach the issue of justification more locally, and consequently the literature on local, or ‘piecemeal’, realist arguments is rich and colourful.¹ Unfortunately it is not always clear what exactly is meant by ‘local’ and its cognates, and in particular, whether the global-local distinction is meant to be absolute. I will now propose a way of understanding the distinction that makes it a matter of degree, and allows us to compare the various self-proclaimed local realist positions on *The Realist Spectrum*.

What characterises the *global orientation* of a realist strategy is the attempt to justify the realist inferences by reference to some rather general attribute unifying all these inferences. The local realist arguments, by contrast, take there to be more justificatory analysis to be done on case-by-case basis. Locality comes in degrees. A set of inferences can be unified by virtue of there being some single characteristic/form of an inference that acts as the vehicle

¹For some examples, recall the earlier quotes by Day & Kincaid, on p. 14 and Magnus & Callender, p. 29.

of justification for each inferential instance featuring that characteristic/form. Corresponding to the level of generality at which such characteristic/form is described—how encompassing the set of *such* inferences is—we have more local and less local realist strategies. This abstract preliminary distinction between global and local gets clarified in the subsections below via concrete examples of the realist positions entertained in the literature. It will turn out that many of the ‘local’ realist positions considered in this section are actually rather global, despite following a general strategy quite different from that of the global explanationism of the No-Miracles Argument. The diagram below indicates the (rough) ordering we will end up with.

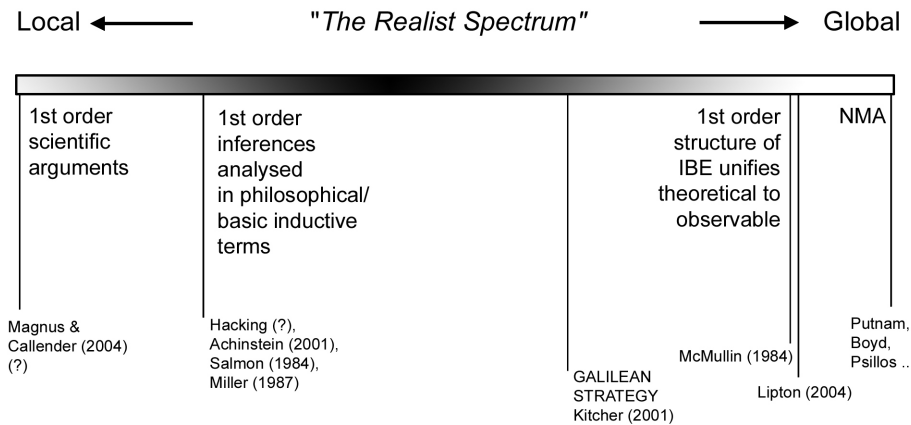


Figure 3.1: It is interesting to note that the term ‘local’ or ‘piecemeal’ has been associated with every realist position left of NMA.

There are two basic intuitions that pull the realists in opposite directions along the spectrum, described below.

3.1.1 Graduation of the realist commitment / warrant

Some realists take there to be a good *prima facie* motivation for incorporating into their realist position a certain *graduation of the realist commitment*: the intuition is that instead of being uniform and indiscriminating, the realist commitment—which unobservable features of the world described by the current science one believes in—comes in degrees. The realist commitment is more warranted for some unobservable items than for others, and some are excluded from the commitment altogether. The further reaches of the realist

commitment—quarks, black holes, dark matter—seem intuitively to be less secure and less warranted than those just below the surface of observability (microscopic living matter, moons of Jupiter (?), flapping wings of a hummingbird). Furthermore, the realist can decide to draw a line so that some scientific ‘knowledge’ can be altogether unwarranted by the realist principles.

This intuition is naturally accommodated by making the justificatory project a more local affair. If the argumentative strategy for realism about atoms, say, does not in and of itself apply to realism about electrons or quarks, a level of graduation ensues. Indeed, some realists use ‘piecemeal’ precisely in this way, to express the graduation of their realist commitment.

Realists...ought to claim that treatment of the elite class [of statements about observables] should sometimes be extended to the broader class [of statements about both observables and unobservables], although they may want to allow for conditions that militate against taking a realist attitude to some parts of science. [footnote: So, for example, a realist might adopt a different attitude towards the Ψ function in the Schrödinger formulation of quantum mechanics and towards the molecules discussed in molecular genetics.] The realism with which I’ll be concerned is a piecemeal realism. (Kitcher, 2001: 152)

Clearly, such graduation of the realist commitment must be also reflected in the justification the realist offers for the ampliative scientific inferences that allegedly give us the knowledge of such unobservables.

One central motivation for the idea that the realist commitment comes in degrees can be understood by considering the two perennial anti-realist challenges: underdetermination and pessimistic induction. The realist may try to avoid some of the force of these challenges by narrowing down her realist commitment by localising (to some degree) her justificatory strategy.²

Let us consider the challenge of underdetermination by data here.³ Justification of any ampliative inference faces this challenge to some degree or

²Realist commitment/warrant can also be graduated (to a degree) by adopting a suitable theory of confirmation. This (alternative) strategy is central to the global realist arguments, as will be discussed below. (cf. Psillos, 1999; Dorling, 1992)

³The problem of pessimistic induction, considered in detail in the next chapter (§4) can be also significantly lessened by graduating the realist commitments. It could be argued that the anti-realist’s historical data of instances of radical theory-change speaks

another. This, in essence, is *the* problem of induction. But regarding the various scientific ampliative inferences to the unobservable, it is plausible to view the problem of theoretical underdetermination as being essentially cumulative, so that the very high-level theoretical presuppositions are riskier than those ‘just below the surface’ of the empirical, by virtue of simply involving more non-demonstrative inferential steps. For example, one may share the intuition of Hacking (1981) about the unreasonably high level of ‘conspiracy’ required for the microscopic data to be what it is *without* there being cells and bacteria, et cetera. Hacking’s arguments from the possibility of intervention, and ‘the grid’, taken together, amount to a demonstration of the close *inferential proximity* of the ampliative realist conclusion about the entities we can observe through a microscope. Arguably we can appeal to this proximity in order to put the scientific belief about some microscopic entities on a par with certain everyday beliefs about the things more directly observable, beliefs which also require inferential justification. (Menuge, 1995; cf. also Psillos, 1999 and Chuchland, 1985) By contrast, one may at the same time feel that the level of ‘conspiracy’ required for something like the theory of Quantum Chromodynamics to be successful is much lower. The abstract highly theoretical mathematical structures of the latter are perhaps realisable by many equally plausible (and equally mysterious) entities and processes. Our intuitions about what the world must to be like in order for such very different scientific successes to take place, are not necessarily on a par.

Thus, a realist with a local bent may try to respond to the empiricist underdetermination challenge by distinguishing and defending the reliability of only scientific inferences of relatively close proximity. Underdetermination of theories in the higher reaches of science—whether constructively demonstrated or just imagined—presumably has no force against such realism. (cf. §4.2)

The promising possibility of dealing effectively with these two perennial anti-realist challenges, together with the undeniable continuum of our infer-

only selectively against some higher levels of realist commitment. The higher this level is, the more the realist can hope to salvage. Entity realists in general have never been that bothered about the theory-shift from ether to electromagnetic field, for example, and it is even harder to imagine a theory-shift that would seriously damage one’s realism about bacteria ‘observable’ through a microscope.

ential practices, provides a good prima facie motivation for pursuing realism in the more local end of the spectrum. The general strategy suggested by this intuition is bottom-up: starting from the kinds of ampliative inferences the anti-realist is happy to indulge in—i.e. inferences about observables (moons of Jupiter?)—and arguing that some knowledge of unobservables is obtainable via inferences that are *relevantly similar* to these vis-à-vis the general justificatory challenge.

3.1.2 The unity of the scientific method

When it comes to spelling out the sense in which certain ampliative inferences to unobservables are ‘relevantly similar’ to the ampliative inferences yielding everyday beliefs about observables, we may suddenly face an intuition that tends to pull the realist to the opposite direction. Thus, as we have seen in section §2.2, Hacking’s arguments from the pragmatics of experimental microscopy have been charged with smuggling in an abductive inference: allegedly the only way to make sense of Hacking’s appeal to the ‘conspiracy’ is via an inference to the best explanation, and hence Hacking needs to justify—just as a more global realist does—the reliability of scientific IBE in general. And Hacking’s and Cartwright’s arguments for entity realism from the experimental practice have met a similar response: these are intelligible only when understood as implementations of IBE. As Psillos insists:

Hacking notes, for instance: ‘We are completely convinced of the reality of electrons when we regularly set out to build and often enough succeed in building new kinds of device that use various well-understood causal properties of electrons to interfere in the more hypothetical parts of nature’ (1983: 265). I take it that the just described process by which ‘we are completely convinced’ that electrons are real involves two steps. The first step is positing a natural kind—electrons—and the second is relying on the ‘well-understood’ causal properties of the members of the kind in order to predict, or produce, certain effects. *Both steps presuppose the very same type of argument—inference to the best explanation.* (Psillos, 1999: 257)

Hence the justificatory commitments of entity realism are the same as those of theory realism, for justifying the reliability of IBE leads to general

theory realism (as long as theories can be construed as resulting from abductive reasoning). This motivates the project of justifying abductive inferences in science in general, and this in turn clearly points to the global end of the spectrum. We have also seen that it has been argued that inductive reasoning in the scientific method can to a significant degree be understood in the framework of inference to be best explanation. (Psillos, 2002; Lipton, 2004) This descriptive framework unifies the scientific method in a way that for some realists overrides the motivation for implementing the graduation of commitments by localising the justificatory strategy itself. Instead, the business of variable commitment/warrant is pushed down to the level of confirmation: a single realist argument applies equally well to all levels of unobservable and theoretical posits, but differences in the evidence (which can include theoretical virtues capturing the explanatory power) yield differences in epistemic warrant. (Psillos, 1999) Thus, the global realist can help herself to the full resources of the Bayesian confirmation theory which is, arguably, compatible with the unified IBE account of the scientific method. (Lipton, 2004)

These realists (e.g. Boyd, Psillos) who are fond of the descriptive unity of the inductive scientific method tend to form very general, global arguments which appeal to the success of the method thus described, the cumulative nature of science and the role of abductive reasoning in all this.

* * *

The two intuitions described above pull in opposite directions. Let us next consider more and less local realist strategies based on directly analysing the nature of ampliative reasoning in science.

3.2 Justification of low-level experimental inferences

The core intuition shared by the different versions of the Galilean strategy (as employed by Magnus and Achinstein), and by Hacking's argument from intervention and manipulability, can be expressed in the form of a kind of enumerative induction. Wesley Salmon (1984: 233) has called it 'argument by analogy', and for him it takes (crudely speaking) the following form

An effect of type E_1 is produced by a cause of type C_1

An effect of type E_2 is produced by a cause of type C_2

...

An effect of type E_k has occurred.

Conclusion: A cause of type C_k produced E_k .

An analogical argument of this sort can take us from premises about observables to a conclusion about unobservables, for C_k may be an unobservable cause that is similar to C_1, C_2, \dots in most respects other than size. (1984: 233)

Salmon is operating with a conception of causation that takes the causal relation to be *events*. By suitably categorising events into *types* we can discern and abstract the relevant causal properties involved in the argument. But since scientific laws relate these properties, it is perhaps more natural to think of the enumerative induction directly in terms of properties. Salmon's talk of analogy seems just misleading here, for there really are not two different domains, one observable and one unobservable, across which some analogy is applied. Rather, the induction operates directly on properties which have nothing per se to do with the observability of the entities instantiating them. For example, as already noted above, all observed accelerations of massive objects are caused by forces, so we inductively infer that all accelerations of massive objects are caused by forces, and not just that all accelerations of observable massive objects are caused by forces.

A hard-core empiricist can, of course, question the warrant for the realist belief that there really are massive bits of matter that cannot be seen by the naked eye. Supposedly it is a logical possibility that the term 'massive' only applies to directly observable objects, and the emergence of this macro-property is either somehow explicable or an ultimately primitive fact. But at this level of pyrrhonism the debate degenerates into questioning some very basic beliefs about the uniformity of the world, and sweeping Humean scepticism beckons.

Whatever justification is to be had for the different instances of Salmon's analogical reasoning and the accompanying basic beliefs, it cannot be derived from some form of inductive logic or from other very global, overarching

arguments.⁴ Instead, we must try to seek for justification of these basic beliefs in their own right.

John Norton (2003) has recently argued for a *material* (as opposed to *formal*) theory of induction. As far as matters of justification go, Norton argues that we must forgo the aspiration to formal universality for locality: ‘All inductions ultimately derive their licenses from facts pertinent to the matter of the induction’ (650). Here we are not concerned with justification of induction *in toto*, but only with the selective attitude of the anti-realists. But we can appropriate Norton’s perspective and adopt his division of an inductive inference into a *formal scheme* and a *material postulate*. When an induction—such as the one above to forces accelerating unobservable material objects—is licit, this is due to the matters of fact represented by our basic beliefs. The conclusions of experimental arguments of the previous section can then be uniformly defended against the anti-realist by seeing that the material postulates required for these realist inferences are not any worse supported than the material postulates required for various ampliative inferences the anti-realist is entirely happy with.

But it now seems that the emphasis on the *experimental* character of these realist arguments is misleading. What really counts is the locality and the *inferential proximity* of the relevant material postulates: some basic beliefs about the nature of the world are more warranted than others, given what we have observed. Analysing the relevant features of these material postulates and comparing them to the material postulates of some ampliative inferences to unobserved observables, amounts to bona fide philosophy that grounds the realist argument. This philosophical analysis applies to some cases of first-order scientific inferences, and thus yields realism about some unobservable scientific posits, but it is obviously a far cry from the more global realist arguments.

The rest of this section will first review and criticise some more global realist strategies which operate by analysing the unified nature of ampliative reasoning in the formal mode. Then it will consider the more local end of the spectrum, to argue that one still needs to philosophically analyse the mate-

⁴Salmon wants to embed the argument by analogy in the framework of objective Bayesianism, so that the analogies can supply some prior probabilities as required. Van Fraassen asks: ‘Whence objective?’ (1985: 299)

rial postulates to give a justification over and above the first-order scientific reasoning.

3.3 Kitcher, Lipton and others on (not-so-) Local Realism

3.3.1 Kitcher's Galilean Strategy

We have seen how the No-Miracles Argument strives to give us a reliable correlation between success and truth by appealing to a second-order IBE. Philip Kitcher (2001) has devised an interesting argument purporting to do much the same with less: his application of the *Galilean strategy* represents an attempt to underwrite the inference of (approximate) truth of a theory from its success (suitably construed), without invoking abductive reasoning at all.

Recall (from section §2.2) that the Galilean strategy is an argumentative schema over justificatory arguments of a particular form. According to this schema a method of justification can be validated by applying it successfully to observable instances that can be independently checked. Kitcher's ambitious application of the strategy is directly to the success-truth connection in a way that seeks to secure a rather strong form of realism.

People find themselves in all sorts of everyday situations in which objects are *temporarily* inaccessible, or are inaccessible to only some parties. Detectives infer the identities of criminals by constructing predictively successful stories about the crime, bridge players make bold contracts by arriving at predictively successful views about the distribution of cards, and in both instances the conclusion they reached can sometimes be verified subsequently. . . . [W]e come to believe that people usually only manage to achieve systematic success in prediction when their views about the underlying entities are roughly right. This belief is, I suggest, the source of our confidence in the 'success to truth' inference. . . (Kitcher, 2001: 176)

The basic idea here is that if we consider theories about *temporarily* unobservable matters of fact—i.e. theories the truth of which we can later independently check—we find a strong positive correlation between success and truth.

Barring any prevailing reason to think that the anthropocentric observable-unobservable boundary would make any difference to this correlation, we obtain, by following the Galilean strategy, a good reason to infer truth directly from success in the case of a theory featuring (permanently) unobservable matters.

I think we should consider Kitcher's strategy to be rather global by virtue of appealing to a very general feature of scientific theories: their success. This attribute must be carefully defined in order to rule out possible scenarios in which success does not entail truth. But this can be done, Kitcher maintains, by considering the various observable success-truth correlations. In this way we learn to value error-intolerant predictive/interventional tasks, for example. Kitcher's strategy does not explicitly appeal to unity in the scientific *methodology*, but it does assume that science is a unified enough phenomenon to which we can justifiably apply the Galilean strategy. It is required that the theorising the results of which we can independently check is *relevantly like* the theorising we are making inferences about.

It is exactly at this junction that worries must be raised about Kitcher's argument.⁵ The fact that we can give a unified *description* of scientific success and the constraints for the applicability of this notion (that we learn through observable theorising) does not amount to *justificatory* unity. The activity of theorising, the success of which is under consideration here, constitutes an extremely heterogeneous class. For example, the string theory may one day provide a set of predictive and interventional successes that fit the descriptive scheme well—extremely fine grained, error-intolerant, etc.—but we just don't

⁵Magnus (2003) validly poses the related question of whether any *positive* reason can be given for this assumed unity in the phenomenon of scientific theorising. There is an important disanalogy between this case and Galileo's and Magnus's application of the strategy to telescopic and microscopic observations, respectively: in the latter cases a positive reason can be allegedly given for there being a relevant sort of unity. Cf. section §2.2.

When telescopes and microscopes are pointed at observables or unobservables, they are the same material instrument;... this provides *prima facie* reason to think that [the requirement of having positive reason for relevant unity] is satisfied. Yet in the case of successful theories, the theories are not instruments made of the same stuff as one another. They are not made of anything at all. Thus, the presumption of continuity of cases for the microscope cannot be extended to the success-to-truth inference. (Magnus, 2003: 472)

have a handle on what is required for a theory at that level of abstractness to latch onto reality. In particular, the kind of success-truth correlation that we learn to trust in at the observable level just may not be generalizable to highly mathematical theorising of this kind. What *positive* reason can we give to the anti-realist agnostic to convince her of such a high-level unity in the world?

There thus seems to be some tension between Kitcher's pretension to 'go piecemeal'—recall the quote in section §3.1.1—and the notion of success he derives from observable theorising. As far as I can see there is nothing in this notion as explicated by Kitcher that would exclude the monumental successes in the further reaches of physics from counting as epistemically relevant. A realist of global bent may, of course, consider this a positive feature of Kitcher's strategy, but in view of the empiricist challenge one should really want to cherish the idea of more local realism. Perhaps some further constraints on the notion of success could be tailored to this end, but the more specific the constraints get the less work the generic, overarching success-truth connection does in the strategy. That is, despite having a unified abstract description of the realist strategy, the actual justification for one's realism about the bacteria is divorced from the justification for realism about the molecules of molecular genetics, due to widely different notions of success at play for each class of unobservable entities. The everyday successes that form the basis of the realist inference do not form a natural class.

3.3.2 Lipton and McMullin.

Unlike Kitcher, Lipton (2004) is a friend of abductive inferences. Yet he, too, views the global, meta-level application of IBE in the form of the No-Miracles Argument to be a dead end for realism. But instead of advocating a retreat to first-order case-by-case considerations, he puts forward a very general argument to unify and justify a significant class of the abductive inferences that scientists make. In particular, Lipton still wants to make good use of the fact that abductive reasoning seems to play a huge role in many scientific inferences to the unobservable.

The miracle argument is an inference to the best explanation but one that is supposed to be distinct from the multifarious inferences to the

best explanation that scientists make. Can explanationism defend realism instead by appeal to the structure of those first-order inferences? ... The structure of causal inferences is the same, whether the cause is observable or not. ... So there is a prima facie case for saying that all these inferences should be construed in the same way: granting the truth-tropism of inferences to observable causes, we ought also all to be realists about inferences to unobservable causes, since the inferences have the same form in both cases. (2004: 199–200)

So Lipton, although avoiding the 2nd-order global abductive inference about science, still advocates rather a global strategy by virtue of providing a very general template for the justification of scientific inferences. For him any scientific first-order instance of causal abduction is (probably) approximately true by virtue of being on a par with ‘structurally similar’ ampliative reasoning at the level of everyday observables.

In very much a similar vein, McMullin’s well-known *Case for Scientific Realism* (1984) disavows the meta-induction, but nevertheless retains a rather global spirit. For example, McMullin states that:

The form of the successful retroductive argument is the same at the micro- as at the macrolevel. If the success of the argument at the macrolevel is to be explained by postulating that something like the entities of the theory exist, the same ought to be true of arguments at the microlevel. (McMullin, 1984:14-5)⁶

The distinctive feature of his argument comes from McMullin’s emphasis on causal-*structural* inferences and explanations, in making the claim that there is a *single form* of retroduction common to macrotheories and microtheories, but the overall strategy is identical to Lipton’s.⁷

I find this kind of appeal to the structural uniformity of IBE problematic. In particular, it is the very generality of the justificatory template that

⁶See also McMullin (1987, 1994, 2003). Unfortunately I cannot do justice to all the subtleties of McMullin’s position here.

⁷Some read McMullin’s endorsement of structural inferences, together with the failure to apply the metaphysical categories of the macroworld to the microworld (of quantum mechanics), as pointing towards structural realism. (cf. Ladyman, 1998)

threatens to weaken the argument beyond its breaking point. As already discussed in section §1.2, IBE is a highly-generalised inferential template of two variables: an actual inference of this form is governed by (i) what counts as explanation and (ii) what counts as loveliness of such explanation. These two variables are context dependent—they are determined by the background. But this entails that the business of justifying a scientific instance of IBE is effectively done at the level of those constraints that determine what counts as a lovely explanation. The mere fact that an inference appeals to explanatory reasoning has nothing to do per se with its reliability on the basis of the track record of previous explanatory inferences. Lipton too (echoing Day & Kincaid, 1994) stresses the sensitivity of explanatory standards to our background beliefs, but he does not view that as a problem vis-à-vis his realist argument. Rather, Lipton’s response is to narrow down the kind of IBE instances that are grouped together for the purpose of generalising from inferences to observables to all inferences of that *kind*.⁸ But the more specific the description of the IBEs thus grouped together is, the more piecemeal the realism becomes and, in particular, the less work the shared-structure-argument does in itself. What rather does the justificatory work are the specific constraints on the general IBE template.

Take Achinstein’s causal-eliminative argument for the existence of atoms, for instance. We can formulate this as an inference to the best causal explanation if we like, but the justification of this inference would still lie at the level of the particular uniformity assumptions that carry the justificatory burden in Achinstein’s argument. These assumptions give us a *positive* reason to think that the loveliest causal explanation is the likeliest one in this case, given that ‘the loveliest’ is determined by the background encoding these assumptions. The applicability of the causal IBE model in itself only amounts to the weaker assumption that we have no *negative* reason to think that the applicability of the model does not warrant its conclusion.

⁸Personal communication

3.4 How local must you go?

Kitcher and Lipton are the ‘more global of the locals’ on the Realist Spectrum (p. 65). To put it in very general terms, my criticism above of both of these strategies derives from the fact that the global range of the arguments is achieved by displaying a high-level descriptive generality that is not coupled to the required kind of justificatory generality in any obvious or necessary way. We can analyse the situation in the abstract by considering how the descriptive generality potentially pulls away from the justificatory generality.

The descriptive generality of the kind appealed to in the above arguments is achieved by *abstraction*. This is how we get the general IBE template and the general argument from success to truth: by leaving out specific information about the background assumptions that fix the loveliness of an explanation, or about how to explain exactly how the success of a theory is on a par with some everyday success. But the price of abstraction is the risk of leaving out something that is relevant vis-à-vis the justificatory task. The more global the realist argument is and the higher the level of abstraction, the higher the risk of making illegitimate generalisations from the observable to the unobservable. To drive the point home we can consider a caricature realist argument of extreme generality.

Theoretical beliefs in science are formed by means of abductive reasoning. But so are most of our every-day commonsense beliefs. Realists have exploited this fact in order to argue that if one has no reason to doubt commonsense abductive reasoning, then one should have no reason to doubt abduction in science. The patten of reasoning, as well as justification, are the same in both cases. (Psillos, 1999: 211)

Such caricature realist argument is naïve, for surely there is much to the justification of everyday beliefs besides the *abstract pattern* of reasoning that can be taken to be descriptively adequate to it. Psillos, of course, is an extreme descriptive foundationalist about IBE unlike Lipton and McMullin, for example, and the latter two take it to be incumbent on the realist to provide more specific descriptions of the kind of abductive reasoning that allows us to generalise from the everyday (or observable) theorising to the

scientific. Hence Lipton stresses the causal-contrastive and McMullin the structural mode of IBE. But is that enough said?

As explained above, the resulting realism is still rather global, and I think the onus is on the realist to show that the naïvete of the caricature argument has been completely eliminated by the specific constraints applied. The inferences of science and everyday life form a colourful and heterogeneous class. The members of this class can be unified by abstract descriptions in many more ways than one. If there is a unified subclass that runs from the observable to the unobservable, we need *not only* to ask whether we can find a negative reason to think that the realist generalisation over that subclass is illegitimate, but *also* to ask whether we can provide a positive reason to think that there is none. This is the force of the empiricist challenge.⁹

Perhaps the intuition behind the arguments reviewed in the previous section is that such a positive reason can be given by meticulously spelling out the constraints for the notion of success, for example, or the exact form of the relevant explanatory considerations (including the background assumptions) of IBE. I have no problem with the spirit of such top-down approach to realism in principle, but I do challenge the realist with those preferences to work out the details of what I consider currently to be a set of promissory notes. Also, I assume that in spelling out the details much of the work will be actually done by low-level material postulates.

3.5 How local can you go?

All the arguments for realism begin with the first-order scientific inferences that, *prima facie*, seem to yield realist commitments to various unobservables by equally various arguments. Taking the conclusions of such arguments at face value without any further justification amounts to *fully local* realism. I have emphasised that the difference between local and global is a matter of degree, and I have criticised above the more global end of the spectrum. But I have at no point advocated fully local realism.

⁹Kitcher takes this reading of the empiricist challenge to be allied with a more skeptical ('Cartesian') epistemological starting point than he is willing to accept (personal communication). In this thesis I do not attempt to compare the realist and the empiricist positions at that level of disagreement.

Some seem to view the comparison of global and local in much more black and white terms. Magnus & Callender (2004) also provide an argument against ‘*wholesale*’ (viz. global) realism—an argument from base-rates that we considered and found wanting in the previous chapter (§1.4)—that takes them into full-on *retail* (viz. local) realism.¹⁰ Their piecemeal realism does not seem willing to go *at all* beyond the first-order scientific inferences. Thus, we ought to be realists about atoms, for instance, for ‘all the usual reasons’ given by Perrin and others, but to the suggestion that we might philosophically analyse the various first-order inferences to produce a realist argument unifying and philosophically justifying the truth-tropicity of the first-order inferences, they respond with surprising pessimism:

We acknowledge that it may be possible to get a kind of wholesale argument by discovering something in common among all good retail arguments for realism. Without trying to settle the larger epistemological issue, we offer a note of caution. Reflecting on the vast complexities of various historical episodes in science, there is no reason to think that the general assumptions one finds will be at all simple, natural, or even non-disjunctive; in short, there is no guarantee that the criterion one finds will be either interesting or useful. So although it is logically possible to turn a retail argument into a kind of wholesale argument, the resulting wholesale argument may appeal to ‘general assumptions’ that are long, gruesome, and can do none of the heavy lifting that wholesale arguments are usually meant to do. (2004: 335)

Although I am sympathetic to their cautiousness regarding the idea of recovering a simple, natural and completely unified global argument from the local level, I also find this dodging of the central philosophical issues slightly unsatisfactory. Surely there is more to be said by the realist about our (local realist) beliefs about atoms than what was said by Perrin and others at the turn of the century. This much is manifested in the just-reviewed literature philosophically analysing, and hence going beyond, the texts of Perrin et Co. And surely the philosophical defence required for realism about atoms,

¹⁰Although the very ends of my spectrum coincide with their ‘retail’ and ‘wholesale’, Magnus & Callender actually draw the distinction between these two attributes differently, in terms of base-rates.

say, contains general elements (Galilean strategy, eliminative-causal strategy...) that are potentially applicable to a wide variety of cases of ampliative reasoning of similar kind. One does not immediately arrive at a global, statistical wholesale realist argument in this way, of course, but neither is it the case that the local first-order scientific inferences are enough on their own. Local realism, too, requires an argument. And the arguments of the kind required—whether these are based on the higher-level form or lower-level material postulates—tend to unify the disparate cases and thus de-localise the resulting realist position in the name of general inductive ‘principles’. I conclude that this end of the spectrum—the 100% local realism—is also a non-starter.

* * *

Now where does this leave us? I have argued that the realist should not follow the very global strategy of NMA, but instead consider the prospects of analysing the nature of various ampliative inferences directly. I have argued that the popular experimental arguments for realism are ultimately dependent on an analysis of the unified nature of the material postulates implicit in our ampliative reasoning, rather than on an analysis of its unified form. I have discerned different degrees of globality in the arguments concerning the nature of ampliative inferences, and I have argued that the more global of these arguments have more work to do in filling in the *missing premise* that takes us from high-level descriptive unification to the required kind of justificatory unification. Hence, I have expressed my preference for the more local way of arguing for the required kind of unity in nature by recognising the relevant material postulates and directly comparing these with the material postulates that underlie the ampliative reasoning to unobserved observables. The general terms in which such comparison is to be effected are still unclear: there is thus much work to be done by the local realist too, for she cannot be content with the first-order arguments pure and simple. But the specific case-studies on realism about microscopic matter and atoms give some idea of the kinds of moves that are fruitful for the realist to consider. This concludes my discussion of the justificatory matters in the foundations of the realism debate.

Two Challenges to the Realist Image

“I dare say that for every highly successful theory in the history of science that we now believe to be genuinely referring theory, one could find half a dozen once successful theories that we now regard as substantially non-referring.”

Larry Laudan, *A Confutation of Convergent Realism*

In the *Introduction* I urged a dichotomy between *the issue of justification*, on one hand, and *the issue of the realist image*, on the other, as a useful way of handling the vastly burgeoning literature on the realism debate. The previous three chapters have dealt with the vexed topic of justification with a modest outcome. The rest of the thesis, beginning with the present chapter, is dedicated to the other delicate issue at the core of the realism debate: how to project a maximally optimistic realist image that is plausible and compatible with our best understanding of science. This chapter opens this new theme by first briefly expounding the general structure of the anti-realist challenge, and then reviewing (i) the challenge from empirical underdetermination and, in more detail, (ii) the challenge posed by the historical record of radical theory changes. Some have claimed that the argument at the heart of this latter dispute is, in its best-known form, actually fallacious. I will rebut such hopes for an easy victory in order to salvage the worthiness of the toil occupying

the second and third parts of this thesis.

4.1 The Polemic

The realist image of science is an epistemically optimistic one: a significant class of the theoretical claims of science are taken to be true; there are evidential relations between the empirical consequences and the explanatory goodness of a theory, on the one hand, and its truth-value on the other; and there is a rational methodology to follow to get a grip on these evidential relations. The polemic about the realist image concerns the most optimistic image we can plausibly portray of science as we know it. The question of justification concerns the *further* issue of whether we are ever warranted in taking such an image to be a truthful one.

As a starting point we can consider the *ultra-optimistic* ideal according to which *all* our present theories with appropriate evidential support are true. Such state of affairs, although possible, is extremely implausible given our best understanding of science. There are two obvious and well-known challenges to this realist image. First of all, it can be objected that empirical underdetermination is an inherent feature of our theories and this makes mockery of the notion of infallible support. Since incompatible theories can bear the same relation to empirical evidence, for one theory to stand in that relation tells us nothing of its truth value. Secondly, in so far as the realist takes a successful prediction or accommodation of some phenomenon to count as evidential support for the truth of a theory, we can see from the history of science that such support has a bad track record. It follows from these two challenges that the ultra-optimistic realist image is not only unjustified, but also implausible.

The real realist is not, of course, an ultra-optimist. The real realist claims that a more plausible optimist image of science is obtained by taking our present theories to be only *approximately* true, true *only in such-and-such respects*, and perhaps only so in *most* cases. Furthermore, the real realist claims that the notion of evidential support can be spelled out in a way that rules out the empirical underdetermination, at least in most cases.

But whatever the realist puts forward as a sophisticated, less naïve realist image, there is always the possibility that our best understanding of science,

informed by both historical and contemporary studies of actual theorising, does not corroborate it. The polemic about the realist image is about formulating different ways to portray an optimistic image of the truth content of our current best science, and analysing their plausibility with respect to actual science, both current and past.

Below I will first briefly review in some detail the challenge from underdetermination (§4.2), to lay down some groundwork for the discussion on *metaphysical underdetermination* that will take place in the next chapter. The main focus in this thesis will be on the challenge from historical theory changes (§4.3), although the underdetermination challenge will crop up later, too. Despite its failure at the level of justification, the No-Miracles Argument remains the psychological driving force for the realist image. The Pessimistic Induction pulls in the opposite direction and the realist needs to find a way of accommodating historical theory-shifts with some level of plausible optimism. One way of doing this—to be examined in Parts II and III—is to carefully define a suitable notion of approximate truth to delineate a level of continuity over otherwise radical theory-shifts. But for some all this seems quite unnecessary: there are responses to the Pessimistic Induction which attempt to show that some optimistic realist images of science are simply compatible with both the historical record of a great number of false past theories and the approximate truth of the current theories. I will argue in this chapter that the images of these particular responses are not plausible.

4.2 Challenge from Empirical Underdetermination

In the previous chapter we discussed how the issue of justification springs from a fundamental and undeniable kind of underdetermination: there are many logically possible ways the world could be, compatible with our knowledge of the observable world. This general underdetermination is very closely linked to the problem of induction, and it is manifested as scepticism or agnosticism about a particular class of ampliative inferences. The challenge from *empirical underdetermination of theories* has obvious connections to the issue of justification thus understood, but the two are not the same. Certainly the empiricist motivation can be partly drawn from the idea of empirical underdetermination—cf. van Fraassen (1980: ch. 3), Duhem (1954 [1906])—

but the latter does not exhaust the former. Indeed, our critical discussion of the selective scepticism about *everything* unobservable by the naked eye (including the wings of a flying hummingbird) had very little to do with, for example, the debate about the impact of a theory of confirmation on the issue of evidentially equivalent *theories*, which lies at the heart of the empirical underdetermination challenge. It is important to distinguish between the two issues, since we may be able to defend a realist image without achieving justification in the face of the more fundamental empiricist challenge.¹

The challenge from the empirical equivalence of theories gains momentum from our best understanding of science. Arguably it is a fact that a theory T committed to unobservables has empirically equivalent rival(s) T^* , and it is further claimed that therefore the semantic component of realism is incompatible with the epistemic component: given the realist's literal reading of theories we have equally well supported theories telling different, incompatible things about the world, which undermines the realist's epistemic optimism. The challenge is not (or at least not completely) a priori: the 'data' comes from a philosophical reading of actual science. For example, it has been claimed that the *holistic* dimension of theorising—the perpetual need for auxiliaries to draw out the empirical consequences of T —is a source of underdetermination (Duhem-Quine thesis). It has also been claimed that the way theories are expressed logico-semantically leaves room for algorithmic tweaking to yield infinitely many empirically equivalent rivals for any theory.² (Kukla, 1998) And, in particular, it has been claimed that there are significant actual examples of empirically equivalent theories which render an optimistic realist image implausible. (Earman, 1993)

Empirical equivalence is understood as underdetermination under *all possible* empirical *evidence* that could be generated for a theory. Regarding the *top-down* strategy that takes the existence of empirically equivalent rivals to be a universal phenomenon, this leaves room for at least two potential responses, both nicely presented in Laudan & Leplin (1991).

¹Indeed, one can deny the problem of empirical underdetermination without even aspiring to a realist image. The idea of having a rational methodology to choose a unique theory with does not by itself commit one to the truth-value of that theory. (cf. Laudan (1996))

²This is the closest thing to the a priori underdetermination fuelling the empiricist challenge, and must be answered along the same lines.

Firstly, we might question whether we have any grounds for ever claiming to know the empirical equivalence of two theories at any given time. Since the range of all possible empirical evidence for a theory depends on the auxiliaries accepted, we would have to take into account all the *possible* auxiliaries we might eventually come up with, which is something we cannot do at any given time. This in a way turns the Duhem-Quine thesis on its head, from vice into virtue. Perhaps an underdetermination indexed to time/auxiliaries is a problem enough for the realist image (Kukla, 1998), but it is not clear whether even this weaker thesis is sustainable (Devitt, 2002).

Secondly, we might question the link between empirical underdetermination and evidential underdetermination. For whereas the former is a matter of the logico-semantic relationship between the theory, its auxiliaries and the observational content, the latter concerns the epistemic status of the theory and depends on the theory of confirmation we adopt. Certainly the hypothetico-deductive view can be blamed for the mistake of equating them, and resources from Bayesianism and abductive inferences can be marshalled to sever the link between the two. Again, the realist manages to turn the challenge on its head by claiming that the reduction of evidential to semantic relations is not supported by actual science.

So not only is the widely assumed thesis of empirical underdetermination questionable, but its link to evidential underdetermination is also unsubstantiated. There are, of course, notorious examples aiming to convince us of the unavoidability of the universal underdetermination. Algorithms such as ‘Take a theory T , assert its observational consequences but deny the theory’ are ad hoc. They only manage to tap into the undeniable logical possibilities that fuel the empiricist challenge and must be responded along the same lines. (Devitt, 2002) But there are individual cases of underdetermination that cannot be thus responded to. These give rise to the *bottom-up* strategy to the underdetermination challenge. For example, van Fraassen (1980: 46–47) gives the example of Newtonian theory TN , together with the hypothesis R that the centre of mass of expanding universe is at rest with respect to absolute space. This is empirically equivalent to $TN + V$, where V is the hypothesis that the centre of mass has the constant absolute speed v . Earman states that ‘it is hard to get excited about this example... since $TN + R$ and $TN + V$ involve exactly the same ontology and ideology for space,

time and motion' (1993: 31), and I agree. Nevertheless, there are also more interesting examples of empirical indistinguishability, like the choice between the flat four-dimensional formulation of *TN* and the theory that displaces gravitational force in favour of non-flat affine connection. (*ibid.*)

But how far towards the general underdetermination thesis can the bottom-up, case-by-case approach take us? Earman expresses his pessimism about realism regarding this issue:

The production of a few concrete examples is enough to generate the worry that only a lack of imagination on our part prevents us from seeing comparable examples of underdetermination all over the map. (*ibid.*)

But I think intuitions will vary in this regard. As far as the *prima facie* plausibility of the realist image is concerned, it seems sufficient for the realist to cut down her realist commitments to certain levels or areas of inquiry. Psillos, for example, is happy to accept that the underdetermination challenge may get purchase *locally*, as long as the global version of the challenge can be kept at bay. (1999: 167) We just have to refine the realist image accordingly.

4.3 Challenge from Historical Theory Changes

Probably the best known argument against the realist image is the argument from Pessimistic Induction. (Poincaré 1952; Putnam 1978; Laudan 1981) This argument in some form or another has been part and parcel of the realism debate for quite some time now. It is therefore interesting to come across two recent papers which both claim that the argument in its best-known form is actually fallacious. (Lange 2002*a*; Lewis 2001) Here I want to re-establish the dignity of the Pessimistic Induction by calling to mind the basic objective of the argument, and hence restore the propriety of the realist program of responding to PMI by undermining one or another of its premises.

I take this argument against the realist image to be in essence the argument employed by Larry Laudan in *A confutation of convergent realism* (1981). Laudan appeals to a historical record of successful yet false theories to argue against the connection that realists like to draw between successfulness of a theory and its approximate truth—the connection that a successful

theory is deemed probably approximately true. The intuition behind the No-Miracles Argument motivates this connection, and in the present context we can take this intuition to motivate the realist image in which the successful theories of current science are approximately true. To sharpen the image we need to fix the meanings of ‘successful’ and ‘approximately true’, to yield a realist thesis that I call here NMA (in *sans serif*, to distinguish it from the justificatory No-Miracles Argument itself). This thesis simply says that a current scientific theory which is successful in a such-and-such way is probably true in such-and-such respects. PMI was devised—in the hands of Laudan, at least—to deliver a lethal blow to this thesis, by rendering it highly implausible in the face of history of science.

Laudan’s version of PMI can be succinctly reconstructed as the following reductio (Lewis 2001: 373, Psillos 1996*b*); call it PMI:

- (1) Assume that success of a theory is a reliable test for its truth.
- (2) So most current successful scientific theories are true.
- (3) Then most past scientific theories are false, since they differ from current successful theories in significant ways.
- (4) Many of these past theories were also successful.
- (5) So successfulness of a theory is not a reliable test for its truth (since this leads to contradiction in (3) and (4))

A typical realist response to this reductio can take issue with, for example, the implicit premise of step (3) by describing (usually via careful case studies) some theoretical elements solely responsible for the successfulness of past theories in a way that renders these theories continuous with otherwise incompatible current theories, and hence candidates of approximate truth in some suitable, restricted sense. I am personally very optimistic about such a line of response, developed in detail in the third Part of this thesis, but let us not question the premises of Laudan’s argument yet. Here my sole purpose is to stand up for the dignity of such premise defeating work against two lines of thought that claim to remove the anti-realist threat of PMI by denying the validity of the argument to begin with.

4.3.1 Lange's Turnover Fallacy

Lange (2002a) presents the *turnover fallacy* as a potential source of invalidity of pessimistic inductions in general (and not just of PMI against the realist). The basic idea of this fallacy can be conveyed by the following example:

Assume there is a board of directors comprising of ten members and that you are introduced as a new member to this board replacing someone else. Someone tells you that the company is in turmoil: there has been a change in the assemblage of the board two hundred and forty times in the past ten years, but you don't know who's been sitting in the board for how long. You pessimistically infer, inductively, that someone is going to be replaced again very soon. It could be you or it could be someone else for all you know.

You might be tempted to pessimistically infer that the probability of most of you getting the boot within a year, say, is quite high. But this would be to commit the turnover fallacy! For it could be that nine out of ten members of the board have actually sat in throughout the past ten years and it is only your 'predecessors', as it were, who came and went. Just by knowing *the number* of personnel changes in the board does not allow you to inductively infer anything about the probability for *any one* individual to get replaced—all you can infer is the high probability for *someone* to get replaced.

Now consider the case of scientific PMI. Looking at the set of current, well-confirmed, successful theories we may want to ask: 'How likely is it that most of these theories will turn out to be false and will be replaced by new theories incompatible with them?' Given a very bad *numerical* historical record of successful yet false theories we may be tempted—vaguely remembering the intuition behind the PMI argument—to answer 'Very likely'. But this would be to commit the turnover fallacy! For it could be that most of the current theories have been stable throughout the historical record tracking period, and all the numerous theory changes involve the 'predecessors', as it were, of only one current theory.

Although this is a point about a type of induction in general, Lange takes it to be telling against Laudan's argument in particular. The alleged lesson is that to validly infer the wanted conclusion—that most current theories are probably false—one needs to use a premise much stronger than (3) above in an argument of slightly different form.

... a pessimistic induction of a somewhat different and less familiar form is made impervious to the turnover fallacy by employing a historical premise that is not cumulative: at most past moments, most of the theories receiving wide acceptance at that moment are false (by current lights). (Lange 2002a: 284)

This is significant 'since the usual premise that most of the theories that have ever been accepted were false is inevitably more plausible than the needed premise: that at most past moments, most of the theories then accepted were false' (2002a: 285). A fallacy is committed, Lange proposes, since a typical statement of PMI (such as Laudan's) only refers to the *number* of past false theories as an inductive basis, and yet draws a conclusion about the high likelihood of any one of our present theories to be found false and replaced in the course of future science.

It must be admitted that Lange makes a fine point about pessimistic inductions in general, but nevertheless it seems that this potential fallacy *cannot* be incorporated against the scientific PMI of interest (that is, Laudan's PMI). Here we need to be more careful about the real objective of the PMI argument—what is the conclusion being inferred exactly? To begin with, note that the conclusion (5) above makes no reference to future times: what will be found false or whether any theory-shifts will take place. This argument PMI is therefore *not* an argument to the *time-dependent* conclusion that most of our current theories will be most likely found false and will be replaced. Rather, in the first place it is an argument to the *timeless* conclusion that '(5) So successfulness of a theory is not a reliable test for its truth'. As a matter of fact, in this conclusion no reference is made even to the probable falsity of any one theory of the current successful science; this conclusion would indeed hold even if the current theories were all likely to be true! And nonetheless the force of the argument is considerable given the key role of the claimed connection between success and approximate truth in the realist's image. It is interesting to notice that in the literature the term 'Pessimistic Induction', originally coined by Putnam in 1978, is invariably tagged on to the anti-realist line of thought the canonical formulation of which is taken as PMI (i.e. Laudan's reductio argument as presented above). The failure to properly distinguish Laudan's argument from Putnam's rhetoric is behind

Lange's undue optimism to be able to sidestep the anti-realist worry about the historical facts of science as they are typically told.

This reading of PMI—viz. merely as something to counter NMA—may feel unintuitively *neutral* to some.³ One may feel that PMI should have some pessimistic force on its own and not just as a reactive opposition to NMA, and we can indeed discern different levels of pessimism which PMI is sometimes taken to be an argument for. For example, witness Psillos' informal summary of Laudan's argument:

Therefore, by a simple (meta-)induction on scientific theories, our current successful theories are likely to be false (or, at any rate, are more likely to be false than true), and many or most of the theoretical terms featuring in them will turn out to be non-referential. (1999: 101)

This sentence perhaps typifies a more customary reading of PMI as entailing the probable falsity of any one of our current theories, and indeed this is the reading that Lange explicitly adopts. Is this reading of the argument, referring to the probable falsity of our current theories, now subject to the *turnover fallacy* as Lange suggests?

I believe not.⁴ First of all, we need to notice that this new argument is no longer just the reductio presented above.⁵ Rather, we now add to the above reductio a statistical argument along the following lines, call it PMI*

(1*) Of all the successful theories, current and past, most are taken to be false by the current lights.

(2*) The current theories are essentially no different from the past successful theories with respect to their 'observable' properties.

³Laudan (1981) does not use the term PMI, but I believe this 'weak' reading of PMI is closest to the use Laudan makes of his pessimistic historical record. This version of the anti-realist's argument is obviously already damaging against the realist, given the respective objectives of the two positions: even if PMI does not conclude that most current successful theories are probably false, the anti-realist has undermined the plausibility of the image according to which truth simply correlates with success.

⁴This is not to say, of course, that there are no other weaknesses or fallacies that the proposed simplistic classical statistical reasoning may succumb to.

⁵The argument is usually presented as a reductio as I have presented it (cf. Laudan 1981, Lewis 2001, Psillos 1999). Lange also refers to Laudan and Psillos in his discussion of the scientific pessimistic induction.

- (3*) Success of a current theory is not a reliable indicator of its truth (by the *reductio* argument above), and there is no other reliable indicator of truth for the current theories.
- (4*) Therefore any current successful theory is probably false by statistical reasoning.

This argument concludes that any one current successful theory, *ceteris paribus*, is probably false for all we know. The *ceteris paribus* clause effectively amounts to the premises (2*) and (3*) above: NMA is taken to be indiscriminating so that the current observer has no advantage over the past observers in evaluating the truthlikeness of a successful theory. Furthermore, this clause should be also taken to rule out all kinds of ‘relativisations’ of NMA to specific scientific domains: scientific methodologies and mechanisms are taken to be homogeneous across the domains and the competing realist and anti-realist arguments apply across the board. I take the content of these premises to be implicit in the standard construal of PMI.

The argument PMI* does not fall foul of the turnover fallacy. However, one may be tempted to *further* infer from such probable falsity the probability of finding a theory false and it getting replaced, but such an inference would go beyond the confines of—and indeed beyond the validity of—this version of pessimistic induction. Hence a *timeless* conclusion (4*) is inferred from timeless premises and no fallacy of turnover is being committed; this fallacy requires a reference to a time-dependent property (e.g. getting the boot within the next two weeks) in the conclusion but ‘being false’ is not such property.⁶ And a further argument to the conclusion that false theories *will be* replaced in the course of future science, whilst perhaps not unthinkable, is surely not part and parcel of the contemporary debate about the plausibility of the realist image.

Moreover, the conclusion of PMI* is clearly compatible with the kind of possible (asymmetric) state of affairs that Lange puts forward as problematic. Assume that all theory changes have taken place within just one domain of scientific enquiry, say. It seems, *pace* Lange, that we nonetheless have reason,

⁶Notice that there is a time-dependent part in the above quote from Psillos (1999) invalidly going beyond the confines of PMI. Curiously enough there is no such explicit mistake to be found in Lange’s exposition of PMI.

ceteris paribus, to believe that all domains of enquiry are currently ridden with false theories. This is because the only feature of theories appealed to in NMA is their successfulness and not, say, the duration of their reign. Once the connection between success and truth has been demolished by PMI, all the current successful theories (including those which we inductively have *no* reason to expect to get replaced) are on a par with all the past successful theories in one big domain of theories most of which are false, and the conclusion (4*) can be drawn. Furthermore, whilst the assumed asymmetric state of affairs undoubtedly begs for *some* explanation, it is not clear that we have any reason to think that *the best* explanation is achieved by hypothesising that the stable theories are true. What the realist needs is an argument to the conclusion that the combination of successfulness *and* long lifespan of a theory is best explained via truthlikeness, or something like that. As far as I know, no such version of NMA has yet been developed. On the other hand, our degree of confidence with respect to realism as a *possible* explanation of the asymmetric state of affairs is significantly lowered by Laudan's PMI and the availability of numerous other explanations, together with the ceteris paribus clause.

One may, of course, have grave doubts about the ceteris paribus clause in the above portrayal of PMI*, and many realists indeed argue that at least some current successful theories are not on a par with the past theories which are employed as the basis of the statistical inference above. But while this may offer a way to encounter this version of PMI, it does so by undermining one significant premise of the argument and not by virtue of showing it to harbour the turnover fallacy.

* * *

I prefer to follow Laudan and read the argument as the reductio PMI. We should notice that Laudan's argument is a somewhat atypical case of induction. Usually induction is described as an inference from the particular to the general, and it typically concerns states of affairs at future times being inferred from states of affairs at past times. But we have seen that PMI is not best characterised in such terms. Rather, PMI should be viewed as a reductio of an indiscriminating realist image—a challenge to the realist's beloved

connection between success and truth. Even if none of our current theories eventually succumbed to some incompatible successors, the anti-realist could nonetheless appeal to PMI as an anti-NMA. To do this, all that is required is a pool of theories all of which are successful at some time or another, yet most of which have turned out to be false.

So perhaps it is better to regard this meta-induction as a *statistical argument* against the realist claim that one ‘observable’ feature of our theories—successfulness—is a *reliable statistical indicator* of another, ‘unobservable’ feature of our theories: their truth(likeness). This is exactly what Peter Lewis (2001) does and claims that it falls victim to another kind of fallacy.

4.3.2 Lewis’s False Positives Fallacy

Lewis presents an altogether different rationale for regarding PMI thus understood as harbouring a fallacy. For Lewis the problem is that ‘the premise that many false past theories were successful does not warrant the assertion that success is not a reliable test for truth’ (2001: 374). More specifically: the *fallacy of false positives* that Lewis has in mind concerns the reliability of successfulness as an indicator of (approximate) truth. The notion of statistical reliability is usually characterised in statistics literature in terms of the rates of false positives and false negatives: a reliable indicator is one for which ‘the false positive rate and false negative rate are both sufficiently small, where what counts as sufficiently small is determined by the context’ (2001: 374–5). An instance of false positive (negative) indication is, of course, one in which the existence (absence) of an indication fails to reflect the existence (absence) of the indicated. The rate of false positives (negatives) is then calculated as the number of such cases per all negative (positive) cases.

With statistical reliability characterised in these terms Lewis then takes successfulness to be a reliable indicator of the (approximate) truth of a theory T (picked at random out of *all* theories at time t) if and only if the rate α of false-yet-successful theories is small and the rate β of true-but-unsuccessful theories is small. With this notion of statistical reliability at hand Lewis explains why Laudan’s *reductio* formulation of PMI is a non sequitur:

At a given time in the past, it may well be that false theories vastly outnumber true theories. In that case, even if only a small propor-

tion of false theories are successful, and even if a large proportion of true theories are successful, the successful false theories may outnumber successful true theories. So the fact that successful false theories outnumber successful true theories at some time does nothing to undermine the reliability of success as a test for truth at *that* time, let alone other times. In other words, the realist can interpret Laudan's historical cases, not as evidence against the reliability of success as a test for truth, but merely as evidence of the scarcity of true theories in the past. (2001: 377)

And to do otherwise is, Lewis proposes, to commit the fallacy of false positives.

The basic intuition behind this argument is made most clear in pictorial terms:

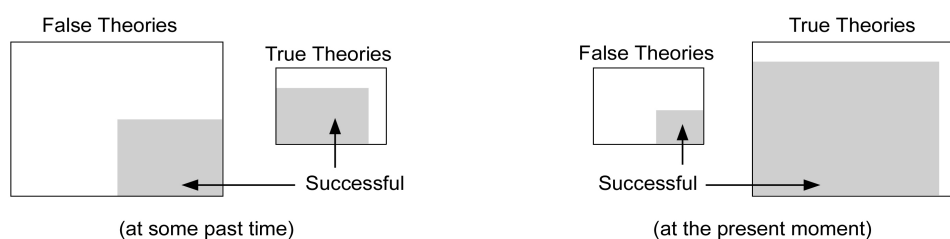


Figure 4.1: Domains compatible with both statistical reliability and 'bad' historical record. The ratio of successful (grey) and unsuccessful (white) is constant for each domain over time.

We can see immediately that by having a big enough domain of *false and unsuccessful* theories we can satisfy the requirement of statistical reliability even in cases in which, somewhat unintuitively perhaps, the probability of a randomly drawn successful theory to be true is small (less than 0.5, say). At both times pictured the requirement of statistical reliability is satisfied. Furthermore, given that we take *most of our current theories to be successful*, it follows 'deductively that most current theories are true, as required by the realist' (2001: 375). This Lewis takes to be a reasonable justification for regarding statistical reliability to be a notion that adequately captures the realist's appeal to the success-versus-truth connection.

So the notion of statistical reliability works for Lewis on the assumption that the statistical reference classes (relative to which the statistical reliability is determined) are of the right kind and vary radically as we move from past to current theories: the domain of all theories at some past time t_p must contain a much higher proportion of false and unsuccessful theories than the domain of all current theories. This immediately raises a couple of worries regarding the overall framework in which Lewis casts his realist image and allegedly sidesteps the challenge of PMI: (1) How are the crucial reference classes defined in the first place? (2) Has there really been a change in the reference classes such as to enable Lewis's response to PMI to get off the ground?

(1) First of all, it is not at all clear that the notion of the reference class of statistical reliability is well-defined in the context of scientific theories. It seems that the relevant domains of *all* true theories and *all* false theories (at some time t) with respect to which the rates of false negatives and false positives are calculated are not straightforwardly definable in the way a pool of people, say, is readily given in a typical case of medical statistics, for example. Not much has been said in the discussion so far about the putative identity conditions of theories—it just has been surmised that they could in principle be given. But whereas this assumption may be reasonable with respect to both the set of successful theories and the set of true theories, I can make no sense of the idea of delineating a non-arbitrary, well-defined collection of *both false and unsuccessful* theories.

Lewis's realism-friendly scenario which makes Laudan's historical record compatible with success being a reliable statistical indicator depends on there having been a large domain of such false and unsuccessful theories relative to which the rate of false positives is small.⁷ But what exactly are the theories which are neither successful nor true? Should we count in only the theory-proposals made by eminent scientists, or perhaps all the proposals actually published in scientific journals, or what? It is easy to imagine a variety of

⁷Lewis's proposal for testing the history of science for the pessimistic conclusion of PMI in a *valid* way consists of taking 'a random sample of theories which are known to be false, and show[ing] that a significant proportion of them are nevertheless successful' (2001: 378). The worry now is that this testing cannot be done since the domain in question is ill-defined.

sociological factors, say, yielding scores of unsuccessful and false theories, directly affecting the notion of reliability at stake. But why should we care about *those* theories? It just seems that the debate between NMA and PMI does not involve unsuccessful and false theories (or true yet unsuccessful, for that matter) in anything like the way Lewis projects.

But it seems that the realist should really give us some rough idea of how many false and unsuccessful theories there are per each successful one, given that *ultimately* the plausibility of the optimistic realist image—the justificatory NMA, being a form of inference to the best explanation (cf. §1.4)—seems to hang on the assumption that this ratio is not high enough to explain away the ‘miracle’ of successful science by the mere number of trials. Indeed, I argued in chapter §1 that it is part of the realist image of science that it is *not* an enterprise correctly depicted by the ‘selectionist’ picture offered by the anti-realist to explain its success. So the realist really needs to say something about the ‘ratio’ of false and unsuccessful theoretical proposals to the successful ones. But I would maintain that the realist arguments to this end can remain qualitative and emphasise the success and unity of the abductive methodology, for example, rather than degenerate into senseless estimates of the *number* of unsuccessful theories required for every successful one. And however we decided to delineate the domain of all theories for this purpose it should not be the case that the realist explanation is held hostage to contingent matters regarding the number of false and unsuccessful theories in the strict manner implied by Lewis’s strategy; realism simply cannot depend on the alleged (contingent) fact that most current theories are successful! Rather, it is implicit in the No-Miracles intuition that any feasible fluctuation in the number of false and unsuccessful theories—feasible to science as we know it—is not large enough to overthrow the justificatory NMA as the *best* explanation around.⁸

(2) So has there been a change in the reference classes of the kind that Lewisian realism requires? The idea is that realism only requires that most

⁸Whether or not the assumptions implicit in this justificatory NMA hold is another matter, of course. The point is that Lewis has not only put forward a response to PMI but also a particular understanding of realism to go with it. The problems with the former really spring from the inadequacy of the latter, already identified in section §1.4.

of our current theories are true which deductively follows, given good statistical reliability of success as an indicator of truth, from the premise that most of our current theories are successful. That is, given any one successful theory—current or past—the best explanation for a Lewisian realist of its successfulness is that *either* it is (approximately) true *or* it is a member of a huge domain of false theories a small portion of which are successful. Regarding past successful science, at least, this is fully amenable to an anti-realist reading. To an anti-realist like Bas van Fraassen—who persistently denies the justificatory force of the No-Miracles Argument—an explanation such as the above is good enough and fully consonant with his Darwinian selectionist image of science. For van Fraassen, of course, this picture fits the bill with respect to current science just as well; that is, he denies the initial premise of Lewis’s that most of our current theories are successful. But the soundness of that premise is neither necessary nor sufficient for the realist to make a case against van Fraassen; what is required is NMA as typically understood and the intuition that (approximate) truth is thus connected to successfulness—and for that intuition to have bite is for it to have bite *at all times*, regardless of the number of false and unsuccessful theories present at the time in question.⁹

As a matter of fact, Lewis’s unorthodox formulation of the realist position seems to beg the question against this point to begin with. According to Lewis ‘convergent realism usually includes the thesis that most of our current theories are true’ (2001: 371). But this is certainly an unreasonably strong thesis for any realist to aspire to: contingent matters regarding the number of false and unsuccessful theories produced by the scientific community depend on factors quite independent from realism and NMA—or so the realist argues—which is why convergence is typically characterised in terms of increasing level of ‘truthlikeness’ in a sequence of *successful* theories of cumulative empirical adequacy. Lewis’s convergent realist is committed ‘to the empirical claim that successful theories were rare in the past and are common today’ (2001: 377). Such commitment is not generally acknowledged to be part of

⁹Unless, of course, that ‘number’ is *so high* as to undermine the credibility of NMA as the best explanation altogether as explained in (1) above. Lewis stresses ‘the inference that the realist wishes to draw from the success of most *current* theories to their truth’ (2001: 378, my italics) but this requires that the realist accounts for some principled difference between the current and the past. And Lewis does not provide such an account.

any contemporary realist position. And it better not be! Keeping in mind how strict a qualification ‘successful’ can be for the realist and casually glancing through *The Journal of Mathematical Physics*, for example, one is bound to be convinced of the sheer incredibility of this premise upon which realism à la Lewis is erected.

4.3.3 What Pessimistic Induction is

Despite Lange’s and Lewis’s respective attempts to short-circuit the Pessimistic Induction it remains a powerful force to be reckoned with. There is no easy way out for the realist; one or another of the premises must be defeated.

Providing a realist image begins by portraying the alleged truth content of some present theories. It is also part of the realist image that scientific evidence of some description is connected to that truth content in a way that allows us to access the latter by attaining the former. All this can be viewed as being independent of the complementary realist project of providing a justificatory argument for the actual faithfulness of the image. Typically these latter arguments are quite global and refer to some general features of a theory: the form of its ampliative inferences or its success, for example (cf. §3.1). Corresponding to each such justificatory argument there is a description of a realist image that makes a general claim about all theories of the relevant kind. In principle a pessimistic induction (a *reductio*, rather) is possible against any one of these realist positions, and such anti-realist arguments should be viewed simply as historically based claims that the image in question is as a matter of fact implausible.

The fact that the PMI argument is non-deductive allows the realist to come up with a distribution of truth values which is consonant with both the realist conclusion (regarding our current theories) and the anti-realist premise concerning the number of (past) false theories of the relevant kind. But the mere logical possibility of such distributions shows nothing in itself: the plausibility of the realist image hangs on the further issue of whether the realist can explicate some principled difference between the domain of false theories and those theories the realist takes to be true. Lange and Lewis have focused on their respective formulations of the anti-realist challenge without

appreciating what these arguments are arguments against. The challenge remains valid against a popular way of framing the realist image, and Lange and Lewis both fail to give a positive indication of how this popular image might be surpassed by something that does not fall foul of the challenge and plausibly coheres with our best understanding of science.

In the case of Laudan's argument the target is the image projected by the No Miracles intuition according to which successful theories are mostly true. Laudan claims that it is apparent from the history of science that most successful theories are actually false. Hence, even if the justificatory argument was fully valid—an assumption Laudan also rejects—it would thus be shown not to warrant the conclusion. Generally speaking, the force of the historical data on which the Pessimistic Induction relies depends on the realist image it attempts to undermine. The discussion in the literature has mostly revolved around global realism fuelled by the intuition that success is a sign of theoretical truth, and the realists have focused on ways to refine the realist image to undercut one or another of the premises of the anti-realist argument.

In the rest of this thesis, Parts II and III, I will shift the focus on explicating and assessing some of these existing realist responses to PMI, as well as developing an interesting variant approach. The responses looked at attempt to undermine the assumption of Laudan's argument that 'Most past scientific theories are false, since they differ from current successful theories in significant ways' (i.e. premise 3 of PMI on p. 87). The general idea of the realist defence is to offer a principled notion of *approximate truth* that allows her to claim that the existence of prima facie significant differences between a past and a present theory does not entail that the former is simply false, i.e. not even approximately true. Depending on our realist image as regards the present theories, the argument goes, we can simply *dismiss* the features of the past theory that do not correspond to the reality as we now perceive it through our current best theories. These different notions of approximate truth are explicated and tested through case studies from the actual science, and the more troublesome the purported problem-case seems to be for the realist, prima facie, the stronger the force of the realist rhetoric is. We can take an ether theory of optics, for instance, as a piece of mature, successful

science, and try to delineate its truth content in some principled manner.¹⁰
(cf. Part III)

The recent years have seen an influential, broadly *structuralist* development in the realist arguments in this regard. The next Part is a study of this movement.

¹⁰But it is not part of the realist argument that each and every case can be dealt with in these terms; the realist image can allow for sporadic falsehoods.

Part II

Structural Realism and its Structure

What is Structural Realism?

“Aside from its importance as a contribution to the literature on approximate truth, structural realism is significant in two other ways. In the first place, it reflects a general tendency in the literature on scientific realism to worry about the extent to which scientific realists must portray scientific knowledge as potentially resolving genuinely metaphysical questions.”

Richard Boyd, ‘*Scientific Realism*’

Unfortunately there is no simple answer to this question. The terms ‘structural realism’ and ‘structuralism’ have many meanings in the philosophy of science, and there are various motivations underlying the positions carrying these labels. Sometimes these meanings and motivations are not clearly distinguished, and obscurity and the risk of equivocation ensues. Historically speaking structuralism is a well-represented philosophical trend with some rather distinguished exponents, and many of the contemporary advocates see themselves as appropriating for some issue of current interest the deep thoughts of their (chosen) predecessor(s). Recently structuralism has enjoyed a great deal of attention in the broad context of the present day scientific realism debate.¹ Indeed, many consider structural realism as *the* main contender

¹See the special issue of *Synthese* **36**(1) (2003), and also *Philosophy of Science* **73**(5)

for a defensible realist image, but the grounds for this belief vary widely.

I will focus here solely on structuralism in the context of epistemological scientific realism, as a solution to the issue of the realist image. The basic question is whether there are good reasons for contemplating on ‘going structural’ in delineating the realist content of theories (this chapter), and if so whether this can really be nicely done by following a broadly structuralist path (the next chapter and §7). I am trying to avoid as much as possible the terminological issue of whether this or that position should be called *structuralist*, given some ‘natural’ reading of the term. Rather, the primary concern here is with possible novel forms of realism. This is worth emphasising because in philosophy the word ‘structure’ carries strong (but variegated) connotations, leading easily to fruitless terminological quarrels.

The objective of this chapter is to review and scrutinise the principal motivations behind structural realism. I will begin with what I think is ultimately the *only* valid motivation for structuralism vis-à-vis the issue of the realist image: the epistemological refinement of realist commitments (§5.1). After that I will criticise an attempt to surpass epistemic structuralism with the so-called metaphysical or ontological structural realism (§5.2), and consider the potential virtues and liabilities of adopting a specific meta-scientific framework in which to address the whole issue (§5.3). The next chapter will then look in detail into the possibility of expressing structural realism via *Ramsey-sentences*, allegedly the most rigorous way of articulating structural content.

5.1 Epistemological Motivations

In the previous chapter we considered the challenge for the realist program posed by the historical record of the Pessimistic Induction. The realist wants to paint an image of the truth content of theories that is plausible regarding our best understanding of science and its method, motivated by the intuition behind the No Miracles Argument, and not falsified in the face of the history of science. This requires a principled way of delineating continuity in the

(2006) for papers presented in a symposium organised in the recent PSA2004 conference, for example.

truth content across radical theory shifts, if such shifts are considered to be so frequent as to threaten the plausibility of the naïve realism. We should now contemplate the possibility of defining the *structural content* of a theory in such a way as to ensure cumulative continuity in that kind of content.

Given the nature of this challenge it seems that it should be approached by reviewing a significant number of the theory shifts that form the basis of the anti-realist worry. Only after such a fantastic historico-philosophical feat could we *really* claim to know the kind of continuity that takes place in science, by and large. Unsurprisingly, this task remains to be completed. The philosophers writing on the subject have tended to employ general arguments that rely on our broad understanding of science, accented with one or two more detailed case studies of some particularly striking (and allegedly particularly problematic) instance of theory change. I plan to follow the same approach here: I will first consider the epistemic motivation for structural realism in the abstract, and in the next Part (chapter §8) I will look at a case study. But the shortcomings of this line of argument should not be understated; the ultimate corroboration or otherwise of any realist image comes only via extensive comparison with the actual history of science.

The clearest intuition driving the structural realist image was beautifully captured by Poincaré in his *Science and Hypothesis*, and later successfully resuscitated and promoted by Worrall (1989) (following Elie Zahar) in the context of the modern realism debate, as a response to PMI.² This motivation very simply springs from the fact that in various instances of theory change there are crucial mathematical equations that are carried over either intact or, more typically, as one set of equations being a limiting case of the other. Hence, we have examples of theory shifts in which the crucial equations of the two theories are (a) formally *identical*, but furnished with divergent interpretations (e.g. the shift mentioned by Poincaré and Worrall from Fresnel's ether theory to Maxwell's electromagnetic theory of optics—cf. §8); (b) formally equivalent apart from one or two new parameters, say, which disappear at some well-defined mathematical limit to yield the old equation (e.g. moving from the Galilean to the Lorentzian inhomogeneous group of transformations). The latter (more typical) cases get arguably further sup-

²Mary Domski has pointed out that Poincaré himself was no realist, but a neo-Kantian.

port from the endorsement of Heinz Post's (1971) *general correspondence principle* according to which any acceptable new theory should explain the well-confirmed part of its predecessor. Although this principle can be viewed to manifest itself *de facto* in a variety of ways (Hartmann, 2002), an important dimension of Post's well-received notion of both descriptive and prescriptive correspondence in modern science is undeniably of the relevant mathematico-structural kind. So Worrall's suggestion was to take the theoretical continuity manifested as such formal mathematical correspondence to be the focus of the realist commitment.

The above is a valid structuralist intuition. The highly mathematical nature of modern science together with the presumed descriptive soundness of the general correspondence principle makes it a very promising idea, at least with respect to some domains of science. But there remains much to be clarified for the intuition to turn into credibility. To begin with the most obvious, the structuralist needs to ensure that the kind of continuity in focus really has to do with the *realist* rather than empiricist content. For surely the instrumentalist or the empiricist is also bound to find a level of continuity in the mathematical structures of a theory—namely those structures that encode the theory's empirical content. (Bueno, 1999) The realist claim is meant go further, of course, to declare a structural correspondence in the relevant theoretical content. Hence, Poincaré (as quoted by Worrall, 1989) asserts that

The differential equations [in Fresnel's theory] are always true [that is, they are carried over into Maxwell's theory], they may always be integrated by the same methods and the results of this integration still preserve their value.

It cannot be said that this is reducing physical theories to practical recipes; these equations express relations, and if the equations remain true, it is because the relations preserve their reality. They teach us now, as they did then, that there is such and such a relation between this thing and that; only the something which we then called *motion*, we now call *electric [displacement] current*. But these are merely the names of the images we substituted for the real objects which Nature will hide for ever from our eyes. The true relations between these real objects are the only reality we can attain... (1906)

This is undoubtedly a beautiful piece of rhetoric and quite suggestive too, but what does it amount to in practice? To really see what ‘the true relations’ of Fresnel’s theory are and how they compare with Maxwell’s theory, we need to engage in some history of science (§8). Worrall (1989, 1994) simply lists Fresnel’s equations for the amplitudes of reflected and refracted polarized light, to point out that they are truly identical to those resulting from Maxwell’s theory. But this is far too simplistic. The all important correspondence principle as characterised by Post emphasises that we should be able to *explain* the success of the predecessor theory from the vantage point of the successor theory:

Roughly speaking, this is the requirement that any acceptable new theory *L* should *account for* its predecessor *S* by ‘degenerating’ into that theory under those conditions under which *S* has been well confirmed by tests. (1971: 228, my italics)

This surely demands more than pointing out the fact that the equations that the two theories ultimately spit out—the equations that are used to test the theory against the experiment—are equivalent or stand in some limit-correspondence. What it demands, rather, is that we can account for the *derivation* of Fresnel’s equation in terms of Maxwell’s theory. For not only is there much to Fresnel’s theorising besides ‘the Fresnel equations’ which represent the very end result of theorising, but we also recall that the plausibility of the realist image, structural or otherwise, comes in part from fulfilling the intuition that success of a theory is connected to its approximate truth in a ‘non-miraculous’ fashion. This means that we should really be considering the relationship between the mathematical *derivations* by which the corresponding equations are arrived at in the first place. I will return to this in the next Part.

Another point to press the structuralist on concerns the sense in which one structure can be said to ‘approximate’ another. By merely appealing to the general correspondence principle this is left open—too open one might say. The worry is that without a precise sense in which one structure corresponds to another we end up finding mathematical continuity where we want it. Even in the cases of intuitively appealing limit-correspondence we often have

grave mathematical discontinuities that mark the theoretical revolution, as Redhead reminds us:

Consider the case of classical neo-Newtonian spacetime being replaced by the Minkowski spacetime of special relativity. We can consider a family of structures $\{S_C\}$ corresponding to varying the velocity of light c . For all finite c we can argue that the structure is stable with respect to changing c , but at $c = \infty$ there is a qualitative singularity in the sense that the metric of spacetime becomes singular in this limit. The existence of qualitative singularities of this type is also apparent in the case of the family of quantum mechanical structures indexed by a variable Planck's constant h . (2001: 346)

Such 'discontinuities' in the evolution of theoretical structures can perhaps be dismissed on the grounds that they are immaterial to the explanation of the success of the antecedent theory from the later perspective, but surely such a claim needs to be made on a case-by-case basis and only after carefully scrutinising the nature of the particular structural continuity in question.

It is obvious that declarations of structural continuity cannot solely refer to the theories' equations, for we must (of course) somehow express the fact that the theories have the same subject matter. Formally equivalent equations are used for various purposes in different domains of physics and any application of the correspondence principle needs to say something about how the equations in question are comparable apart from the shared logico-syntactic form. In some cases we can relate the terms of the equations to the same observable phenomenon (e.g. the Fresnel-Maxwell case), but we should ask whether there is a principled and preferred way of making the comparison. Is there a principled (logical) way of teasing out the structural content? In particular, do theories have to be regimented in some way for the comparison to be justified? For example, within the *syntactic-axiomatic framework* the prima facie possibility of using Ramsey sentences arises: these leave only formal structure to supplement the content expressible with the terms left outside of the Ramsey-elimination (as will be explicated in the next chapter). But arguably there are various reasons for preferring the alternative *semantic framework* of theories, and it turns out that this framework lends itself to a very different structuralist reading. I will return to these issues below (§5.3), after critically considering a very different motivation for structural realism.

5.2 Ontological Motivations

The thesis of epistemic structural realism is in serious need of sharpening as it stands. Furthermore, it will be argued in the next Part that the central case study cited by Worrall and others in support of structural realism actually turns out to support realism of a somewhat different kind. Nevertheless, I consider the epistemological motivation for a structural realist image to be a valid one; the challenge of first making the structuralist proposal more precise and then comparing it to various instances of historical theory change is surely a worthwhile endeavour. But some consider the above epistemological reasoning to be only part of the driving force behind structural realism: allegedly our best understanding (or lack of it!) of modern physics is on a par as an impetus for structural realism. And allegedly structural realism, properly construed, is not a purely epistemological thesis, but metaphysical! We should now try to make sense of this proposition.³

5.2.1 Metaphysical Underdetermination.

James Ladyman has asked about structural realism: ‘is it metaphysics or epistemology?’ (1998: 410) As explicated above the answer seems clear: it is epistemology. There is, however, an interesting argument that might at first seem to lead to a different conclusion.

Consider a ‘standard’ (non-structural) realist interpretation of quantum mechanics, for example. Setting aside momentarily all the problems having to do with providing a realist interpretation of the collapse of the wave function to begin with, the realist should presumably say of this most successful mature theory that it is probably approximately true in its claims about the unobservable world. So the quantum particles and fields, for example, are approximately like the theory tells us they are. But what does the theory tell us, *exactly*? According to our best understanding of the quantum theory these particles can just as well be *individuals* (‘cheese’) or *non-individuals*

³Actually there are two strands of structuralism in the recent philosophy of physics. One directly concerns the realist image, whilst the other belongs to philosophy of physics proper. Regarding the latter I do not wish to consider or criticise structuralist interpretations of physics per se, but only highlight the illegitimate inference often drawn to structural *scientific* realism (§5.2.2).

(‘chalk’), this metaphysical nature of the particles being underdetermined by the theory. Both interpretations of the physics are equally compatible with the phenomena as well as the formalism. (French 1989, 1998; Huggett 1997; French and Rickles 2003) So the standard realist is arguably in a pickle: she wants to say that the nature of the quantum particles is as the theory says it is, but the theory doesn’t say what it is.

We need to recognise the failure of our best theories to determine even the most fundamental ontological characteristic of the purported entities they feature. It is an *ersatz* form of realism that recommends belief in the existence of entities that have such ambiguous metaphysical status. What is required is a shift to a different ontological basis altogether, one for which questions of individuality simply do not arise. (Ladyman, 1998: 419–420)

Let us assume that the premise of *metaphysical underdetermination* holds at least with respect to some entities featured in our best physical theories.⁴ How big a blow is this for the standard realist? And what is the appropriate response? In order to avoid begging the question here, I follow Ladyman & French in taking *standard realism* to have a metaphysical dimension.⁵ Given what I have said of the project of painting and defending a realist image (§4.1), it should be clear that *prima facie* I do not take such metaphysical dimension to be a necessary part of a realist position. What I aim to show now is that the move from standard realism to ontic structural realism is unnecessarily radical, and not supported by the premise of metaphysical underdetermination.

The accusation against the standard realist is that her commitment to quantum particles and fields referred to in our best theories is deflated if she

⁴Ladyman (1998), French & Ladyman (2003) and French & Rickles (2003) defend this premise particularly for quantum particles and quantum fields, and tentatively point towards the nature of spacetime. Pooley (2006) dissents, especially regarding the underdetermined status of spacetime points. See also Redhead & Teller (1992) and Saunders (2003b) for criticism of the underdetermination thesis, and French & Krause (2006) for further defence.

⁵What exactly standard realism amounts to is unclear, but at least the ability to spell out our realist commitments in terms of ‘fundamental metaphysical categories’ is assumed. It is immaterial to my argument whether Psillos (1999), for example, represents standard realism thus characterised, as Ladyman & French read him.

cannot specify ‘the most fundamental metaphysical categories’ exemplified by the referents. Hence it is an ‘ersatz’ commitment that does not do any work for the realist, but merely represents an extraneous metaphysical image of the microworld given in terms of categories applicable to the macroworld. What *does* the work for the realist is *structure*, metaphysically construed. I am afraid, however, that this line of thought embodies a gross misrepresentation of the aims and the basic character of the epistemic realist position. Several points can be made in this regard. First of all, it is not the case that the question of realism is an all or nothing affair regarding our epistemic powers. Labeling standard realism ‘ersatz’ may be appropriate in exposing the superfluousness of a certain level metaphysical commitments. But moving directly from *standard* to *structural* is a non sequitur conclusion. Secondly, to make sense of the impact of quantum mechanics in realist terms we do not need to defend realism about the metaphysical nature of quantum particles, say. Thirdly, there is something fishy about the idea of avoiding metaphysical underdetermination by promising to provide *yet another* metaphysical framework of structures primitively understood. Let me elaborate.

French and Ladyman press the standard realist on the nature of quantum particles:

[T]he (standard) realist is unable to give a *full answer* to [the question:] ‘what is a quantum object?’, where a ‘full’ answer will involve the metaphysical nature explicated in terms of such fundamental categories as individuality, identity, etc. Van Fraassen rightly sees this as a challenge to standard realism (and it is regrettable that the standard realist has not seen fit to respond) expressing his conclusion as a waving ‘good-bye to metaphysics’^X (1991, 480–482), leaving the field clear for constructive empiricism. (2003: 36, my italics)

To demand a ‘full’ answer is to demand too much. Van Fraassen sees the kind of metaphysical underdetermination at issue to set a challenge for metaphysics, not epistemic realism per se. Underdetermination considerations in general do motivate van Fraassen’s empiricism, no doubt, but various degrees of epistemic confidence about the results of our ampliative practices can be had without giving up van Fraassen’s distaste for wholesale metaphysics. In a footnote (marked by *X* in the quote above) French & Ladyman insist that

‘if the realist refuses to be drawn on the metaphysics at least at the level of individuality versus non-individuality then how are we supposed to make sense of the impact of quantum mechanics?’ (*ibid.*, 50) But the question is ambiguous: there are two separate explanatory endeavours at stake for the realist.

- (1) Defend realism by explaining the success of a theory by its *approximate* truth.
- (2) Explain what the world could be like to make the theory true *simpliciter*.

The latter endeavour answers to the question of what the world *could be* like according to our theory, whilst the first answers to the question of what the world *must be* like according to our theory in order for the success of science (and that theory in particular) not to appear ‘miraculous’. Only the latter endeavour, providing a full-blown metaphysical interpretation of the theory, is affected by the professed metaphysical underdetermination at hand. The realist image—not being ultra-optimistic about our current science (cf. §4.1)—can avoid the force of the underdetermination by appealing to approximate truth. This notion (to be properly discussed in chapter §7) can be analysed in terms of theoretical *properties* responsible for the successful derivations in science. Knowledge of these properties is independent of the knowledge of the fundamental categories relevant to the explanandum (2) above.

What, then, *can* a realist claim to know of the quantum world? To properly answer this question would require having a close look at the various intricate issues in quantum mechanics, not least the outstanding measurement problem.⁶ But I would tentatively assert the following, regarding something like the prediction of the lamb shift and the anomalous magnetic moment of the electron from QED, for example: the crucial properties at play in these derivations can be understood and attributed to quantum fields at the level

⁶Saunders (2003c) takes the measurement problem to be one of the chief motivations for structural realism. This is yet another structural realist claim, based on an interpretation of the theory of decoherence, to be regarded as distinct from the metaphysical underdetermination argument under consideration.

of epistemic realism without answering the *ultimate* metaphysical question ‘What is a quantum field?’⁷

Consider, by way of analogy, a feasible metaphysical underdetermination vis-à-vis the nature of spacetime. The realist wants to explain the successfulness of the general theory of relativity by claiming it to have correctly identified the curvature of spacetime as the source of gravitational phenomena. Explaining the theoretical accommodation of the precession of the Mercury perihelion in these terms is independent of the metaphysical question of whether the spacetime points of the substantivalist interpretation of GTR are to be understood *haecceitistically* or *anti-haecceitistically*. In both of these metaphysical pictures the theory is true about the crucial unobservable features of the world, so that the concepts of curvature and geodesic, for example, apply to properties of substantival spacetime.⁸ This kind of metaphysical underdetermination just considered is quite different from the more old-fashioned empirical underdetermination that could take place in the spacetime context: one theory having a curved spacetime and the other having extra forces in its ontology. Then we would really not know what to believe in. And, as already emphasised in connection of the empirical underdetermination problem (§4.2), *even if* the metaphysical underdetermination prevented us from getting to standard realism in some domains or at some levels of enquiry, in as far as this can be regarded as an idiosyncratic rather than a universal limitation it does not make the standard realist image

⁷Chakravartty (2004) has likened the metaphysical underdetermination of individuality and non-individuality to the kind of ‘metaphysical underdetermination’ we face at the level of everyday objects: are tables and chairs ultimately just bundles of properties, or are they substances instantiating universals. French & Ladyman (2003) insist that the two cases differ since ‘in the case of unobservables the content of belief in them is exhausted by their theoretical description—if that underdetermines their metaphysical nature then our belief is empty’ (51). I acknowledge that there is a difference between the cases but would also insist that there is *much* more to the content of belief of electrons, say, besides their metaphysical natures as (non-)individuals or ‘nodes in a structure’. The realist belief is not empty, but half-full!

⁸The radical structuralist may respond by claiming that we don’t *really* understand how these concepts apply to substantival spacetime unless we can explicate in non-ambiguous terms the fundamental nature of GTR-spacetime. But whatever the status of such metaphysical understanding of GTR is (regarding the explanandum (2) above), surely what matters for the realist image is the hope that those concepts and properties doing the work in the GTR will correspond to the theoretical elements in the future theory of quantum gravity, describing the fundamental nature of *our* spacetime.

unappealing *in toto*.

Coming back to the two explananda above, it seems that even at the level of the second explanatory endeavour the underdetermination does not fully motivate the truly radical step to *ontic structural realism*, regarded as ‘offering a reconceptualisation of ontology, at the most basic metaphysical level, which effects a shift from objects to structures’ (*ibid.*, 37). Such a *metaphysical* project is in itself fully legitimate, of course, but cannot in my view gain any extra impetus from the metaphysical underdetermination. An ontological structuralist conclusion (regarding (2)) could perhaps be argued for by saying that structuralist metaphysics provides the only way to make sense of the notion of objecthood at the level of quantum particles (Saunders 2003a, 2003b), but this is not the claim presently evaluated. Indeed, such a claim directly contradicts the underdetermination premise which is conditional on both horns being intelligible *bona fide* possibilities. If anything, it seems that the structuralist proposal only makes matters worse, for with such an alternative structuralist ontology available there would be three instead of two to choose from!⁹ The choice between these would presumably be done on the grounds of general metaphysical preferences. And in any case, if *the realist image* needs truncating due to some inescapable metaphysical underdetermination, it is *epistemic* structural realism that is in the offing in the first place, and even that only as regards those specific areas of enquiry that are underdetermined. Standard realism about molecules and atoms and the cell mitochondria is not threatened by underdetermination at the level presently discussed.¹⁰

I conclude that the motivation gained from the metaphysical underdetermination for structural realism, and for ontological structural realism in particular, is highly problematic. I will next briefly look at an oblique line of enquiry that is sometimes (mistakenly) taken to provide further grounds for taking ontological structural realism to supplant epistemic structural realism.

⁹It has been suggested that the individuals and non-individuals packages could be viewed as different representations of the common ‘structuralist core’ but this intuition must be substantiated in order to show how the underdetermined options go over and above the common core, instead of just being metaphysical alternatives.

¹⁰The radical ontic structural realist aims to supply a new metaphysical framework which accounts for the ‘emergence’ of such unobservable entities. This should be regarded as a different project from that of painting a realist image of science.

5.2.2 Structuralism in physics vs. epistemology.

I now want to argue in more general terms for a distinction that should be made between two levels of structuralist philosophy which are often run together in a synergistic fashion.¹¹

One family of broadly structuralist ideas belongs to the philosophy of physics proper: the unifying theme is the conviction that the ontology of physics (at some level) is best conceived in structural terms. This line of thought is well represented in the history of philosophy by the likes of Cassirer and Eddington, for example, as a way of philosophically refining the worldviews imposed upon us by quantum mechanics and the general theory of relativity. (French 2003, French & Rickles 2006) *Very* broadly speaking this movement can be characterised as the attempt to shift one's ontology away from objects, as traditionally conceived, and towards structures relationally understood. The historical as well as the contemporary literature on structuralism in the philosophy of physics is by and large spurred on by the central role of fundamental symmetries exhibited by our best physical theories: the diffeomorphism invariance of the general theory of relativity, permutation symmetries in quantum mechanics, gauge symmetries of gauge field theories, and so on. Very crudely put, these symmetries can be understood in a sense to 'relationally define' the objects that are invariant under these symmetries, and hence are ontologically prior to the objects in some sense.

Another family of broadly structuralist ideas belongs to the epistemology, and concern the question of what we can claim to know of the (mind independent) world. Again, there are eminent historical figures to draw on—such as Russell (1927), fighting against phenomenalism about the external world—but in the contemporary context (of the No-Miracles Intuition and the Pessimistic Induction) the epistemological motivation, as already outlined above (§5.1), boils down to something quite specific. It is the attempt to craft a plausible image of science, motivated by the intuition behind the No-Miracles Argument and not refuted by the history of science. Once again,

¹¹It is not always easy to prise apart the different motivations running in parallel, but in my view an illegitimately close connection between different structuralist motivations is implied in Ladyman (1998), French & Ladyman (2003), Saunders (2003b), Lyre (2004), French & Rickles (2006), and Dorato & Pauri (2006), for example.

the idea simply is, crudely put, that theories by and large get the ‘structure’ right but often say falsehoods about the rest.

On the face of it, it is not easy to say how exactly the above structuralist project in the metaphysics of physics can interact with this latter idea. One might at first think that if the preferred ontology of physics is at the bottom structural, so that one is an ontological structural realist at the level of philosophy of physics, then one must also be a structuralist with respect to one’s scientific image, since all theoretical truths are ultimately truths about structure. But the connection cannot be this straightforward, for reasons already touched upon. After all, the structuralist ontology is inspired by metaphysical questions regarding a *literal* reading of our best theories—questions such as: what are the spacetime points quantified over in GTR like; how to understand the nature of quantum particles in the face of the permutation symmetry, or the gauge symmetry behind the Bohm-Aharonov effect—whilst the epistemological humility of the realist image is based on the belief that our theories may only be *approximately* true. Therefore it depends on the notion of approximate truth that the realist adopts whether or not the literal reading of our present theories has any input on the realist’s epistemic commitments. For example, it might be part of the realist image that there really is a curved spacetime and free particles move along the shortest paths as mathematically represented by the geodesics on a manifold—i.e. the theoretical terms ‘curvature of spacetime’ and ‘shortest path’ do refer—irrespective of whether the most fundamental spacetime ontology consists of dimensionless points or of something completely different. GTR might be a true representation of the curvature properties of spacetime whilst being a false representation of its ‘fine structure’. Indeed, being a classical (non-quantised) theory this is most probably the case.

This example is enough to sever the intimate link between ontological and epistemological structuralism suggested above. Structuralism in metaphysics might be appropriate for an interpretation of some theory T , but if T may be strictly speaking false it is not clear what epistemological lessons we should draw from it. And we may have a good reason to believe that a theory is strictly speaking false; for GTR it is the lack of a quantum field theoretical aspect that vindicates this. (cf. French & Rickles 2006) But how about the quantum field theory itself? Do we have any warrant for thinking that

this theory is only approximately true, and if not, what are the epistemological consequences of the (conjectured) structuralist metaphysics of QFT? Is it not enough to motivate epistemic structuralism that there is the *possibility* of some strictly speaking true theory T_S to have a preferred structuralist ontology? Surely the realist should have the resources necessary to express what T_S tells about the world? This line of thought misrepresents the realist project once again, however. The realist only needs the resources required to capture those aspects of the world that were latched onto by the scientific practice in producing the successes of T_S . I believe that those features can be described independently of the underlying ‘fundamental metaphysical categories’, as the haecceitist vs. anti-haecceitist example above is meant to indicate. Strangely enough, Ladyman and French also repeatedly stress that the preferred ontology cannot be drawn directly from physics: ‘we cannot infer the appropriate metaphysics for describing the world from the physics itself.’ (da Costa & French 2003: 188)¹² So what becomes of the idea that the realist needs to be a structural realist to latch onto the world as described in physics?

Finally, I want to dismiss the idea (as expressed in Lyre (2004), for example) that a structuralist ontology at the level of physics should not only lead us to epistemic structural realism, but surpass epistemic in favour of ontic structural realism altogether! This is a grand non sequitur: the epistemic structuralist à la Worrall (1989), for example, would not consider one or another interpretation of the ‘literal reading’ of quantum mechanics to be of significance to his realism, since he is only committed to the claim that (a) the relevant theoretical structure suitably approximates the structure of any future theory of quantum phenomena, and (b) we can ultimately make sense of the success of quantum theory in his epistemic structuralist terms. I regard this claim as indefensible—cf. chapter §7—but its shortcomings are epistemic, not ontological.

The failure to properly distinguish between the ontological and epistemological levels of structuralist endeavour has landed the realism discussion in this context in a muddle. Witness Lyre, for example, who takes struc-

¹²They underscore this in connection of the metaphysical underdetermination, of course, to motivate the alternative structuralist ontology. As far as I can see, the alluded to ontological notion of structure forms just another fundamental metaphysical category.

tural realism to be a monolithic position supported by arguments from both philosophy of physics and philosophy of science.

[A] philosophical view such as structural realism gains by far more credence if supported by arguments from science directly than by mere indirect and notoriously debatable considerations of the philosophy of science. This is the difference in style between the Worrall-type of arguments in favor of structural realism and the French-type of arguments—on the basis of the ontology of quantum theory—or the Stachel-type—on the basis of general relativity. (Lyre 2004: 621)

These two sets of arguments simply do not have the same objective.

5.3 Meta-Scientific Frameworks

In addition to promoting the radical shift from an epistemic to a metaphysical conception of structural realism, James Ladyman and Steven French have forcefully advocated a thoroughgoing shift in the meta-scientific framework in which to cast the position to begin with. We should now consider the merits and liabilities of thus committing oneself to one or another of the frameworks on offer.

In rough outline, the argument for moving (in the present context) from the *syntactic* (or ‘received’) to the *semantic* (or ‘model-theoretic’) view of theories goes as follows.

Historically the term ‘structural realism’ was coined by Grover Maxwell in the late 1960s and early 1970s, for a philosophical view that takes the cognitive content of a scientific theory to be contained in the Ramsey-sentence of a theory expressed as a partially interpreted system of axioms in the first-order logic. (1966, 1970*a*, 1970*b*) Maxwell’s motivation originates from Russell’s writings on what could also be called epistemic structural realism (about the external world) but, as propounded by Russell and Maxwell, such a view has little to do with Worrall’s (1989) intuition of a structural realist image as portrayed above. (More on this in the next chapter, §6.4) Nevertheless, it is reasonable to ask whether Ramsey-sentences could provide a natural way of formally spelling out Worrall’s somewhat vague suggestion about the structural content and continuity, especially since Worrall himself has later opted

for this route. (Worrall & Zahar, 2001) But arguably this is not an easy route to take. Russell's formulation of epistemological structural realism (about the external world) in *The Analysis of Matter* (1927) immediately faced a challenge articulated by the Cambridge mathematician M.H.A. Newman in his review of Russell's book. (Newman, 1928) This challenge—dubbed 'Newman's Problem' in the contemporary literature—was later resuscitated against the view advanced by Maxwell, and later still formulated as a problem for the realist view espoused by (the later) Worrall. (Demopoulos & Friedman, 1985; Ketland, 2004) This problem is allegedly an insurmountable matter of logic, a 'theorem one cannot argue with'.¹³

Implicit in this 'Ramseyfication' approach to structural realism is a construal of theories in the *syntactic* framework: theories are represented as partially interpreted axiom systems the axioms of which can be conjoined to a single sentence over which the Ramsey-sentence of the theory is to be taken.¹⁴ By contrast, the alternative *semantic* view of theories takes *models*, rather than the systems of axioms satisfied by those models, to be the primary representative elements in the philosophical depiction of the theory-theory and the theory-data relationships. (cf. for example, French & Ladyman, 1999) The considerable representative power of model-theory can be further supplemented by introducing richer logical machineries such as that of 'partial structures', and arguably the representation of science thus achieved is more naturalistic and unifying with respect to various kinds of scientific modelling and theoretical representations of the world. (da Costa & French, 2003) Hence the semantic approach is arguably preferable on general grounds, at least with respect to some philosophical ends and purposes.

Ladyman and French (*ibid.*) take the semantic view to provide the preferred framework in which to capture the structural realist image. The argument for this preference attempts to get leverage from (i) Newman's problem, (ii) the general superiority and naturalness of the semantic approach, and also (iii) from the inherent 'structuralism' of the semantic view:

The alternative 'semantic' or 'model-theoretic' approach to theories,

¹³Jeffrey Ketland famously declared this whilst visiting at the University of Leeds during the term 2003-4.

¹⁴I will give a more formal explication of Ramseyfication in the next chapter (§6.1).

which is to be preferred on independent grounds, is particularly appropriate for the structural realist. This is because the semantic approach itself contains an emphasis on *structures*. (Ladyman, 1998: 416)

Hence, the argument goes, it is only natural to locate the structural realist gambit fully within the confines of the semantic view.

I think there are several reasons to be unhappy with the above story line. Firstly, the much discussed Newman's problem is not a problem for the structural realist working in the framework of the syntactic view, as long as a reasonable semantics of theoretical terms is adopted. Secondly, the structuralism inherent in the semantic view is not in itself by any means tantamount to structural realism *qua* a realist image; more needs to be said about the notion of approximate truth, in particular. And, thirdly, what little has been said about the notion of approximation applicable to consecutive theories represented in the model-theoretic/structural terms is not enough to take advantage of the claimed general preferability of the semantic view. I will now explain the last two points, before analysing in detail and rebutting the Newman's problem -attack on Ramseyfying realism, in the next chapter (§6.3).

Let us first consider the claim that the emphasis on structures in the semantic view renders it a particularly appropriate framework for structural realism. Perhaps there is something to this claim—I'm personally still undecided about it—but whatever it might be it cannot be anything as straightforward as the quote above makes it seem. Recall the discussion above of the epistemic motivation for structural realism (§5.1). The basic idea there was, in very general terms, that theoretical content might bifurcate in a principled manner so as to leave a level of abstracted structure intact in theory change, whilst being accompanied by radical shifts in the 'nature' over which the abstraction is taken. Hence, structure in *this* sense involves abstraction and the loss of more detailed theoretical content. By contrast, the notion of structure at play in the semantic view does *not* involve such abstraction and bifurcation. Rather, the inherent structuralism of the semantic view has to do with the way that theories are analysed to represent and latch onto the world, not so much through language and linguistic correspondence rules *per se*, but through relational hierarchies between higher and lower order theo-

retical models and, ultimately, models of the empirical data. This sense of structure is fully applicable to an analysandum theory that is taken to be true *simpliciter*, whilst the sense of structure relevant to Worrall's structuralist intuition is not: the latter is a matter of a theory being *approximately* truth, a notion which completely collapses for fully true theories. And it cannot be the case that the representation of a fully true theory in the semantic approach makes it weaker content wise—speaking 'only of structure'—than it would be in the syntactic approach, say, for then the semantic approach would not offer a complete representation of science.

So the equation here cannot simply be 'Semantic Approach + Realism = Structural Realism'. Indeed, this would make any advocate of the semantic approach, if a realist, also a structural realist by fiat! I do not want to attribute to anyone such a naïve equation of the two sides, but I think we can justifiably press the question of what the connection here is exactly meant to be. What exactly would structural realism in the semantic view amount to?

The crux of the matter can be found in the way the semantic approach can be employed to represent the 'horizontal' theory-theory relations, in addition to the 'vertical' theory-phenomena relations. The semantic approach offers a general abstract representational framework for theory comparison that operates in extra-linguistic terms via the mathematics of model-theory. This framework, the idea is, allows us to discount the kinds of discontinuities that occur at the 'surface' linguistic level (referential variance). Structural-realism-in-the-semantic-approach then, in my view, amounts to the following claim. By adopting the appropriate meta-scientific representational stance, we can always find continuity across theory change, and hence there are no radical theory-shifts that would falsify the realist commitment to the approximate truth of a theory understood as the truth content of the theory expressible in the pertinent model-theoretic terms (of e.g. partial structures—cf. da Costa & French, 2003). This claim embodies a form of *structural* realism, since the crucial form of continuity, and the corresponding notion of approximate truth, are best understood in holistic model-theoretic, structural (as opposed to linguistic) terms.

This is a nice starting point for a realist position, but it is *only* a starting point. What is missing is an account of how this general representational framework manages to latch onto the kind of continuity in the theoretical con-

tent that is significant vis-à-vis the realist project. What is it, in particular, that allows us to discount the discontinuities marked by referential variance as completely unimportant? The challenge issued for the realist by the Pessimistic Induction is not one of finding theoretical continuity simpliciter, but one of finding continuity of *the right kind*. The worry here is that theoretical correspondence becomes too cheap in the semantic approach, and this worry is heightened by the very flexibility of the framework of partial structures. Whereas the semantic approach can provide the formal underpinnings for the notion of approximate truth, an informal analysis of the notion is badly needed to render the notion fit for the realist use. In my view such analysis is philosophically prior to any project of formalisation, and without such analysis the structural realist claim above is rather deflated: regarding our present theories, for example, the realist is committed to their approximate truth only in the sense that there will be a mathematico-structural relationship to any subsequent future theory that (i) can be formally represented in the semantic approach, and (ii) coheres suitably with the predecessor theories.

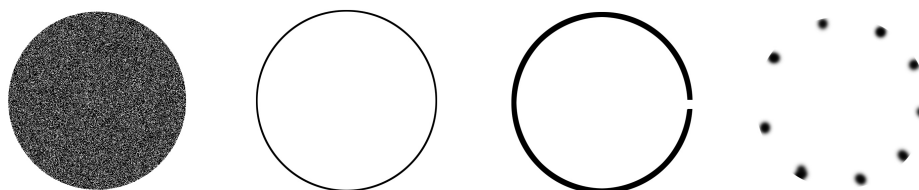


Figure 5.1: Radical discontinuity in some aspects is compatible with continuity in others. But which is important?

The notion of structure gets very differently interpreted when developed in the syntactic and the semantic view. The Ramseyfication approach attempts to provide a recipe for weakening theoretical content of a theory, to leave only ‘structure’. This yields ‘structuralism’ about a *particular* theory, but nothing has been said of the way the Ramsey-sentences of successive theories should be compared. By contrast, the semantic approach in the hands of the structural realists has yielded a formal framework for comparing subsequent theories, but it says very little of the structural content (understood

as a restriction on the full content) of some particular theory. The advocates of this latter framework have appealed to the inherently structural character of this way of representing science, but the connection is far from straightforward. Perhaps both of these meta-scientific frameworks have in principle the resources to capture an interesting sense of structural continuity. Staying closer to scientific practice is a big bonus for the semantic view, but there is something dissatisfying with the arguments relying on this formal approach, too. In both cases the formal rigour is gained by moving away from the actual science, and something relevant is lost in the idealisation. In the Ramseyfication approach this is particularly unencouraging: we can't in practice even write the Ramsey sentence for any truly interesting theory! This makes the whole approach seem slightly fruitless and quixotic in the least, from the point of view of the realist project as delineated in chapter §4. In the semantic approach we seem to be able to nicely carry out various instances of theory comparison, but the formal results have little realist interest purely on their own without an accompanying philosophical analysis of the concepts relevant to the notion of approximate truth and realist explanation of success.¹⁵

I will re-approach the all-important topic of approximate truth from a more informal angle in the next Part (chapter §7), after investing the next chapter on clarifying the vices and virtues of the Ramseyfication approach. Despite the disparaging remarks above, the various issues involved have a wealth of intrinsic philosophical interest.

¹⁵It should be emphasised that I have nothing against formal representations of science as such. The point here merely is to lay stress on the conceptual analysis that I regard to be prior to formalisation of concepts. Finding continuity in a representation does not help unless we already know exactly what kind of continuity we are looking for.

Structural Realism and Ramseyfication

“A delicate question concerns the construction of the Ramsey sentence, since our interpreted language refers to ‘mixed’ or ‘bridge’ relations.”

Jeffrey Ketland, *‘Empirical Adequacy and Ramsification’*

It has been suggested in the context of the traditional, axiomatic-syntactic view of theories that the theoretical content in structural realism could be captured by the logical procedure of Ramsey-elimination (or Ramseyfication, for short). But prima facie it is not by any means obvious how Ramseyfication is to be harnessed to structural realist ends. I shall begin by looking at this question (§6.2). Against the Ramseyfying structuralist proposal there is the so-called Newman’s objection. This is rebutted in section §6.3, completed by a discussion of a finessed model-theoretical challenge. After sketching a preliminary answer to this finessed challenge, we shall look at some exegetical issues to be clarified, with regard to the development of this whole debate. How did Ramseyfication get affiliated with epistemic structural realism in the first place? What is the connection of the debate between Russell and Newman to that of Worrall and Ketland? These questions will be discussed in the final section §6.4.

6.1 Defining Ramseyfication

To apply the Ramseyfication procedure we need to have our theory logically regimented as a finitely axiomatised theory θ . A *Ramsey sentence* $\mathfrak{R}(\theta)$ of θ is then obtained by first replacing some of the predicates in θ by second-order (predicate) variables, and then prefixing the resulting open formula by a string of second-order existential quantifiers, one for each variable. The resulting sentence is relative to the set of predicates replaced by the existentially quantified variables, or ‘Ramseyfied over’.

$$\theta(A_1, \dots, A_n; B_1, \dots, B_m) \Rightarrow (\exists X_1) \dots (\exists X_n) \theta(X_1, \dots, X_n; B_1, \dots, B_m)$$

Frank Plumpton Ramsey (1929) operated with the logical positivist construal of scientific theories, with distinct observational and theoretical vocabularies related via a dictionary of correspondence rules. At the time it was of interest to consider Ramseyfication over the whole of the theoretical vocabulary, and much of the early literature on Ramsey sentences revolve around the question of the significance of the fact that $R(T)$ implies the same set of sentences in the purely observational vocabulary as does the original theory T . Nowadays we understand the semantics of scientific terms quite differently, and the notion of meaningful Ramseyfication has correspondingly shifted and is dependent on one’s views on the semantics of theoretical terms. The discussion below (§6.3) on Newman’s Problem will turn on this matter.

6.2 Ramseyfication and Structure?

Philosophers have made great use of Ramseyfication in answering questions of semantics. One line of thought originates from Carnap’s work on theoretical terms (1956), gets a golden realist coating in the hands of David Lewis (1970), and continues to provide useful machinery in the realist’s arsenal in the present day (Papineau 1996, Nola & Kroon 2001). It is beneficial for us to begin by considering this ‘received’ use of Ramseyfication, before moving on to evaluate the contemporary structuralist proposition.

Lewis’s (1970) proposal for defining theoretical terms is well-known. Any

n -tuple $\langle e_1, \dots, e_n \rangle$ of entities which satisfies $\mathfrak{R}(\theta)$ is called a *realisation sequence* of $\mathfrak{R}(\theta)$. The Ramsey sentence of θ says that θ has at least one realisation. Carnap (1966) proposed to implicitly define the theoretical terms $t_1 \dots t_n$ of θ by the analytical *Carnap sentence* $(\exists X_1) \dots (\exists X_n) \theta(X_1 \dots X_n) \rightarrow \theta(t_1 \dots t_n)$. Lewis (1970, 1972) suggested strengthening this scheme by the requirement that each theoretical term t_i gets defined as denoting the i^{th} component of the realisation sequence $\langle e_1, \dots, e_n \rangle$ of $\theta(x_1 \dots x_n)$ *only if* there is a *unique* such sequence, and denoting nothing otherwise.¹ Thus we end up with the *Lewis sentence* $(\exists! X_1) \dots (\exists! X_n) \theta(X_1 \dots X_n) \leftrightarrow \theta(t_1 \dots t_n)$ for the whole theory. All the theoretical content is contained in the antecedent—a Ramsey sentence with the uniqueness operators.² It should be emphasised that the Carnap-Lewis program is a *purely semantic* one, a kind of descriptive theory of reference for theoretical terms, defining a set of new terms by using some antecedently understood terms, some of which can denote relations between the new terms to be defined.

Lewis improved Carnap’s account in crucial respects. Implicitly defining theoretical terms via the Carnap sentence $(\exists X_1) \dots (\exists X_n) \theta(X_1 \dots X_n) \rightarrow \theta(t_1 \dots t_n)$ may be a good suggestion for understanding how the stipulating nature of the implicit definition is compatible with the contingent, synthetic observational content of the theory. But it leaves the following serious problems intact: (a) it does not guarantee that theoretical terms denote *unique* set of physical entities; (b) it does not guarantee that theoretical terms denote physical entities *at all*. Lewis’s account steers clear of these two problems by imposing the uniqueness requirement together with the rejection of the logical empiricist style understanding of the observational/theoretical term distinction.

The first problem above is entailed by the possibility of multiple realisation: there may be more than one realisation sequence for a Ramsey sentence. By simply imposing the uniqueness requirement it is ensured that the semantic roles of theoretical terms are unique, if these terms denote *at all*. Lewis takes it to be inherent in the scientific practice that a theorist proposing θ as a well-defined theory is implicitly asserting that θ is uniquely realised.

¹But see also Lewis (1994).

²The requirement of unique realisation marks the difference to the proposed ‘structuralist’ use of Ramseyfication, considered below.

Let us now consider the second problem, however. In this regard John Winnie (1967) has shown that—given a division of predicates into three non-overlapping classes of Observable, Mixed and Theoretical predicates³—the idea of implicitly defining theoretical terms along the lines of Carnap leads not only to the existence of multiple realisation within a given domain, but also to the possibility of construing theoretical entities as numbers. Therefore it may be asked on what basis Lewis thinks that (i) term introducing theories are *ever* uniquely realised in a domain of *physical* theoretical entities; and (ii) the uniqueness requirement could be extended to rule out realisation in a cardinality consistent domain of *mathematical* entities?

Lewis (1970) responds explicitly to (i), but in this response there lies an implicit answer to (ii) as well. The relevant section of the paper is worth quoting in full. (This is closely related to the model-theoretic arguments examined in the next section.)

Finally, I should say again that we are talking only about realizations that make T true under a fixed interpretation of all of its O -vocabulary. And this O -vocabulary may be as miscellaneous as you please; in practice it is likely to be very miscellaneous indeed. An O -term is *any* term, of any character, which we have already understood before the new theory T came along. It does not have to belong to an observational language. If anyone hopes to adapt my proposal to the task of interpreting theoretical terms using only an observational language—if there is such thing—I would not be at all surprised if he ran into trouble with multiple realizations. But his project and his troubles are not mine.

John Winnie has announced a proof that scientific theories cannot be uniquely realized. Though his proof is sound, it goes against nothing I want to say. . . . I am concerned only with realizations under a fixed interpretation of the O -vocabulary; whereas Winnie permits variation in the interpretation of certain O -terms from one realization to another, provided that the variation is confined to theoretical entities. For instance, he would permit variation in the extension of the O -predicate ‘— is bigger than —’ so long as the extension among

³Only a mixed predicate can have both observable and unobservable entities in its extension; cf. §6.3.1 for details.

observational entities remained fixed. Winnie's proof does not show that a theory is multiply realized in my sense unless the postulate of the theory is free of 'mixed' *O*-terms... (1970: 430)

The first paragraph of the quote stresses the significance of adopting an understanding of the semantic categories at work which radically differs from that of the logical empiricists. Lewis's well-known 'Old' term versus 'New' term dichotomy replaces the observational/theoretical term distinction with the defining/defined distinction. Thus the *O*-vocabulary, for Lewis, includes a miscellaneous collection of (quasi)-logical and mathematical predicates that we take to be antecedently understood prior to *any* scientific theory; examples of such predicates would be 'is individual', 'has a property', 'is between of', et cetera. Within these antecedently understood *O*-terms there are some which correspond to some of Winnie's *mixed predicates* applying to both observable and theoretical entities. By insisting on a *fixed interpretation* not only of Winnie's 'observational' predicates but of these mixed predicates as well, Lewis at least partially blocks the physical multiple realisation worry raised by Winnie: by discounting all the alternative interpretations which fail to keep the extensions of all the 'Old' predicates of a theory fixed, the mechanical procedure employed by Winnie to generate alternative realisations is no longer eligible.⁴ The existence of multiple realisations is now to be decided on a case-by-case basis; it depends on the structure of the theory and the role that mixed predicates play in it. All this is obviously compatible with Lewis, who only claims that given the nature of scientific theories 'we can reasonably hope that there is only one way in which [a theory] *is* realized', not that it would be impossible to have multiple realisations.

There is also an implicit answer in Lewis's reply to the second worry raised in the foregoing—the worry that even if we have a good reason to hope for a unique realisation in a domain of physical entities, and even if we define theoretical terms via such unique physical interpretation, we are

⁴Lewis also stresses that we need not assume 'that the language of T is an extensional language', so that 'among the O-terms there may be nonextensional operators, for instance "it is a law that —"; nonextensional connectives, for instance "— because —"; and so on' (*ibid.*, 80). Incidentally, John Winnie already hints at the possible role of causation in bringing in intensional content to save Ramseyfication from the unsatisfactory conclusion he had arrived at by purely extensional reasoning.

nevertheless able to give a true interpretation of the theory in a domain of mathematical entities. What then would become of the alleged truth of the theory, regarding the semantic roles of implicitly defined theoretical terms? Luckily the answer implicit in Lewis's construal of mixed predicates is rather obvious: no mathematical entity could belong to the extension of a scientific mixed predicate denoting a physical property or relation!

But despite all the advances Lewis's account makes to the 'received view' of theories, it can still be considered to be a problematic framework for the realist to adopt. In particular, the combination of the uniqueness requirement and the 'Old' vs. 'New' term distinction seems to lead to problems with theory change: if 'New' theoretical terms are simply defined at each theoretical stage by this scheme we automatically end up with problematic meaning variance.⁵ Furthermore, it is not yet clear whether the fixed interpretation of mixed predicates can actually promote the uniqueness of scientific theories to the extent optimistically hoped by Lewis.

An alternative is to drop the uniqueness requirement together with the Lewis sentence and focus on the content given by the Ramsey sentence *simpliciter*. This is what some structural realists have proposed to do. (Worrall & Zahar, 2001) Problems of referential variance presumably evaporate if we forget about the theoretical terms and reference altogether! However, we still have theory change encoded in the Ramsey sentences: logically regimented Fresnel's ether theory presumably looks sufficiently dissimilar to the likewise regimented Maxwell's electromagnetic theory for there to be a significant shift in the corresponding Ramseyfications.⁶ Hence Ramseyfication cannot all by itself offer a realist solution to the problem of theory change. What we need is a theory of approximate truth for Ramsey sentences. Such a theory should supplement Ramseyfication in two ways. It should explain, to begin with, how an *idealised* theory, say, can be false yet approximately true

⁵Lewis realises this problem, of course, and introduces the notion of 'near-realisation' to overcome it. Intuitive though this notion is, it is not enough for the realist project to state that this notion is 'hard to analyze, but easy to understand' (1972: 252). Providing an analysis of this notion would effectively amount to a theory of approximate truth for scientific theories.

⁶Exactly how big a shift? It is hard to say, given that we don't have these theories logically regimented in a way that would allow Ramseyfication, and the possibility of achieving this is far removed in practice!

by virtue of leaving out some complicating parameters, and how a theory can be approximately true by virtue of featuring a slightly off-the-mark numerical value of some constant of nature, for example. This is the ‘trivial’ bit. But more importantly it should provide a way of comparing inequivalent Ramsey sentences. This crucial, very much non-trivial component of the Ramseyfying structuralist account of theory change remains at the present completely undeveloped.

But why have the modern day structural realists such as Worrall adopted Ramseyfication in the first place?⁷ Given the decidedly non-structural, ‘received’ semantic use of this logical procedure, Ramseyfication does not seem like an obvious way to develop Worrall’s epistemic structural realist intuitions examined in the previous chapter (§5.1). Although the connection is by no means straightforward, we can motivate it through the following considerations. First of all, the fact that $\mathfrak{R}(\theta)$ can have the same content as θ about the observable world, but yet be strictly weaker in its theoretical content, gives rise to the possibility that $\mathfrak{R}(\theta)$ is (approximately) true even if θ is false (simpliciter). Now, whether or not there is an obvious sense of ‘theoretical structure’ which Ramseyfication naturally captures is a moot question here. What matters is that this way of driving a logical wedge between some theoretical content and the rest promises to give a rise to an interesting novel realist position. Secondly, perhaps it is actually not that far fetched to call the Ramseyfied content structural, given that Ramsey sentences *can be* employed to capture what is naturally considered to be multiply realisable structural content.⁸ (Shapiro, 1997: 106–108)

The motivation for Ramseyfying structuralism should be based on these kinds of general considerations, and be independent of both the particular thoughts advanced by Worrall along the revival of the Poincaréan rhetoric, as well as the connection—advertised by Maxwell, and to be scrutinised below—of Ramseyfication to Russellian epistemic structural realism. Although a great deal of work remains to be done to consolidate the position hinted

⁷There is an historical story to tell about this: cf. §6.4 below.

⁸There is a difference, it must be admitted, between claiming that Ramsey sentences *can* capture structural content and claiming that *all* Ramseyfication capture structural content. But perhaps the weaker claim is all that is required for motivating a piece of terminology.

at, I find the starting point worthy of close scrutiny. The formal rigour and intriguing connections to the idea of multiple realisability are auspicious, even if somewhat offset by the pragmatic difficulties in implementing the logical regimentation required to apply the position to real science. But there are those who claim—and this really seems to be the current consensus—that the Ramseyfication approach is a complete non-starter, due to certain model-theoretic considerations. It is these considerations that we shall now turn to.

6.3 Model-Theoretic Arguments

The basic model-theoretic argument against the Ramseyfying structural realist boils down to the following claim: it is a matter of logic that a Ramsey sentence $\mathfrak{R}(\theta)$ is true of the world if and only if $\mathfrak{R}(\theta)$ is both empirically adequate and true of the number of unobservable entities in the world. So what becomes of the idea that one's realist commitments with respect to a theory θ are fully captured by its Ramsey sentence? (Demopoulos & Friedman 1985, Ketland 2004)

Clearly such a weak commitment is realism in name only. In particular, we cannot explain the success of a past theory by referring merely to the cardinality of the unobservable world. But, luckily for the Ramseyfying realist, the logical considerations of the basic model-theoretic argument, although valid, are based on questionable premises.

To expose these premises it is worth looking in some detail at the recent formalisation of the basic argument (§6.3.1). The formal façade of this presentation hides some delicate philosophical assumptions that can, and should, be questioned (§6.3.2). In particular, the following three assumptions are exposed: (a) the structural realist must eliminate all predicates that can apply to unobservables; (b) quantification over properties is correctly formalised by a model theory which treats the domain of the second order quantifiers as full; (c) the scientific theories to be Ramseyfied are formulated in an extensional logical framework. Rejecting the assumption (a) allows us to undermine the basic model-theoretic challenge. But there is a further, 'finessed' model-theoretic challenge that purports to show that even if (a) is rejected, $\mathfrak{R}(\theta)$ can still be true too easily to carry substantial realist commitment (§6.3.3).

To find a way out, the Ramseyfying realist needs to incorporate content into her logical formalisation that goes beyond what can be expressed in a purely extensional second-order logic.⁹

6.3.1 Formalising the Challenge

If $\Re(\theta)$ were to be formed by Ramseyfying over *every* predicate in θ , then certainly triviality follows. For the truth of $\Re(\theta)$, conditional only on the cardinality of the intended domain S of the theory θ , now follows directly from the model-theoretic clauses for second order existential quantification. For there is a model \mathbf{M} of a consistent theory θ , and we can use any bijection between the domain of that model and the intended domain S to form an isomorphic model \mathbf{M}^* over S . Since isomorphic models make the same sentences true, the theory θ is true in \mathbf{M}^* , and hence $\Re(\theta)$ is true *simpliciter*.

This may be a troubling result for some structuralist positions—indeed, it can be viewed as a formalisation for Ramsey sentences of the very point that Newman (1928) made against Russell’s structuralism—but the Ramseyfying structural realist should clearly adopt a more conservative position. For the all-inclusive Ramseyfication above does not encode any empirical content whatsoever! The Ramseyfication of a theory is relative to the choice of predicates to be eliminated, and the task for the Ramseyfying realist is to divide the predicates of θ into those eliminated and those left intact. And she clearly wants to do that in a way that allows her to retain as much as possible of the uncontroversial and unproblematic content. In particular, she surely wants to retain all of the empirical content of the theory.

Jeffrey Ketland (2004) has proved that *one way* of thus making that division yields a triviality result much like Newman’s. Ketland formalises the proof in a two-sorted second-order language. The first-order variables are divided in two sorts: variables of the first sort range over observable entities, and variables of the second sort range over unobservable entities. Correspondingly, the predicates of a theory expressed in this language are divided

⁹The material of this section has borne out of the supervisory collaboration with Joseph Melia, and consequently owes much to Joseph’s considerable analytical faculties. See Melia & Saatsi (forthcoming). The exquisite presentation of Ketland’s proof in §6.3.1, in particular, is completely due to Melia. I take the full responsibility of any remaining errors, of course.

into *three* sorts: those predicates whose extension is drawn entirely from the domain of observable entities (O-predicates), those whose extension is drawn entirely from the domain of unobservable entities (T-predicates), and those whose extension is drawn from both domains (M(ixed)-predicates).

A theory θ expressed in this two-sorted language has models of the form

$$((D_1, D_2), O_i, M_i, T_i)$$

Here D_1 is the domain over which the variables of the first sort range ('observable entities'); D_2 is the domain over which the variables of the second sort range ('unobservable entities'). O_i is a sequence of subsets of D_1^n , i.e. the extension of the particular O-predicate in the model. Similarly, M_i and T_i are sequences of $(D_1 \cup D_2)^n$ and D_2^n , respectively.

We can take a theory to be *empirically adequate* if it has a model which contains all the appearances. Formally, a theory is empirically adequate if it has a model \mathbf{M} such that (a) the domain D_1 of \mathbf{M} is D_{Obs} , the set of observable objects, and (b) for all O-predicates $O(x_1 \dots x_n)$ and observable objects $a_1 \dots a_n$, $\langle a_1 \dots a_n \rangle \in \text{val}(O)$ if and only if it is true that $O(a_1 \dots a_n)$. We say that such models are themselves *empirically correct*. If a model has a theoretical domain D_2 of the same cardinality as the theoretical domain of the actual intended model D_T —the set of unobservable entities in the world—then we say that the model is *T-cardinality correct*.

Ketland has proved within this framework the following result: $R(\theta)$ is true if and only if θ has a model which is empirically correct and T-cardinality correct.

The proof is as follows. Suppose that θ is empirically adequate and T-cardinality correct. Then θ has a model $\mathbf{M} = ((D_1, D_2), O_i, M_i, T_i)$ which is both empirically correct and T-cardinality correct. By empirical correctness: (1) the observable domain of \mathbf{M} , D_1 , just is the set of all observable objects D_{Obs} , and (2) for each sequence of observable objects $\langle a_1 \dots a_n \rangle$ in D_1 , and for each n-place observational predicate O , $\langle a_1 \dots a_n \rangle \in O$ if and only if $a_1 \dots a_n$ really do stand in relation O . For \mathbf{M} to be T-cardinality correct is for D_2 to have the same cardinality as D_T —the set of unobservable entities. Thus there is a 1-1 function f from D_2 to D_T .

By hypothesis, $((D_1, D_2), O_i, M_i, T_i) \models \theta(O_i, M_i, T_i)$. Let g be a 1-1 func-

tion from $(D_1 \cup D_2)$ to $(D_{Obs} \cup D_T)$ which leaves every element of D_{Obs} as it is (this is possible as $D_1 = D_{Obs}$ in this case). Let $g(O_i)$, $g(M_i)$ and $g(T_i)$ be the obvious extension of g to n-tuples of $(D_{Obs} \cup D_T)$. Since g is 1-1 and onto, $\mathbf{M}^* = ((gD_1, gD_2), g(O_i), g(M_i), g(T_i))$ is isomorphic to \mathbf{M} . By the defining properties of g , $\mathbf{M}^* = ((D_{Obs}, D_T), O_i, g(M_i), g(T_i))$. Since \mathbf{M}^* is isomorphic to \mathbf{M} , $((D_{Obs}, D_T), O_i, g(M_i), g(T_i)) \models \theta(O_i, M_i, T_i)$. So, by the definition of truth in a model for the second order existential quantifiers, $((D_{Obs}, D_T), O_i, g(M_i)) \models \exists X_i \theta(O_i, M_i)$. Again, by the definition of truth in a model for the second order existential quantifiers, $((D_{Obs}, D_T), O_i) \models \exists X_i \exists X_j \theta(O_i)$. This final formula is none other than $R(\theta)$. Thus we have $((D_{Obs}, D_T), O_i) \models R(\theta)$.

Now, $((D_{Obs}, D_T), O_i)$ isn't just any old interpretation of $R(\theta)$ —it is the intended interpretation of $R(\theta)$. The domain of the first sorted variables is exactly the set of observable objects; the domain of the second sorted variables is exactly the set of unobservable objects—but these are simply the intended domains of such variables. Similarly, all predicates of the theory have their intended interpretation—each predicate O_i is assigned the set O_i which is the set of observational objects that satisfy O_i . But if $R(\theta)$ is true in the intended interpretation then $R(\theta)$ is true *simpliciter*, for truth and truth-in-the-intended-interpretation are, at the very least, co-extensive. QED

6.3.2 Ramseyfication and Theoretical Predicates

It was emphasised above that the end result of Ramseyfication is relative to the class of predicates thus eliminated. The challenge for the realist is to provide a principled division between the eliminable predicates and the rest. Ketland's proof demonstrates that *one way* of drawing the dividing line results in deflation of practically all theoretical content. But is there anything to recommend this particular dichotomy as the *only* such division? Indeed, is there anything to recommend this particular dichotomy as a reasonable division *at all*? I think not.

Ketland's proof is reminiscent of Winnie's (1967) challenge. Winnie, too, proceeds by making a formal division between the predicates that (i) apply only to observable entities, (ii) apply only to unobservable entities, and (iii) apply to both. Lewis (1970)—quoted at some length above (§6.2)—dismisses

such formal division as a trick that does not respect the semantic categories at work. Lewis (1970) is, of course, concerned explicitly with defining theoretical terms on the basis of terms antecedently understood. But it seems that the Ramseyfying structural realist can equally dismiss the formal premise of Ketland's proof merely as a formal trick.

Ketland acknowledges in a footnote (2004: 289, footnote 5) the oddity of gerrymandering the predicates in this way. But he bites the bullet and states that 'in order to make sense of Ramsification [sic] one is forced into some such distinction' (*ibid.*). But why so? If one understood Ramseyfying structuralism to be the position that of the unobservable world only the purely formal logico-mathematical structure can be known, *then* one would be right in demanding that every predicate applicable to unobservable world is to be Ramseyfied away. While there are those who have taken this to be the default position, implied by the word 'structural' in 'Ramseyfying structural realism', this really is a strawman position the untenability of which Ketland has quite rightly shown.¹⁰ We are much better off viewing Ketland's result as a *reductio* of such 'purist' structuralism, a constraint for tenable forms of Ramseyfying structural realism. And it certainly does not follow that if we leave some of the M- or T-predicates un-Ramseyfied, then Ramseyfying realism just collapses to standard realism. The theoretical content of a Ramseyfied theory is simply logically weaker than the content of an un-Ramseyfied theory, regardless of how the class of eliminated predicates is delineated.¹¹

It is clear that the predicates of scientific theories do not naturally divide up in anything like the way assumed in Ketland's proof. The one-place predicates 'has mass', 'has energy', 'is spatio-temporally located' apply to things that are observable and unobservable alike, whilst two-place predicates such as ' x is a part of y ' and ' x is larger than y ', can relate observables to observables, unobservables to unobservables and observables to unobservables.

¹⁰Cf. Psillos (2001, 2005*b*, 2006*b*) for his purist understanding of what structuralism entails, and French & Saatsi (2006) for a criticism of such purism.

¹¹There is a tricky issue with nature-structure distinction here: as far as the structuralist claims to know merely the structure as opposed to nature of the unobservable world, the un-Ramseyfied M- and T-predicates seem to count against this claim. Nevertheless, I think this issue is at the bottom merely terminological. What is crucial is that a tenable form of Ramseyfying realism is a novel proposal to the problem of pessimistic induction. What is *not* crucial is whether or not such Ramseyfying realism fulfils one's prior expectations of what structuralism amounts to!

Now, no sensible structural realist would simply Ramseyfy away all occurrences of the predicate ‘ x is a part of y ’, say. Since the predicate applies to observables as well as unobservables, such wholesale Ramseyfication could *rob* his theory of empirical content. Technically, this is avoided in the above proof by employing a two-sorted language where one sort of variable ranges over observables whilst the other sort of variable ranges over unobservables. At the purely formal level such move is fully legitimate, of course, but it remains to be argued for that it best represents the structural realist’s ambitions.

The consequences of adopting this move are so clearly problematic that we don’t even need fancy model-theoretic considerations to see it. Since *all* the predicates applying to unobservables get eliminated by default, the Ramseyfying realist cannot express anything interesting about the unobservable world. For example, the realist might have it as part of her theory that there are unobservable objects located in spacetime. The predicate that expresses that an entity is spatiotemporally located gets Ramseyfied away if it applies to an unobservable object. Hence the *very weak* claim that there are unobservable objects located in spacetime cannot be part of the theoretical content of the Ramseyfying realist, for the Ramsey sentence says only that there is *a* property these unobservable objects have! Similarly, the realist is unable to express various other rather weak theoretical propositions involving mixed predicates: that some unobservable entities are *part of* some observable entities, that some unobservable entities *move*, etc. Given that so much of the theoretical content is eliminated by the formal division that Ketland’s proof adopts, it is no wonder that triviality beckons!

It should be clear that the realist, structural or otherwise, aims to project an image of science that retains as much as possible of the theoretical content as normally understood. In particular, there is nothing in the spirit of structural realism that implies that all predicates for unobservables should be Ramseyfied away. The structural realist thinks we cannot know certain aspects of the nature of the unobservable world, but that other aspects of the unobservable world can be known. This is quite compatible with the structural realist retaining some interpreted predicates for unobservables. Consider, for example, the predicate ‘ x is part of y ’. That a is a part of b leaves completely open the exact *nature* of a or of b . Whether or not atoms are tiny indivisible Newtonian balls, or whether or not they are complexes of

charged and uncharged entities, or whether they are tiny waves oscillating in an aether, it can still be true that these unobservable atoms are parts of observable objects. Or consider the predicate ‘ x is located in region r ’. That a is located in region r again leaves open the exact nature of a . Whether or not light is a wave or a particle, it may still be true that a light ray is located in a region r . Consider for example ‘ x has velocity v ’. Whatever the nature of light may be it is still true that light can have a particular velocity, or that the velocity may change in certain observable circumstances. By retaining one-place predicates for extrinsic properties and n -place predicates for external relations, structural realists can say definite things about unobservables, such as that they are located or that they form parts of other objects, which does not commit them to any definite thesis about these objects’ nature.

Of course, structural realism is not merely a formal articulation of a philosophical intuition; it is offered as the solution to a particular philosophical problem: ‘how can one accommodate the intuition behind the No-Miracles Argument whilst avoiding the pessimistic meta-induction?’ It is lessons from the history of science that tell us which parts of a theory we can believe and which parts of a theory we should be agnostic about. Whether our main motivation for Ramseyfying is to articulate a notion of structure or whether it is to avoid the pessimistic meta-induction, there is no reason to think that Ramseyfication must follow the Ketland prescription and eliminate each and every predicate that applies to unobservables. The structural realist can and should show discretion in the predicates for unobservables that he chooses to eliminate. In doing so, the structural realist blocks the basic model theoretic argument.

6.3.3 Model-Theoretic Argument Finessed

Although the moral of the last section is a happy one for the structural realist, it is not the final word. Even when predicates that apply to observables and unobservables alike are handled with the requisite care, it may still be the case that a Ramseyfied theory demands too little of the theoretical world.

Given the nature of the basic model-theoretic challenge, it is possible to generalise the problem so as to deflate the theoretical content that goes over and above the un-Ramseyfied predicates that apply to unobservable

entities. The proof of the triviality result above turns on the existence of a bijective function $g : D_1 \cup D_2 \mapsto D_{Obs} \cup D_T$ that is an identity for the elements of D_{Obs} and (inductively) also for the extensions of O-predicates and of those M-predicates that apply to observables. This bijection is then deployed to ‘carve out’ a model for $\mathfrak{R}(\theta)$ which, furthermore, turns out to be the intended interpretation of $\mathfrak{R}(\theta)$. Assume now that we have a mixed predicate which is left completely un-Ramseyfied. Assume, for example, that it is part of the formalisation of our theory that the unobservable entities are located in spacetime, and we leave the predicate ‘is spatiotemporally located’ un-Ramseyfied. Now Ketland’s proof does not go through, since there is undeniable content going beyond the empirical consequences of the theory plus the cardinality constraint. But the proof generalises in an obvious way. That is, we can easily prove that if the theory Θ has a model that is empirically correct, T-cardinality correct, *and* has a domain of theoretical entities each one of which satisfies the predicate ‘is spatiotemporally located’, then $\mathfrak{R}(\Theta)$ is true *simpliciter*. If the intended domain D_T consists of entities spatiotemporally located, then the bijection $g : D_1 \cup D_2 \mapsto D_{Obs} \cup D_T$ indeed satisfies the new constraint automatically.

Theoretical content in the above example is still extremely impoverished, of course, and the Ramseyfying realist probably has a number of other mixed predicates to provide further non-trivial content. But the proof generalises. If we have a theory that is not only empirically adequate, but true also in this or that theoretical proposition, then we can engineer a bijective function G from \mathbf{M} (the model that satisfies the theory) to \mathbf{M}^* (the intended model), in such a way that G is an identity map not only for the relations O_i but also for those M_i that are the extensions of the un-Ramseyfied mixed predicates M_i . The truth of the Ramsey sentence $\mathfrak{R}(\Theta)$ follows from the truth of those theoretical propositions that are expressible in terms of the un-Ramseyfied predicates without existential quantification.

Although Ketland’s proof can be undermined by finding merely *some* M-predicates that are judiciously left outside Ramseyfication by the realist, the above considerations shows that the realist is by no means in the clear. What is also required is that the realist finds a *sufficient* number of such predicates to express theoretical content substantial enough to yield an interesting and substantial realist image in which we can explain the success of past theories

in terms of what they got right about the world. I consider it to be very much an open, outstanding question whether the class of terms to be Ramseyfied can be delineated in such a way that allows the realist to undermine the triviality worry on the one hand—to be true to her basic realist motivations—and sidestep the pessimistic induction on the other. What is badly needed is a painstaking case study of some piece of actual science, and unfortunately the syntactic framework does not really lend itself to easily modelling actual theories to this end. It seems that what is gained in the formal clarity of the Ramsey-sentence logic is lost in the dissociation of the required meta-scientific framework from scientific practice.

If this path is followed nonetheless, then closer attention needs to be paid to what is actually required of the formal framework. There is an important class of scientific predicates that are not naturally modelled at all in the extensional second-order logic in which the debate has been couched. Many of our theories feature unobservable properties that are postulated as the *cause*, or *explanation* of something observable.¹² Perhaps microscopic constituent particles of *that kind* are the best explanation behind *this* observable phenomena. Such intensional idioms can be crucial for expressing theoretical content and unfortunately the extensional framework cannot by itself accommodate this intensional content. (For an example demonstrative of this limitation, see Melia & Saatsi (forthcoming))

We should want to make room in our formalisation for theoretical content featuring *natural* properties, or *qualitative* properties, or *causal* properties. This involves a departure from the strictly extensional framework in which a property is modelled simply as a subset of the first-order domain. In *full* second-order models the second order quantifiers range over all the subsets, which does not seem appropriate for modelling the much more restricted domain of scientific properties of causal-explanatory significance. For a preliminary investigation of strategies the Ramseyfying realist might appeal to, see Melia & Saatsi (forthcoming).

¹²See also the quote from Lewis (1970) in §6.2, above.

6.4 Exegetical Commentary

The forefathers of the contemporary forms of structural realism are many and varied. Worrall originally drew his epistemic motivation from Poincaré, but this link can only be seen as inspirational given Poincaré's Kantian inclinations. Another eminent figure featuring in the prehistory of structural realism is Bertrand Russell. (Worrall & Zahar 2001, Votsis 2003) It is particularly interesting how Russell gets appealed to in order to motivate the adoption of the Ramseyfication procedure to the modern structural realist arsenal, given the 'received' purely semantic use of Ramsey sentences by Carnap, Lewis and others. In the literature the connection from the modern realists to Russell's epistemic structuralism, as expounded in *The Analysis of Matter*, goes through Grover Maxwell (1970a,b). And the just-repudiated model-theoretic challenge too originates from the Russellian connection (Demopoulos & Friedman, 1985), as already mentioned above. It is this connection that I now want to criticise in some more depth.

Russell was a structuralist about our knowledge of the external world: his response to idealism was to present a philosophical theory that explained how from the direct knowledge of percepts by acquaintance one could derive true propositions expressing our knowledge of the external material world by description. From the knowledge of percepts and their relations one could derive *the structure* of the material world by virtue of a principle stating that different percepts require different material events as their causes. The notion of *structural similarity* pertinent to Russell's analysis is a notion of pure logic: a class of percepts with various monadic properties and relations between them has an abstract (logico-mathematical) higher-order structure defined roughly speaking as the equivalence class of isomorphic (first-order) structures. (cf. Demopoulos 2003b, Votsis, 2003) This notion together with the above principle implies the existence of a (partially) isomorphic structure in the domain of external material events. (Russell, 1927: ch. 20) According to this structuralist '*causal theory of perception*', if one acquires knowledge by acquaintance of a spatial geometrical figure—a simple (infinite) line, say—then one can infer of the external world material events causally responsible for this perception that they also have the logico-mathematical structure of *linear order*. That is, whatever (first-order) properties and relations hold between these material

events the intrinsic nature of which remains unknowable, they instantiate the abstract higher-order structure isomorphic to natural numbers. This is the position against which Newman (1928) objected. Newman showed conclusively that the content of our knowledge of the external world cannot be only of its logico-mathematical structure.

We should ask to what extent Russell's notion of structural knowledge is continuous with that of the contemporary structural realists' who appeal to Ramseyfied theoretical content. The idea that Ramsey sentences could capture a notion of structure useful for delineating a scientific realist position first surfaced in Maxwell (1970a,b). Incidentally, Maxwell also regarded the resulting position as a realisation of Russellian epistemic structuralism. The latter succumbing to Newman's challenge, it is not too far fetched to think that Ramseyfying realism must fall foul of the same problem (Demopoulos & Friedman, 1985).¹³ This negative conclusion was rejected in the previous section. I now want to argue that despite some vague intuitions there is no natural link between Russell and the contemporary Ramseyfying realists, and that the adoption of Ramsey sentences by some structural realists has been motivated by an illusory connection to Russell's structuralism.

Russell did not employ Ramsey's elimination method in *The Analysis of Matter*.¹⁴ We can, of course, nevertheless recognise Ramseyfication as something that Russell perhaps could have used to express his notion of structural knowledge; after all, Ramsey sentences (without any extra-logical names) naturally exhibit purely logico-mathematical structures. (Cf. Shapiro, 1997: 108) But Ramsey elimination is an extremely general logical manoeuvre applicable to many ends; the role this operation can play depends entirely on what one is Ramseyfying over. To express Russell's epistemological idea of the causal theory of perception in these terms, one should first construct a logical formalisation of the knowledge contained in the field of percepts. Ramseyfying this perceptual content would then yield an abstraction to be taken as the abstract structure realised by the field of percepts in question. As the final

¹³Actually Demopoulos & Friedman (1985) argue for the more general conclusions that Newman's problem is a problem for a 'theory of theories' in which Ramsey sentences are used to interpret theoretical terms.

¹⁴The extent to which Ramsey influenced Russell's structuralism is an interesting question, though.

step one would then declare a partial isomorphism between this structure and the external world of material events causing these percepts.

In contrast to such expression of Russell's structuralism in terms of Ramseyfication, Maxwell's use of Ramsey sentences is premised on the logical empiricist conception of theories with its notorious observational/theoretical term distinction. What Maxwell wants to Ramseyfy away are not the observational terms—'the percepts'—but the theoretical terms implicitly defined via observational and logico-mathematical terms. But the resulting Ramsey sentence has very little to do with the Russellian notion of structure, since the Ramseyfied part is not hooked up with the empirical world via a (partial) isomorphism. If the observational terms were to be Ramseyfied over as well, the result would undeniably resemble the Russellian structure in being a totality of logico-mathematical relations between otherwise unknown relata, but this sentence would not be an abstraction of the field of percepts. Maxwell simply cannot assimilate Russellian structuralism into his Ramseyfying framework by mere stipulation. In Maxwell the notion of 'structural characteristic' as a property that is 'not intrinsic and can be described by means of logical terms and observation terms' (where 'intrinsic properties are those that are ... direct referents of predicates') does not seem to be quite the same as Russell's idea of purely structural knowledge appealed to in the causal theory of perception.

This is not very surprising, given that Russell's causal theory of perception concerns the *external*, rather than the *unobservable* world, and is independent of the notion of scientific theory. Russell's structuralism here is wholly contained within the very general epistemological idea about the inferred knowledge of the external world behind both the common sense perceptions of tables and chairs, as well as the theoretical knowledge. It is undeniable that in *The Analysis of Matter* Russell was also concerned about the nature of scientific theories, how they anchor onto our observations, and how statements about unobservable entities should be construed. But as far as the very clearly stated *explicit* idea of the purely structural knowledge of external world is concerned, Russell's structuralism should not be identified with the idea of Ramseyfied theoretical content.

Demopoulos & Friedman (1985) take a broader perspective on Russell's structuralism. They claim that in addition to the foregoing epistemological

idea, ‘Russell’s structuralism can be viewed as a theory of how the reference of the theoretical vocabulary is fixed’.

Russell wishes to exploit the notion of logical form or structure to introduce scientific objects and relations by means of so-called *axiomatic* or *implicit* definitions. [...] Russell is prepared to accept the Ramsey-sentence $\mathfrak{R}(T)$ as the proper statement of our scientific knowledge. (1985: 622)

This claim is based on Russell’s work on descriptions and propositional understanding, and rightly draws a connection between Russell on the one hand, and Ramsey, Carnap and some other Ramseyfiers on the other. (cf. Demopoulos, 2003a) But this is not the connection that motivated the contemporary epistemological structural realists to adopt Ramseyfication in the first place! Rather, the motivation was Russell’s explicit structuralism about our knowledge of the external world, together with the impression—acquired from Maxwell (1970a,b) and Demopoulos & Friedman (1985)—that Ramseyfication is the canonical expression of this idea. In the contemporary context we translate ‘items of acquaintance’ as ‘observables’ and ‘external world’ as ‘unobservables’ and, voilà, we have a Ramsey sentence formalisation of structural realism. (Votsis 2003, Worrall & Zahar 2001) I think we now can justifiably question such ‘inherited’ structuralist connotation of Ramseyfication. Indeed, Demopoulos (2003b: 395–396) sets the record straight by clarifying the connection between Ramseyfication and Russell’s structuralism: the former can be used to express the latter, but not every Ramsey-sentence expresses structural content in Russell’s sense.¹⁵

Despite all this, there are good reasons for exploring in detail the potential realist virtues of Ramseyfication, due to the fact that $\mathfrak{R}(\theta)$ is logically weaker than θ , as explained in §6.2. Perhaps there is even a decent sense in which the Ramseyfied content *could* to be termed structural, after all. But these reasons are quite independent of the connection of the contemporary Ramseyfying realist position to Russell’s epistemic structuralism as represented in the literature.

¹⁵In a footnote Demopoulos also stresses that ‘the secondary literature is frequently misleading on this point, often suggesting a closer conceptual link between Russell’s structuralism and the notion of a Ramsey sentence than in fact exists’, but he does not give references (note 7, p. 416).

* * *

I began this second Part with a favourable appraisal of the epistemological structuralist idea that Worrall (1989) successfully revived from Poincaré's writings. Ramseyfication, on the other hand, is an oblique spin-off from Russell. But it is still ambiguous how exactly these two intuitions behind Worrall's structural realism are meant to fit together. Is Ramseyfication to be employed to uncover continuity in the formal logico-mathematical structures, as one reading of Worrall's presentation of the Fresnel-Maxwell example suggests? Or is it to be employed to uncover continuity in the physical content of the two theories, as another reading of Worrall's rhetoric might submit? Given our discussion of the model-theoretic triviality worries above, it seems that only the latter interpretation can pay dividends here. But how should this physical content be construed, and how exactly can Ramseyfication be employed to express it? Perhaps, if we were to logically regiment both Fresnel's and Maxwell's theories, and to compare the resulting Ramsey sentences, then perhaps we could see a way of spelling out a complete realist image in these terms. Perhaps.

In the absence of the required details I would recommend changing tack here, and looking at the possible ways of delineating continuity in the theoretical content in logically more informal, but perhaps philosophically more informed terms. In particular, we should be able to analyse theoretical content worthy of realist commitment *conceptually* prior to any formal representation of it. This will be my approach in the following third and final Part of this thesis.

Part III

Beyond Structuralism

Explanatory Approximate Truth

“If the realist is going to make his case for convergent epistemological realism, it seems that it will have to hinge on approximate truth, rather than reference.”

Larry Laudan, *A Confutation of Convergent Realism*

7.1 Responding to Pessimistic Induction

Responding to the argument from pessimistic induction has been a central task in the realist agenda. Most realists have attacked this argument (examined in chapter §4) by attempting to directly undermine its rather pessimistic premise that there are a significant number of past theories which were successful but cannot be considered to be approximately true in any reasonable sense. Although there is no clear verdict to be found on what this significant number would be—i.e. how large the ‘inductive basis’ of the pessimistic ‘induction’ needs to be—it seems that many realists are bothered by the argument to the extent that they want to deal with *each and every* item on Laudan’s pessimistic list in order to leave nothing for the anti-realist to work with.

Allegedly a particularly troublesome item on Laudan's list consists of the 19th-century ether theories of optics, featuring both bona fide novel successful predictions and a prime example of a completely rejected key theoretical concept. The current consensus is that the realist wishing to deal with this 'counter example' to the realist image head-on needs to clarify her conception of approximate truth and/or of reference in order to deal with the intuition that the ether theories are false simpliciter. This intuition, shared by Laudan, is simply due to the fact that in our world there is no pervasive elastic mechanical medium the existence of which the successful predictions seem to have relied on—i.e. the theoretical term 'ether' seems prima facie non-referring. The challenge, in other words, is to delineate *some* theoretical content to accompany the undeniable cumulative continuity at the empirical level, enough to *explain* the success of past science in terms of its approximate truth. The challenge thus posed immediately presents itself in form of the following two questions: (1) What does it take to explain a particular success of science? (2) If appeal to *some* theoretical content is indeed required as the explanans (as the realist argues), then on exactly what *principled grounds* should this content be delineated? These questions are considered in this chapter.

A notable array of realists have taken pains at elaborating their position in the face of this challenge, producing a variety of responses to both Laudan as well as to each other. (Hardin and Rosenberg 1982, Kitcher 1993, Worrall 1989, 1994, Chakravartty 1998, Psillos 1999) In the next chapter I will take yet another look at a case study which has been particularly popular in this context: the theory shift from Fresnel's prediction of reflection/refraction amplitude relations for polarised light to their modern derivation from Maxwell's electrodynamics. This is a case well featured in recent realist dialectic between Worrall and Psillos about how we should properly respond to Laudan.

These two authors, with their respective commentaries on the theory shift, arrive at formulations of realism radically at odds with each other. Worrall's *epistemic structural realist* knows next to nothing about the nature of unobservable entities in themselves and is not bothered by abandonment of central theoretical terms like 'ether' in scientific revolutions, as long as the *formal structure* of the theory exhibits appropriate continuity in theory change. Psillos's more orthodox formulation of realism, on the other hand, allows evolving

conception of the nature of the ether only to the extent that it allows him to argue to the prima facie surprising conclusion that ‘ether’ (*pace* Laudan) is not a non-referring term after all! My claim is that both of these positions are ultimately too extreme to be tenable: Worrall’s structuralism tends to collapse into tracking a trivial kind of continuity with no real explanatory power, whilst Psillos is inclined to find unwarranted levels of continuity over and above Worrall’s structure. But there is no need for pessimism either, for a more natural realist position—at least as far as the key case study of the Fresnel-Maxwell theory-shift is concerned—is to be found somewhere between these two opposites.

Despite arriving at two positions so widely at variance, Psillos and Worrall can actually both be seen to follow the same overall strategy to undermine the contribution of this particular item to Laudan’s pessimistic premise. Namely, their shared objective is, broadly speaking, to arrive at a notion of approximate truth that would provide the realist with the necessary resources required to explain the success of the earlier theory in light of the present theory. Psillos dubs this strategy ‘*divide et impera*’, and explicates that

... it is enough to show that the theoretical laws and mechanisms which generated the successes of past theories have been retained in our current scientific image. I shall call this the ‘*divide et impera*’ move. (Psillos, 1999: 108)

The relevant notion involved is called *explanatory approximate truth* (EAT) here, and a careful analysis of this is a prerequisite for fruitfully considering the proper implications for realism of the historical case study. The present Part hence begins by analysing the notion of explanatory approximate truth in this chapter. After critically reviewing both Worrall’s and Psillos’s elaboration of this notion (§7.3 and §7.4), it moves on to reconsider the details of the case study itself. The lesson to be learned from the Worrall-Psillos juxtaposition is that in order to avoid the triviality of Worrall’s structuralist construal of explanatory approximate truth, the realist does not have to defend a (problematic) reference invariance formulation of EAT as Psillos would have it. Rather, the analysis of EAT exposes the possibility of elaborating a variant of this notion which is sensitive to the hierarchy of theoretical properties appealed to in scientific theorising. The kind of hierarchy referred to

is best exemplified by the details of the case study itself (chapter §8). In the final chapter §9 I will elaborate on these findings by framing the issue in terms of reductive explanation. It will be concluded that the details of the case study, properly understood, point towards a novel formulation of scientific realism.

7.2 Explaining theoretical success of rejected theories

The basic realist intuition behind explanatory approximate truth is readily understood on the basis of the No-Miracles Argument. The intuition behind this argument is that the best (or only) explanation for a theory's success is its approximate truth. As far as PMI threatens to pose a problem to this intuition, it is natural for the realist to try and respond by refining her notion of approximate truth. Worrall, for example, suggests that

... the realist needs to show that, from the point of view of the later theory, the fundamental claims of the earlier theory (in so far as they played integral roles in that theory's empirical success) were—though false—nonetheless in some clear sense 'approximately correct'. He needs to show that, from the point of view of the later theory, we can still explain the success enjoyed by the earlier one. (1994: 339)

The basic idea is hence to show that the success of past science, by and large, did not depend on what we now take to be fundamentally flawed theoretical claims, and coupled with the No-Miracles Argument such an understanding of past successes licences a realist belief in the theoretical elements involved in the explanation of these successes. This rough idea of responding to PMI by elaborating a workable account of EAT is subscribed to by many contemporary realists, albeit with diverging details. There is disagreement on the exact form this strategy should take and—regarding Fresnel's theory in particular—whether or not 'ether' refers and whether this matters or not. (Worrall 1989, Kitcher 1993, Chakravartty 1998, Psillos 1999)

Before attending to the form these elaborations take in Psillos and Worrall, the crucial notion of explanation should be clarified further in the abstract. Two questions arise immediately. First, since the spirit of EAT is to

restrict the realist commitment to those and only those things which are used to explain the success of a theory, we need to ask how to characterize ‘the theoretical laws and mechanisms responsible for success’ to begin with. And this is clearly closely tied up with the question of what explaining the success of a past theory exactly amounts to.

To begin with the obvious, I take it to be indisputable that the realist really needs to provide a *realist* explanation of the success of past theories on pain of inconsistency: success is a success, whether of a past or present theory, and the best explanation simply ought to be the same for both.¹ So arguably the best explanation is that a past, rejected but successful theory is false yet approximately true. The chronological order of explanation here is from the present to the past: the success of a *past* theory is to be explained *in the light of* our current best theories, the latter assumed to be approximately true.² Given that the past theories, taken at face value, can be really wide of the mark in presenting the world as we now take it to be, the critics really are right in demanding an explication of the sense of approximate truth applicable here. Luckily for the realist, the anti-realist intuitions about approximate truth are gratuitously pessimistic; EAT suffices for the realist purposes.

The realist’s philosophical explanation of a theory’s success is a logical one: valid arguments with true premises always lead to true conclusions. Explaining the success of a past theory in this spirit now involves identifying truth-content in the theory which enables this form of logical explanation to be applied, without being compromised by the immanent falsehoods.³

Explanation thus understood, let us initially characterise EAT broadly as ‘truth of a theory in those qualitative aspects directly reflecting the unob-

¹Recall chapter §4 where we scrutinised arguments attempting to sidestep this assumption.

²Although this may at first seem circular, this worry is unfounded. The reason is that the realist assumes that the present theory is closer to the truth, or at least more comprehensive than the past theory. Also, the task of extracting the exact truth content of theories falls on the scientist, not the philosopher. (cf. Psillos 1999: 113)

³I emphasise again that here we are *not* concerned with defending the justificatory issue of explaining the successes of our *present* theories in this way. Rather, we are only concerned with explaining the successes of past theories on the basis of the assumption that the present ones are true. In chapter §1 we saw that the best case for NMA regarding the latter assumption was not particularly strong. This result does not render the presently evaluated realist project in any way meaningless, however.

servable reality which can explain the theory's success'. The generality of this depiction is a virtue since it allows that the character of EAT can be to a large extent specific to each particular success to be explained, rather than being given *tout court*. On the other hand, however, the characterisation above seems so weak as to deflate realism into a virtually empty promissory note! Are our realist commitments with respect to some presently entertained successful theory to be summarised merely as 'truth in those aspects which will *turn out* to explain (in realist fashion) the theory's success from any later vantage point'? Let us call this position *minimal explanatory realism*. Although rather deflated, it is still a realist position by virtue of both preferring a realist explanation of success over miracles, and sticking its neck out vis-à-vis historical as well as future science. For all the PMI instances the realist claims to be able to identify elements of reality in the past theories which explain their success. But the minimalist requires initially nothing from these elements other than that they can fulfil the essential explanatory function, and hence they may in principle form a gerrymandered lot. The only principled constraint is for theories which themselves are used to explain the success of their predecessor(s): the success of the current theory must be explicable (in light of any future theory) in a way that is *compatible* with the explanation(s) offered by the present theory.⁴

The realist positions examined in the rest of this thesis, my own proposed variant included, can be viewed as building on such minimal explanatory realism by fixing a principled framework which provides initial guidance for the explanatory endeavour. Instead of being potentially gerrymandered in form, the various instances of EAT are unified in order to pronounce more definitive realist commitments with respect to the current science. Fixing this framework is a matter of answering the initial question of how the potential success-fuelling theoretical constituents should be conceived and characterised to begin with. I believe that this question is a non-trivial one and that a refined attitude towards it will be for the benefit of the realism debate. The remainder of this chapter serves to motivate this belief which, it will turn out,

⁴Perhaps this constraint is actually stronger than it initially seems. Most current theories are needed to explain various theoretical success-stories of the past, and the mere requirement of explanatory consistency may fix the realist commitments of minimal explanatory realism surprisingly precisely.

is fulfilled by the case study considered later.

The realist faces the task of extracting truth content from past theories to explain their successes. Clearly there is nothing in this kind of logical explanation itself that dictates the nature of the truth content eligible: we only need to be able to derive a conclusion which constitutes the theory's success from the truth content, by way of a valid argument. This suggests that we really need to look at particular theoretical derivations in detail to discern the kind of content they depend on.⁵

Glancing over various derivations in physics immediately indicates that characterising and extracting explanatory success-fuelling elements is a highly non-trivial matter. It is commonplace that theories typically speak of various unobservable entities, kinds of physical particulars endowed with qualitative properties and relations, interacting in various ways. This is usually our preferred way of conceptualising theoretical content. What is less well appreciated is the fact that our access to theoretical properties and relations is a complex one. Typical fundamental theoretical properties (e.g. charge, spin, magnetic field amplitude) are defined not only through some dynamical, causal force laws describing how the entities instantiating them interact and are ultimately indirectly observed, but also through a mixture of theoretical principles of conservation and symmetry, for example. Acknowledging the complexity of a typical theoretical description gives gravity to the question of how best to characterise the success-fuelling theoretical constituents. In particular, it is not by any means obvious that successful derivations operate on theoretical propositions best conceived of in terms of interacting entities with causal properties. Are the explanatory ingredients approximately true descriptions of the kinds of unobservable objects the theory speaks of, or approximately true descriptions of the properties and the theoretical principles involved in the derivation of novel successes?

Different answers to this question can lead to alternative realist positions. The explanatory account of derivation congruent with 'standard' realism has always been construed as approximate truth about unobservable objects and processes denoted by the central kind terms. From Psillos (1999), for ex-

⁵I will mainly focus on logico-mathematical derivations here, to lay ground for the Fresnel-Maxwell case study, but I believe that my analysis generalises naturally to logical scientific reasoning more generally.

ample, something akin to the following can be distilled.⁶ The referents of kind terms are seen (informally, not necessarily metaphysically) as bundles of properties, and EAT gets spelled out in terms of sets of properties: if a subset s of all properties S attributed to a posit E of a past theory T_P is found to be fuelling a novel derivation in T_P , in light of the current best theory T_C , and s is also a subset of essential properties S' attributed to a posit E' of T_C , then T_P is approximately true assuming that ' E ' can be seen as referring to E' . Explaining a derivation in these terms boils down to finding a continuous set of properties s that (from the current perspective) were necessary for the derivation. There are then two options in the *standard realist account* of EAT. Either some entity allegedly referred to in the past theory does not actually have *any* of the properties that play a part in producing a particular success taken as the explanandum—in which case the term in question is non-referring but unproblematically so—or this entity has these significant properties but was also incorrectly attributed other 'misleading' properties in the past theory, and the term in question refers to whatever satisfies the correct description. To ensure that continuity in the success-fuelling properties entails also referential continuity an appropriate theory of reference of theoretical terms is required as a fundamental part of this framework. Implicit in this account of EAT is the idea that theories cannot be approximately true (in a sense worthy of realist attention) unless the central terms denote something out there in the actual world—where 'centrality' is understood as having a success-fuelling referent.

But it is not *prima facie* clear that this standard account of EAT with its referring-kind-terms is at all optimal in trying to understand why the logico-mathematical reasoning leading to some novel success took place within a particular theoretical world view, say. In practice one attempts to under-

⁶Psillos's standard formulation of realism is adequately summarised for the time being as follows: scientific theories describe unobservable entities, their properties and causal interactions / processes. Theoretical terms have putative factual reference (the semantic component of realism), successful theories are approximately true and 'entities posited by them, or, at any rate, entities very similar to those posited, inhabit the world' (the epistemic component of realism). This formulation seeks to keep EAT as close to the intuitive correspondence notion of approximate truth as possible. The above discussion on EAT should make it clear that the referential, semantic component of realism does not come automatically with the epistemic component, but depends on the way the latter is spelled out.

stand why the reasoning followed in the derivation leads from one step to the next in a way that ultimately explains the end result by appealing to approximate truth of that reasoning. Applying the *divide et impera* at each step throughout the derivation, we may find that the derivation of a mathematical relation suggested by the theoretical picture is best explained by *approximately true description of the properties and theoretical principles involved*, rather than of kinds of objects as conglomerates of these properties. That is, approximately true theories in this sense give a false characterisation of the right property, and only (some of) the true aspects of this false characterisation are essential for the derivation. The referential question of whether we have two different descriptions of the same *entity* instantiating these right properties, or descriptions of different entities, is a red herring here. Rather, the explanatory work would be done at the level of *properties*.

These initial general remarks are best explicated further via the actual case study undertaken in the next chapter §8. Before proceeding to this, however, we should review in more detail the elucidating views of Worrall and Psillos on EAT.

7.3 Worrall on Explanatory Approximate Truth

As already mentioned in section §5.1, Worrall (1989, 1994) appeals to the Fresnel-Maxwell theory shift in his argument for a structural realist position. After defending the *divide et impera* intuition (as captured by the quote above, p. 152) the argument proceeds by ruling out the standard account of EAT as appropriately capturing the case study in question.

A natural assumption is that such an explanation requires a demonstration either that the parts of the earlier theory rejected by the later one were redundant or that no real ‘rejection’ was involved (but only a ‘re-description’). However, in this particular historical case at least, the most straightforward and least revisionary account of the explanation ... fits neither of those patterns. (1994: 339)

In its stead, Worrall suggests that the most natural explanation of the success of Fresnel’s theory is given in terms of a structuralist construal of

EAT: that Fresnel ‘misidentified the *nature* of light, but his theory nonetheless accurately described not just light’s observable effects but also its *structure*’ (1994, 340).

Worrall is right in claiming that Fresnel should not be interpreted as having spoken about the electromagnetic field all along, or so it will be argued later on (against Psillos). But it is not clear that Worrall has given a fair run to the alternative possibility of demonstrating that ‘the parts rejected were redundant’. The problem is that evaluating this option requires a careful consideration of how these ‘parts’ should be construed in the first place, and Worrall provides no discussion of this matter.

For example, if we take the ether as a theoretically posited entity to be a rejected part of the theory, then indeed it is difficult to argue that this part was (‘as a whole’) redundant for the derivation of the success of Fresnel’s theory. Such a demonstration, I must agree with Worrall, would have a serious air of ‘whiggishness’ about it. But why would the identification of rejected parts and the ensuing demonstration that those parts were redundant operate at the level of objects and kinds—as the standard realist account would have it—as opposed to the level of properties these objects were taken to instantiate? Theoretical reference, of course, is to the unobservable entities and the existence of non-referring terms has been traditionally taken to constitute a strong indication of the failure of approximate truth. But Worrall, with his structuralist theory of approximate truth, already marks a departure from this tradition; the structuralist is not bothered about non-referring central terms, anyway, since the structuralist’s explanatory endeavour takes place not at the deep level of entities, but at the level of ‘structure’. Similarly, if we managed to explain the success of Fresnel’s theory by demonstrating that those *properties* of the ether which fuelled the derivation of the novel empirical prediction were retained, whilst those rejected were redundant, then that presumably would amount to a legitimate realist elucidation of EAT despite ‘ether’ possibly turning out to be a non-referring term. But this would still be in the spirit of the first strategy that Worrall rejects! If this approach were to go through there would be no need to be radical and adopt a structuralist theory of EAT.⁷

⁷If we choose to ignore this option and opt for a more radical line of thought, as Worrall does, then there are problems to be faced with the suggestion that, in general, truth about

Psillos (1995, 1999, 2001) has argued against the nature-structure distinction that Worrall commits to. Although I am far from convinced that there is no useful such distinction to be had, against Worrall this criticism bites since there is no proper explication of the central notion of structure to be found, but rather some serious ambiguity. Mostly Worrall speaks of ‘formal’ or ‘mathematical’ similarities in Fresnel’s and Maxwell’s theories. He says, for example, that

...disturbances in Maxwell’s field do obey formally similar (mathematically identical) laws (1994: 340)

and that

... there is structural, mathematical continuity between the two theories. (*loc.cit.*)

But there are also remarks of quite a different spirit to be found, such as the idea that although

... Fresnel was as wrong as he could have been about what oscillates, he was right, not just about the optical phenomena, but right also that those phenomena depend on the oscillations of something or other at right angles to the light. (*loc.cit.*)

In the usual philosophical terminology a truth about spatiotemporal relations such as those expressed by ‘oscillation at right angles to the direction of propagation’ is *not* a formal, mathematical truth. It is a truth about some *properties* being instantiated where there is phenomenon of light. Oscillating, like ‘having a velocity’, is a property that many different kinds of entities can instantiate; both water and the electromagnetic field can oscillate, and so can the distribution of colour intensity on a computer screen. Oscillating is a *higher-order spatiotemporal property* of a system, and it can be *multiply*

formal, mathematical structure of a theory is truth enough for the realist, and that this construal of EAT adequately *explains* the success of at least Fresnel’s theory, in particular. Worrall does not explicate in any detail this structuralist notion of approximate truth or the sense of explaining this notion fulfils—rather, he just relies on our intuition on this matter. I am afraid these intuitions do not bear a closer scrutiny. See Psillos (1999: 151ff) for criticism.

realised by various lower-order properties (e.g. location of water molecules, electric and magnetic field amplitude, colour intensity).⁸ If there is a decent sense in which such higher-order properties could be termed structural, it is not at all apparent but rather in serious need of explication. Furthermore, it is not clear which one of these divergent readings of structure-as-opposed-to-nature is meant to undertake the explanatory work of EAT for Worrall. In the next chapter an explanation of the success of Fresnel's theory is put forward according to which this can be understood in terms of such multiply realisable spatiotemporal properties which are correctly circumscribed by Fresnel's theorising. Whether or not there is a decent sense in which this explanation could be called a structuralist one, the fact remains that Worrall's move from our knowledge of such properties to the knowledge of formal, mathematical structure is a non sequitur.

7.4 Psillos on Explanatory Approximate Truth

After criticising Worrall's attempt to put a structuralist spin on explanatory approximate truth, Psillos proceeds to salvage the standard account of EAT from Worrall's (and others') criticism, thereby doing away with the principal motivation to 'go structural' in the first place.

Psillos advocates the standard view through a set of detailed case studies. The case study of *dynamical optical ether theories*, in particular, 'suggests that the most general theory—in terms of Lagrangian dynamics and the satisfaction of the principle of the conservation of energy—which was the backbone of the research programme . . . has been retained' (1999: 113). This, together with a suitably tailored theory of reference, provides the grounds for Psillos to argue that 'ether' can be taken to refer to the electromagnetic field.

Prior to presenting the case studies, Psillos does not much elaborate on the notion of EAT in general, apart from the obvious logical characterisation of indispensability of truth-like success fuelling constituent.

Suppose that H together with another set of hypotheses H' (and some auxiliaries A) entail a prediction P . H indispensably contributes to

⁸In a broader sense the term is sometimes used without the non-spatiotemporal connotation, referring not to periodic change in spatial properties but in any quantifiable property, as in 'oscillating exchange rates', for example.

the generation of P if H' and A alone cannot yield P and no other available hypothesis H^* which is consistent with H' and A can replace H without loss in the relevant derivation P . (1999: 110)

This admittedly makes good sense but does not give much of a handle on how the actual case studies should be conducted. Psillos, of course, aims to extract from the case studies a level of continuity required to defend his orthodox realist position. That is, Psillos aims to retain referential invariance for those posited *kinds*—like the luminiferous ether—which cannot be ruled outside of success-fuelling constituents without violating the standard account of EAT. In the next chapter I demonstrate that these case studies, at least as far as the central Fresnel-Maxwell example goes, require a more open-minded construal of the explanatory success-fuelling constituents. When looked at in closer detail this case actually does not conform to the mould offered by the standard realist. More specifically, issue will be taken with the reference invariance claim of the standard account of EAT—the claim that armed with a suitable causal-descriptive theory of reference we can plausibly take ‘ether’ to refer to the electromagnetic field.⁹ Before entering the case study itself the source of the problem will be expressed in more general terms, below.

7.5 EAT and success-fuelling properties

It is my contention that a more natural explanation of the success of Fresnel’s theory can be had by adopting a more refined notion of success-fuelling theoretical constituent. In particular, it seems to be a mistake to view these as *propositions about unobservable objects and kinds* featuring in the theory, in such a way that an indispensable hypothesis must be linked to successful reference of the kind names featured in the hypothesis in question. Rather, the realist should adopt a more open-minded characterisation of these constituents, and make the working/idle –distinction at the level of *properties* instantiated by these kinds, for those properties involved in the derivations.

⁹There is also a general worry about this kind of reference-insisting construal of EAT, raised by some interesting recent arguments to the conclusion that the realist is playing an illegitimate semantic game in trying to salvage her realism by tailoring theories of reference to ensure referential invariance. Cf. Bishop & Stich (1997), Bishop (2003) and Cruse (2004).

Something along these lines has been proposed by Chakravartty (1998), also commenting on Worrall's structuralism and the Fresnel-Maxwell case study. Chakravartty's demarcation between *detection* and *auxiliary* properties provides a useful reference point regarding the present suggestion, and he, too, declares that 'distinguishing kinds of properties may in fact distinguish forms of realist commitment' (394). The crucial distinction for Chakravartty is between those properties 'upon which the causal regularities of our detections depend, or in virtue of which these regularities are manifested' on one hand, and those 'associated with the object under consideration, but not essential (in the sense that we do not appeal to them) in establishing existence claims', on the other (394–5). Properties of the latter kind—the auxiliary ones—then only 'supplement our descriptions, helping to fill out our conceptual pictures of objects under investigation', whereas 'only the former are tied to perceptual experience' (395). The detection properties hence form the subject matter of realist commitment. The insightful observation that the explanatory identification of success fuelling constituents should operate directly at the level of properties involved in a theoretical derivation is certainly a move to the right direction. But the distinction between the auxiliary properties and the rest, as it stands, is only rough and ready and it is not quite clear how this differentiation at the property level is meant to interact with the object level talk. The rest of this thesis is an attempt to flesh out the details of this basic idea, by looking in closer detail at the various distinctions that can be made regarding properties as the explanatory constituents of EAT.

For one thing, it seems that Chakravartty's analysis of detection properties does not probe deep enough into the explanatory structure. The characterisation of these properties as directly contributing into causal regularities, as opposed to purely metaphysical images of the entities thus contributing, provides a useful starting point, but does not exhaust the dimensions of EAT. A crucial facet of properties often ignored in this context is their hierarchical nature (for lack of a better word). Theoretically posited objects and kinds, described as bundles of first-order properties, do not *as such* fuel a successful derivation. These bundles come and go in radical theory change, and the realist is naturally led to consider which particular kind-defining properties are invariant over an otherwise radical theory shift. But it is

not necessarily these properties *as such* that fuel the derivation either, for an instantiation of a property can in turn necessitate the instantiation of higher-level properties and *these* can be the explanatory elements in the final analysis, whilst the lower-level properties exhibit variance in theory change. For example, spatiotemporally distributed systems typically instantiate important spatiotemporal higher-level properties: a spatiotemporal distribution of lower-level properties fluctuating in a particular manner, for instance.

Once again, these (somewhat cryptic) preliminary remarks are best clarified in terms of revisiting the actual case study at the heart of this polemic. After doing that in the next chapter a general framework to accommodate the emerging picture is offered in terms of reductive explanations in science (chapter §9).

Explaining the success of Fresnel's theory

The realist is challenged to explain how Fresnel was able to predict the intensity of reflected and refracted polarised light from seemingly false presuppositions. This chapter looks in more detail at Fresnel's work and offers an explanation which draws solely on those premises of Fresnel's derivation which are not refuted by the modern understanding of the phenomenon. It is imperative to first examine the theory-shift and the electromagnetic understanding of optics in some detail (§8.1), to gain a grip on the kind of continuity this explanation capitalises on (§8.2–8.3). The notion of EAT considered in the abstract in the previous chapter is finally applied to the case study in section §8.4.

8.1 Reflection and refraction from Fresnel to Maxwell

Augustin Fresnel's successful derivation of the reflection and refraction amplitudes for polarised light in the early 1820s stands out as a significant chapter in the development of optical ether theories. Fresnel's successful conception of light as a *transverse* oscillatory mechanism was considered as one of the most fundamental discoveries about the nature of the elastic ether, and it was a crucial test for any proposed ether model that it could reproduce *Fresnel's*

equations (1823) for the amplitudes of reflected and refracted light. For the two components of the reflection amplitudes $A^{reflected}$, for instance, Fresnel derived the equations

$$A_{\perp}^{reflected} = \frac{-\sin(i - i')}{\sin(i + i')} A_{\perp}^{incident} \quad (8.1)$$

$$A_{\parallel}^{reflected} = \frac{-\tan(i - i')}{\tan(i + i')} A_{\parallel}^{incident} \quad (8.2)$$

The terms here refer to: A 's are amplitudes, with A_{\parallel} and A_{\perp} the components parallel and perpendicular to the plane of incidence (spanned by incident and reflected rays); i is the angle of incidence (as well as reflection) and i' is the angle of refracted light (cf. Figure 8.1 in §8.2). The amplitudes (the vector sum of components) are linked to observations so that $A^2 \propto$ intensity.

An ether theory that eventually managed to reproduce the laws of reflection and refraction was that of McCullagh's *rotational ether*, but this in fact turned out to be just part of Maxwell's theory in disguise, since the dynamical assumptions employed by McCullagh turned out to be unrealisable by a system of material, elastic ether, but satisfied exactly by the field of Maxwell's theory as shown by Fitzgerald in 1878. (cf. Stein, 1982) Maxwell's theory indeed produces equations formally equivalent to those of Fresnel's theory, as first shown by Lorentz in his Doctorate thesis (1875). It is this formal correspondence that was appealed to by Poincaré (1952) and Worrall (1989, 1994) in defence of their structuralist positions (cf. §5.1).

The equations (1) and (2) of Fresnel's theory were derived via theoretical principles and heuristic analogies from what was known about elastic mechanics (e.g. conservation laws of energy and momentum in oscillations) and material wave motion, insisting all the time on purely transverse oscillatory mechanism (despite thus creating a severe disanalogy with the nature of mechanical wave motion known). The result is a quasi-mechanical theory of the ether which is not based on a set of exact, consistent principles and force functions, but rather on some fitting dynamical boundary conditions applied to a close geometrical analysis of the problem. It is the interpretation of these boundary conditions that renders Fresnel's construction specifically an ether theory. We will come back to this below in more detail.

In Maxwell's theory, on the other hand, the corresponding equations follow by imposing a natural continuity condition on a dielectric interface of reflection and refraction. There exist solutions to Maxwell's equations which are taken to correspond to waves which propagate at constant velocity $c = 1/\sqrt{\varepsilon_0\mu_0}$ in free space (regardless of reference frame) and with velocity $v = 1/\sqrt{\varepsilon\mu} < c$ in a dielectric (non-conducting) medium. The simplest such solutions represent plane waves, i.e. monochromatic light. Any electromagnetic field must satisfy all the four equations everywhere and reflection and refraction of light on an interface of two dielectric media (air and glass, say) is handled by fitting the solutions of Maxwell's equations in both domains together at the boundary. These boundary conditions follow from the requirement that the fields obey all Maxwell's equations everywhere. For example, *Faraday's law* for the electromagnetic induction can be written in integral form

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S}$$

where S is a complete surface spanning the loop over which the line integral on the left-hand side is taken. For this integral identity to hold smoothly in the limit at which a loop intersecting the boundary approaches to a line lying at the boundary (so that the surface integral on the right-hand side vanishes) it must be the case that the components of the electric field tangential to the interface are continuous across it. Similarly we can derive continuity conditions for the tangential component of the magnetic field intensity $\mathbf{H} = 1/\mu\mathbf{B}$, and for the normal components of both \mathbf{B} and the electric displacement field $\mathbf{D} = \varepsilon \cdot \mathbf{E}$. Imposing these continuity conditions for plane wave solutions hitting the surface of a dielectric material eventually gives us the modern Fresnel's equations, where the amplitudes are, of course, electric field amplitudes. (cf. Jackson, 1999)

In order to properly assess the correspondence between the two theories it will prove useful to go through the equations involved in broad outline.¹ For example, for a plane wave

$$\mathbf{E} = \mathbf{E}_0 e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t}$$

¹Cf. Jackson (1999), for example, for more detail.

$$\mathbf{B} = \sqrt{\mu\varepsilon} \frac{\mathbf{k} \times \mathbf{E}}{k}$$

we get the boundary conditions

$$(\mathbf{E}_0^{incident} + \mathbf{E}_0^{reflected} - \mathbf{E}_0^{refracted}) \times \mathbf{n} = 0$$

$$\left[\frac{1}{\mu} (\mathbf{k}^{incident} \times \mathbf{E}_0^{incident} + \mathbf{k}^{reflected} \times \mathbf{E}_0^{reflected}) - \frac{1}{\mu'} (\mathbf{k}^{refracted} \times \mathbf{E}_0^{refracted}) \right] \times \mathbf{n} = 0$$

for the tangential components of \mathbf{E} and \mathbf{H} , where \mathbf{n} is a unit vector normal to the interface. If the light is linearly polarised with \mathbf{E} perpendicular to the plane of incidence, these conditions yield

$$E_0^{incident} + E_0^{reflected} - E_0^{refracted} = 0$$

$$\sqrt{\frac{\varepsilon}{\mu}} (E_0^{incident} - E_0^{reflected}) \cos i - \sqrt{\frac{\varepsilon'}{\mu'}} E_0^{refracted} \cos i' = 0$$

These give as the relative amplitudes

$$\frac{E_0^{reflected}}{E_0^{incident}} = \frac{n \cos i - \frac{\mu}{\mu'} n' \cos i'}{n \cos i + \frac{\mu}{\mu'} n' \cos i'} \cong \frac{\sin(i - i')}{\sin(i + i')}$$

where we have taken $\frac{\mu}{\mu'} \cong 1$ (which holds for dielectrics and optical frequencies) and used Snell's law $n' = n \sin i / \sin i'$. Apart from the minus sign—resulting from a particular choice of orientation of \mathbf{E} and \mathbf{B} —this is equivalent to Fresnel's result (1), assuming that the electric amplitude E_0 squared is proportional to intensity, which is the case (for each medium of propagation).

In Maxwell's theory there is also a continuity equation for a quantity that is interpreted as the energy of the electromagnetic field. The flow of this quantity is represented by the so-called *Poynting vector* which depends on the electric and magnetic fields:

$$\mathbf{N} = \mathbf{E} \times \mathbf{H}$$

The interpretation of this vector as representing energy flow through unit area per unit time agrees with the idea that the energy contained in static

electric and magnetic fields (in volume V) is

$$U = \frac{1}{2} \int_V (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) d\tau$$

as well as with the continuity equation

$$\frac{\partial U}{\partial t} = - \int_S \mathbf{N} \times d\mathbf{S}$$

For plane waves we simply get

$$u = \frac{\varepsilon}{2} |E_0|^2$$

as the time-averaged energy density. It should be finally noted—and this is relevant for our later assessment of Psillos's analysis of the Fresnel-Maxwell theory shift—that this continuity equation for energy is not *as such* employed in the derivation of Fresnel's laws from Maxwell's equations (although it undeniably follows from these equations just as well).² This is in stark contrast to Fresnel's original theorising in which a continuity equation for a quantity he interpreted as kinetic energy plays a central role, as will be seen below.

8.2 Deriving Fresnel's equations

The modern derivation of Fresnel's equations was sketched in some detail above, but very little was said about the original theorising in question. The present objective is to explain, on the basis of our current best understanding of light phenomena, how Fresnel was able to derive his equations for the amplitudes of reflected polarised light. If such an endeavour manages to succeed by unveiling theoretical constituents of Fresnel's derivation which can be taken to directly reflect the unobservable reality, then there is a decent sense in which Fresnel's theoretical construction is approximately true. Many of the assumptions Fresnel made about the nature of light may have to be left outside of such explanation, which only appeals to those properties and principles that are not only of heuristic use but truly indispensable for the

²There are also some interpretation difficulties associated with the interpretation of \mathbf{N} as energy flow, cf. Lange (2002, 136 ff.).

derivation. When it comes to distinguishing these elements in practise, I whole-heartedly agree with Chakravartty who suggests that

... we must turn to the equations with which we attempt to capture phenomenal regularities, and ask: what do these mathematical relations minimally demand. We must consider not what possible metaphysical pictures are consistent with these equations, but rather what kinds of property attributions are essential to their satisfaction—i.e. to consider not what is possible, but what is required. (1998: 396)

To implement this idea of *minimal interpretation*, the following strategy is adopted: an abstracted reconstruction of Fresnel's derivation of the equation (1) is initially considered from premises which say as little as possible about the nature of light (below). It is then argued (§8.3) that the theoretical properties appealed to in this reconstructed derivation are also realised in Maxwell's theorising, as well as in a multitude of other possible theoretical constructions which all agree on the explanatory basis required for the derivation in a way that is compatible with the Maxwellian understanding sketched above. Finally, the notion of EAT is revisited in the light of this case study (§8.4).

Fresnel derived his equations from some rather elegant and simple premises.³ The molecules of elastic ether were taken to have mass, so that in oscillation they obtain a mixture of kinetic and potential energy presumably very much like a harmonic oscillator. They also obtain some momentum. The key assumption is that the maximum velocity of the oscillating ether molecules is directly proportional to the amplitude of light, which in turn is proportional to the square root of intensity. Armed with this '*minimal mechanical assumption*' (Psillos 1999: 158) Fresnel goes on to exploit a mechanical analogy of elastic collision (in which both momentum and energy are conserved) to derive the equations (1) and (2) above.

In his earlier attempt Fresnel derived (1) from the assumption of momentum conservation for *longitudinal* oscillations, but this result is mitigated by checking the results against the assumption of energy conservation. On his later attempt he took the ether oscillations to be *transverse* in character,

³Cf. Fresnel (1923). A useful secondary source is Buchwald (1989).

and he derived the equation (1) (without the minus sign) from the principle of conservation of energy (or *force vive*). A further continuity equation (or boundary condition) requiring continuity in the components of oscillation parallel to the interface yields, together with the energy equation, eventually the equations (1) and (2).

Here the original derivation is not rehearsed verbatim; rather, we proceed by deriving the equation (1) from some truly minimal metaphysical premises. This derivation is nevertheless in the spirit of Fresnel: the mathematical relationships employed are the very same, and they are motivated by physical continuity principles abstracted from Fresnel's theorising. But it is also in the spirit of Lorentz's derivation of these equations from Maxwell's theory, in that it represents a kind of common core upon which different metaphysical interpretations are then stacked.

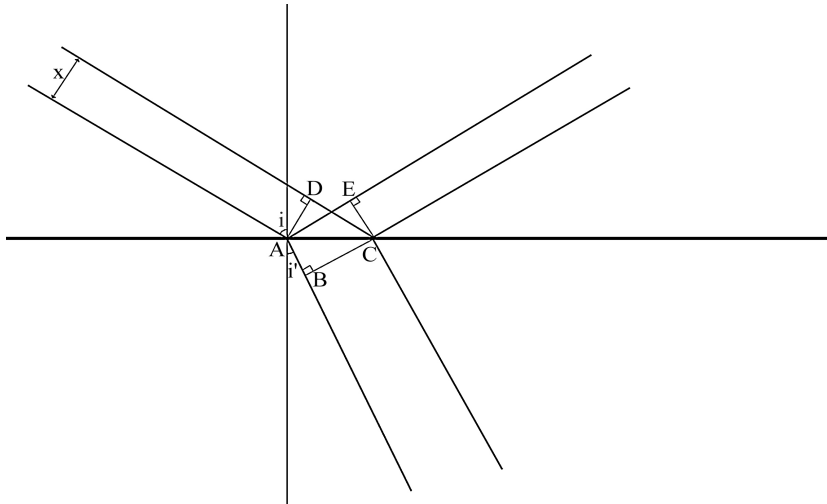


Figure 8.1: The geometrical reasoning behind Fresnel's derivation. As light front propagates the distance $y=DC$ above the interface, an area of $a_0 = \Delta(ADC)$ of the front reflects and refracts, producing areas $a_1 = \Delta(AEC) = a_0$ and $a_2 = \Delta(ABC)$, respectively.

Fresnel's Energy Equation. Let's start with the following basic assumptions. The speed of light depends on the medium of propagation. Snell's law of refraction, $n_1 \sin i = n_2 \sin i'$, quantifies refraction phenomena in terms of experimentally determined refractive indices n_i . Let's attribute quantity Q to an area a (suppressing the third dimension for simplicity) swept by a

beam (of width x) of light in some time interval Δt . Let Q depend not only of the area a , but also of some quantifiable ‘density property’ q of light which is proportional to some attribute characterising the medium of propagation and to some attribute characterising the quality of light. For example, we might take refractive index n to characterise the medium and light intensity I to characterise the quality of light. One particularly simple way to assign Q is then simply

$$Q \propto a \cdot I \cdot n^2 = a \cdot q \quad (8.3)$$

Looking at the geometry of the situation (*Figure 8.1*)—given that the speed of light depends on the medium—it is natural to assume that a_0 is the area of the beam above the interface which through the process of partial reflection and partial refraction gives rise to the reflected area of $a_1 = a_0$ and the refracted area of a_2 , respectively. Now, let’s proceed by proposing a simple *continuity equation* for Q with respect to this process. Since the incident light ‘splits into two’ it could be assigned quantity Q_0 , which is then divided into Q_1 and Q_2 , proportional to areas a_1 and a_2 , respectively. Hence the continuity in Q , over time interval Δt , area-wise, is captured as

$$Q_0 = Q_1 + Q_2 \quad (8.4)$$

From the above assumptions alone one can derive, starting with this continuity equation, *Fresnel’s energy equation*⁴

$$\sin i' \cdot \cos i \cdot (I_0 - I_1) = \sin i \cdot \cos i' \cdot I_2 \quad (8.5)$$

Fresnel, of course, derived this from the aforementioned mechanical principles. For him the conserved quantity Q was proportional to the density of the ether (molecules) and the vibration amplitude A of the ether (molecules) squared. Given the *minimal mechanical assumption* the continuity equation (8.4) clearly amounts to conservation of energy.

It is undeniable that Fresnel’s mechanical analogies provided extremely

⁴The derivation (which is very short) proceeds by first writing the ratio of the areas $a_0(=a_1)$ and a_2 in terms of the angles of incidence i and refraction i' , and then eliminating the refractive indices by using Snell’s equation.

helpful heuristics for the assumptions needed to arrive at (8.5), but even then the derivation is not a straightforward deduction. The properties of the elastic ether, as is well-known, present a somewhat curious mixture of analogies and disanalogies to the known properties of elastic solids. But even if heuristically crucial, it is obvious that these assumptions about the energy stored in some density of vibrating ether are not indispensable, for the above assumptions spelled out in much less specific terms led to the same equation! Here reference is made only to features of the continuous property of light (quantified in Q) such that Q is proportional to intensity I and the square of refractive index n^2 . These features can be satisfied by energy, perhaps, or some other character of light, for all we know. All that is observed (or directly inferred) of light are its geometrical paths and intensity and state of polarisation. As far as light is said to carry energy or momentum, for example, this should be viewed only in a hypothetical or metaphorical sense, since these ‘theoretical properties’ were in no way linked to observations at Fresnel’s time, in the sense of not having the status of Chakravartty’s *detection properties*. Nowadays we can meaningfully talk about the efficiency of solar powered machines, for example, but in the early 19th century there was obviously no way to transform the alleged mechanical energy of light into some better understood form. For sure there were characteristics of light *suggesting* it carried some form of energy—the typical warmth of the sun, say—but the point is that these characteristics were not systematically and scientifically linked to the observed regularities understood as quantifiable forms of energy.

Fresnel’s Amplitude Equations. So far there has been no mention of anything having to do with the transverse amplitude of light or its polarisation. The continuity equation thus far presented does not at all refer to the underlying mechanical character of light or what happens at the interface. To derive Fresnel’s equation (1) one needs to appeal to the transverse vectorial property \mathbf{A} (‘amplitude’) of light related to intensity so that intensity is proportional to \mathbf{A} squared. But again we need to assume *very little* about this property of light!

We want to set up another continuity equation for a particular component of a vector quantifying this property, and in effect *all* we need in order to do this is to assume that polarised light is somehow ‘spread out’ (i.e. not fully

represented by 1-dimensional rays) asymmetrically so that this spatial asymmetry is quantified by a transverse vector. *Nothing* needs to be said about what the direction in question relates to in the underlying allegedly mechanical mode of propagation. We further assume that this vectorial quantity satisfies the principle of superposition; i.e. that it can be broken down to components each of which describes a possible state of light and that these can be added together to get the original vector.

We now impose another continuity condition for the components of amplitude \mathbf{A} : the components of \mathbf{A} parallel to the interface A_{\perp} , that is, perpendicular to the plane of incidence, must satisfy ('the no-slip condition')

$$A_{\perp}^{incident} + A_{\perp}^{reflected} = A_{\perp}^{refracted} \quad (8.6)$$

That is, Fresnel demands that 'the horizontal velocity of the incident wave added to the horizontal velocity of the reflected wave must be equal to the horizontal velocity of the transmitted wave' (1923: 773, my translation). For light polarised in the plane of incidence we get, combining this with the energy equation, by elementary algebra and trigonometry⁵

$$A_{\perp}^{reflected} = \frac{-\sin(i - i')}{\sin(i + i')} A_{\perp}^{incident}$$

The equation for light polarised parallel to the plane of incidence is obtained similarly, although with some extra trigonometric manipulation to get the form (2).

Again, the heuristics used by Fresnel to come up with the 'no-slip condition' for the vector components A_{\perp} may well have been indispensable heuristically at the time, although in this case our intuitions about mechanics surely do not suggest this boundary condition as anything obvious, and the striking disparity in the treatment of the 'vertical' versus 'horizontal' velocities is a notorious weakness in Fresnel's argument. The point is, however, that we can naturally describe this continuity principle in more abstract and less specific terms: to impose this fairly natural boundary condition we only need

⁵The trick is to choose unit incident amplitude, for which we have $(1-v)^2 = (1-v)(1+v)$ on the left-hand side and $(1+v)^2$ on the right hand side of (8.5). To get (1) the trigonometric identities $\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$ are used.

to subscribe to the asymmetric polarisation of light that can be represented vector-wise; what this asymmetry consists of, that we need not say.

8.3 Comparing Fresnel to Maxwell

Fresnel's original derivation relied heavily on crucial assumptions about the geometrical configuration of light rays: that the incident, reflected and refracted light lie on a plane, and that the angle of incidence equals the angle of reflection. It also incorporated Snell's law—relating the refractive indices and the angles of incidence and refraction—which together with the geometrical reasoning encapsulated in the Figure 8.1 ties together assumptions about the speed of light, refractive indices and the two angles. By way of contrast, in the modern derivation these 'observed' facts about the behaviour of light all flow out from the field equations. But apart from the fact that the Maxwellian derivation is 'deeper' in that way, what can be said of the correspondence between the two derivations? And in particular, can the latter be used to explain the former in the way earlier alluded to, in terms of multiply realisable success-fuelling properties (cf. §7.5)?

First of all, it is clearly not the case that the modern derivation just follows the path of Fresnel's deduction with some substituted set of theoretical properties. For example, no use is made of the conservation of energy and there are not one but two boundary conditions imposed for the vectorial quantities \mathbf{E} and \mathbf{B} related through Maxwell's equations. These equations tie the two vector fields together in a way that describes a self-inducing orthogonally oscillating system of fields propagating at the speed $v = 1/\sqrt{\epsilon\mu}$. We have seen that although Fresnel also spoke of *waves* of the ether, the real work in his derivation is done by the tangential boundary condition for \mathbf{A} plugged into the continuity condition for Q , regardless of what these two quantities *exactly quantify*—as long as they relate to observations and the geometrical nature of the phenomena by the constraints $\mathbf{A}^2 \propto I$ and $Q \propto a \cdot I \cdot n^2$. So even though both Fresnel's and Maxwell's theory speak of the oscillatory nature of light, in our understanding of EAT this constitutes a kind of accidental correlation that is not to be appealed to as an explanatory correspondence.⁶ On the

⁶Explanatory approximate truth thus *diverges* here from the intuitive notion of approx-

other hand, for the realist appealing to explanatory approximate truth—for her to have her cake and eat it too—it is required that the quantities \mathbf{A} and Q are there to be found in both theories.

The claim now is that these minimal explanatory properties could be realised by different kinds of systems. *Wave theory* of light represents one such possible realisation.⁷ Understood in a wave theoretical framework (regardless of whether these waves are *further* described by Maxwell's equations and regardless of what these waves are understood to be waves *of*), the success of Fresnel's theory can be explained in a very straightforward fashion: we can point out that the boundary condition for \mathbf{A} can be effectively arrived at by purely wave theoretical reasoning based on the principle of superposition. Furthermore, we do not need to employ the geometrical reasoning (Figure 8.1 above) involving the variable speed of light, or Snell's law for that matter, to use the conservation of energy condition in conjunction with the superposition principle. Hence the mere fact that in Maxwell's theory light is understood in terms of electromagnetic waves obeying the superposition principle—manifested as linearity of Maxwell's equations, of course—can be already used to explain why the theoretical assumptions of the minimal derivation lead to the right prediction! Fresnel himself, however, did not employ such explicitly wave theoretical superposition reasoning to arrive at the boundary condition, and therefore should not be considered as propounding a properly wave theoretical understanding along these lines.

The electromagnetic theory of light represents a further step in pinning down the properties that can realise the minimal description of Fresnel's derivation. We saw above that although the electromagnetic waves have, well, a wave nature too, and although Maxwell's equations do obey the linear superposition principle as well as the energy conservation law, the derivation of Fresnel's results here is not just a matter of mimicking Fresnel by deriving one boundary condition and coupling it to energy conservation. Actually, we saw that both the wave nature of light and the energy conservation only

imate truth: whilst the latter is a matter of sufficient matching of the theoretical story and the world, *simpliciter*, the former takes into account the *explanatory value* of the matching. Fresnel's hypothesis of the oscillatory nature of light is not explanatory of his successful derivation in the required sense of being essential to it.

⁷Cf. Feynman 1964, Vol. 1:33–6 for a wave theoretical derivation of Fresnel's equations along the lines described here.

come into play in so far as these facts also undeniably follow from Maxwell's equations, and therefore the respective derivations have in fact widely different theoretical bases. But we can nevertheless easily find a vectorial property \mathbf{A} in the electromagnetic theory of light which (approximately) realises the boundary condition $A_{\perp}^{incident} + A_{\perp}^{reflected} = A_{\perp}^{refracted}$ and the conservation law $Q \propto a \cdot I \cdot n^2 = a \cdot q$ connected by $\mathbf{A}^2 \propto I$, which moreover relates \mathbf{A} to the observable intensity of light I and its state of polarisation. Looking at the sketch of the modern derivation in §8.1 we see immediately that a natural candidate for such property is the electric field \mathbf{E} . This satisfies $(\mathbf{E}_0^{incident} + \mathbf{E}_0^{reflected} - \mathbf{E}_0^{refracted}) \times \mathbf{n} = 0$ and is related to the time averaged energy density u by $u = \varepsilon/2 |E_0|^2$. The former is clearly equivalent to the boundary condition for \mathbf{A} , and the latter is directly proportional to q with good approximation—that is, when we put $n = \sqrt{\varepsilon\mu/\varepsilon_0\mu_0} \cong \sqrt{\varepsilon} \cdot \sqrt{1/\varepsilon_0}$, which holds for optical frequencies in typical dielectric matter for which $\mu/\mu_0 \cong 1$. By directly tapping into these properties of the electric vector field and the energy of electromagnetic field, Fresnel's derivation managed to predict the correct observable intensity relations in a way that has thus proved to be well short of a miracle.

8.4 Fresnel and explanatory approximate truth

How does the minimal derivation of Fresnel's equations now accord with the various accounts of EAT reviewed in the previous chapter? Beginning with Worrall, it should be obvious that the minimal derivation appeals to crucial unobservable properties and theoretical principles besides formal, logico-mathematical structure, and that we appeal to these crucial theoretical constituents in our explanation of Fresnel's derivation from the modern perspective. In terms of EAT, to say it again, it is certainly not the case that Fresnel's theory is only true about the structure, as opposed to nature, of light.

Perhaps some find it difficult to appreciate the distinction between the minimal derivation and Worrall; perhaps the minimal derivation appears *too minimal*—so much so that it threatens to collapse into triviality. The worry might be that in the derivation of the energy equation (8.5), for example, the theoretical assumptions employed are so minimal that all the substantial theoretical explanatory content just evaporates! If all there is to the 'density

property' q is expressed in terms of proportionality to broadly speaking observable attributes of light—its intensity I and refractive index n —then where is the theoretical content proper to be found? Where is the explanatory causal mechanism, for example? It may appear, that is, that the explanatory derivation proposed is *so* minimal that the anti-realist could just as well buy into it!

But actually it is not the case that *nothing* theoretical is said about Q , for example. It is not a trivial theoretical assumption to make that there is a quantifiable attribute of light which is thus distributed across space. The continuity equation (8.4) for Q expresses a property of Q that is minimal—yes—but certainly not trivial. Similarly the component-wise required continuity in \mathbf{A} expresses a higher-order property of light: whatever the asymmetry of polarised light amounts to, at the rock bottom, it satisfies the boundary condition $A_{\perp}^{incident} + A_{\perp}^{reflected} = A_{\perp}^{refracted}$. And the logically prior requirement that this asymmetry can be described in linear *vectorial* terms in the first place is already a theoretical assumption about the *nature* of light: satisfying the principle of superposition with respect to spatial components is not a matter of triviality or just a formal logico-mathematical fact about our description of a system, but rather best understood as a higher-order property.

Coming now to Psillos's account of EAT, differentiating the present proposal from his is a more subtle affair. In criticising Worrall's analysis, Psillos looks in some detail into Fresnel's derivation with the conclusion that the theoretical assumptions employed in conjunction with the minimal mechanical assumption were (1) *the principle of conservation of energy*, and (2) a *geometrical analysis* of the configuration of the light-rays in the two media. Hence, according to Psillos, there

...is no sense in which Fresnel was 'just' right about the structure of light-propagation and wrong about the nature of light, unless of course one understands 'structure' so broadly as to include the principle of the conservation of energy and the theoretical mechanism of light-propagation. ... At any rate, all of these properties of light-propagation were carried over in Maxwell's theory, even though Maxwell's theory dispenses for good with ethereal molecules. (1999: 159)

Small differences apart what we have said above seems to agree by and

large with these sentiments. Indeed, one cannot but fully agree with Psillos when he says that ‘we can clearly say that [Fresnel] was right about *some* of the fundamental properties of the light-waves, and wrong about some others’ (159).

Where the real disparity comes in is the set of conclusions drawn by Psillos regarding the most defensible form of realism *vis-à-vis* the case study in question. Psillos argues for *standard realism*, and part and parcel of that is the semantic, referential component: scientific theories are to be taken at face value and the central theoretical terms featuring in them have putative factual reference. Hence it is crucial for him to ensure that theoretical terms such as ‘ether’—which is arguably a basic term in the theoretical part responsible for the successfulness of optical ether theories—turn out in one way or another to invariably refer! Thus he goes to great pains to argue that *ether referred to the electromagnetic field*. It is part of Psillos’s realist project to show that once we adopt a particular causal-descriptive theory of reference in which referential continuity is guaranteed by ‘substantial continuity in those properties which ground the causal role attributed the posited entities’ (294)—so that a ‘term which is employed to denote the posited entity is associated with a *core causal description* of the properties by virtue of which it plays its causal role *vis-à-vis* the set of phenomena’ (295)—then it turns out that ‘luminiferous ether’ and ‘electromagnetic field’ ‘refer to the same entity precisely because their referents share the same core explanatory structure’ (297). In other words: reference of a term is fixed by ‘kind constitutive properties by virtue of which the entity denoted by the term is intended to play its causal role’.

Prima facie it seems that the idea of ‘ether’ and ‘electromagnetic field’ denoting the same class of entities is ludicrous—given what we know of the role of the ether in the transition from Galilean to Special Relativity—but it is good to spell out in some detail just how far-fetched this idea becomes once we realise how *minimal* the explanation of Fresnel’s derivation actually is.⁸ Given the foregoing derivation of Fresnel’s equation, can we really

⁸da Costa and French (2003: 169-170) acutely press the point that in taking ‘ether’ to refer to the electromagnetic field one draws an illegitimate line between the kinematical and dynamical properties as opposed to mechanical properties of the ether, a manoeuvre which obscures the significance of the transition from classical to relativistic physics. They take this as a point in favour of a version of structural realism, as opposed to standard realism with its focus on entities—if the mechanical properties are shunted off to models,

take seriously the idea that ‘the core causal description’ of the elastic ether in Fresnel’s theory, for example, just consists of spatial transverse vectorial asymmetry and a boundary condition with respect to that asymmetry in one component? Is *everything else* in the connotation of the elastic ether to be taken as merely heuristic and dumped into the heuristic models? Psillos wants to say something like that with respect to the mechanical ether models in general, but for him the core causal description of the ether has a significant dynamical component: ‘the luminiferous ether was the repository of potential and kinetic energy during light-propagation’. But we have seen that, apart from heuristics, this component plays no role in explaining the derivation of Fresnel’s law! Furthermore, that notion of mechanical energy had nothing to do with the notion of energy that is nowadays attributed theoretically to the electromagnetic fields.

Perhaps it is worth delving into this last point in a bit more detail. Whilst it is true that Fresnel appealed to a well-known energy conservation law of mechanics in his formulation of (8.4) and (8.5), it is not really the case that we can make sense of the ‘energy of the ether’ in Fresnel’s framework as anything but an *auxiliary* property—to now borrow a useful piece of terminology from Chakravartty. When speaking of the energy of the electromagnetic field, on the other hand, we have at our disposal various ways to link it to observations and other forms of energy, by virtue of which we can elevate this property of the electromagnetic field to the *detection* -category. It is quite possible of course that an auxiliary property matures into a respectable detection one as science advances. Chakravartty is fully sympathetic with this, and indeed regards the auxiliary properties of theories, as far as his *semirealism* goes, ‘not as substantive knowledge, but as methodological catalyst’ (1998: 404). A closely associated point was also made above in explicating the distinction between EAT and the intuitive notion of approximate truth: not all the conserved theoretical elements necessarily play an explanatory role and hence automatically gain the epistemic warrant that goes with it. (cf. footnote 6)

Moving now on to Chakravartty’s account, he insists on a minimal interpretation of Fresnel’s derivation much like I do.⁹

as it were, in what sense can we still say that the scientist is still referring to the ether *as an entity?*—but in my view this is going too far.

⁹Actually Chakravartty speaks of interpreting Fresnel’s *equations* (1) and (2), but the

What, then, do Fresnel's equations require? They demand some kind of influence, propagated rectilinearly and resolvable into two components, oscillating at right angles to one another and to the direction of propagation. The property or properties of light in virtue of which such influences are realized are detection properties. (1998: 396)

Here some care should be exercised, however. First of all, the minimal derivation offered above does not even go as far as speculating about the oscillatory nature of light. Secondly, the properties which fuel the derivation are exactly the density property Q and the vectorial property \mathbf{A} , defined by (8.3), (8.4) and (8.6), respectively. The properties of light in virtue of which these are realized, on the other hand, are undetermined by the derivation. A third and related point has to do with the general characterisation of these *explanatory properties*. Chakravartty emphasises the role of causal relations in his demarcation of detection properties from the auxiliary ones, and he goes as far as claiming that 'all structures of interest may be accounted for in terms of causal relations which identify specific entities' (401). But this is surely too narrow a construal of the explanatory constituents, given the case study above. The two continuity equations supplying the crucial ingredients for successful prediction are not directly causal in any straightforward sense, for example. The question of how best to characterise these properties in general terms is taken up in the next chapter.

real point of focus is, of course, the derivation of these.

Towards Eclectic Realism

This thesis began with a critical discussion of explanationism, the idea that one can justify realism about unobservables by an argument that capitalises on the explanatory aspect of science. After advocating more local and lower-level inductive arguments in this regard, I left justificatory considerations behind and moved on to the context of the realist image. Explanatory issues have assumed a central position in this latter discussion as well, but in this context I have been critical of the *lack of* explicit explanatory considerations in the realist arguments! It seems that philosophers arguing for a realist image have failed to acknowledge the full significance of the notion of *philosophical explanation* for the realist project. To draw attention to this I termed the crucial notion ‘explanatory approximate truth’.

But it is not the case that these realists have forgotten the importance of explanatory considerations *in science*. Indeed, appealing to the explanatory dimension of science as extra-empirical evidence is the standard response to the challenge from empirical underdetermination (§4.2), and a description of such ‘explanatory evidence’ is part and parcel of the realist image. My critique of explanationism in Part I was (roughly speaking) based on the plurality of explanations that are only ostensibly unified by the explanationist arguments. I now want to consider a somewhat similar worry in the context of the realist image. But I will argue that this worry can be dealt with by adopting an appropriate conception of EAT, the kind of conception that I have begun to develop in the previous two chapters.

9.1 From scientific explanation to scientific understanding

It is part of the realist image to describe how we grasp the truth-values of our theories. Delineating a criterion of success purely in terms of *predictive success* is not enough—due to the underdetermination challenge—and the realist must incorporate explanatory, ‘super-empirical’ epistemic virtues.¹ But explanatory values arguably shift across the history of science, as well as across the concurrent scientific domains, creating a threat of a kind of PMI-reductio against the realist image. For example, if action-at-a-distance explanations in science are valued as providing understanding at one time but devalued at another time, there seems to be a level arbitrariness, or contextuality, in the *de facto* super-empirical virtues of scientific practice.² The realist image should capture such contextuality, and depending on how radical such shifts are in the actual science we can either have a plausible piecemeal image, with a list of possible explanatory virtues, or a less plausible one with gerrymandered, perhaps even inconsistent virtues.

De Regt & Dieks (2005) make a case for the diversity of scientific explanations, and (consequently) for pluralism in the philosophical conceptions of explanatory understanding. In addition to the action-at-a-distance example they cite other instances of historical variation in the explanatory values. Moreover, they argue that in the contemporary science the explanatory preferences can vary from one theoretical domain to another, and the existence of non-causal explanations, in particular, speaks against the universality of the causal conception of explanation.³ It seems that by and large realists have failed to take notice of such plurality, since the realist arguments, especially those with global ambitions, typically proceed by assuming that the explanatory values operating in science are, and have been, relatively constant. In other words, to make a link to the discussion of IBE in sections §1.5 and

¹Doppelt (2005) nicely makes the point that the realist’s standard of success must include explanatory success.

²After being criticised for not conforming to the Cartesian intelligibility ideal of contact action, Newton’s theory of gravitation paved the way for *actio in distans* explanations in the pre-ether theories. (De Regt & Dieks, 2005: 146)

³Consider, for example, mechanistic explanations in cognitive science (Wright & Bechtel (forthcoming)), a quantum mechanical explanation of the EPR-phenomena, or a geometrical explanation of some general relativistic phenomenon.

§3.3.2, De Regt & Dieks draw our attention to some manifest evidence for my earlier contention that the two parameters of the IBE schema—what counts as an explanation and what counts as the loveliest explanation—are highly variable. And it is not only that there is a significant underlying disparity between the scientific and philosophical instances of IBE, affecting the justificatory force of explanationism (§1.5), but there is enough disparity between the scientific instances of IBE themselves to threaten the project of the realist image as well. Let us call this the *diversity of explanations* -problem. To ensure the plausibility of the realist image—and in any case if unification is desired by virtue of following the global intuition behind the No-Miracles Argument—it is worth asking whether we could say something general about explanation that would reveal the required unity in the explanatory practice of science.

De Regt & Dieks (2005) articulate a unifying notion of *understanding* that answers to the crucial question: ‘By virtue of what is an explanation *E* explanatory?’. Two of the main features of their account, both plausible and intuitive, are as follows.

Understanding as an ‘*Inextricable Element of Science*’. The relationship between theories and phenomena is not one of logical deduction, but a more complex affair. There is no algorithm to follow to get from theoretical suppositions to empirical results; only by understanding the theory (in a sense to be clarified) can one arrive from the former to the latter. In particular, one needs to know how to *use* the theory, to derive predictions or descriptions of the phenomenon in question. Acknowledging this introduces the pragmatic dimension of *skills*, or *abilities*.

Intelligibility of theories is needed because scientists have to be able to use theories in order to generate predictions and explanations. (2005: 150)

Limitations in the scientists’ skills can delimit the theoretical solutions available at the time. If these abilities shift over time, or across social boundaries, understanding becomes contextual.

Understanding as *Contextual*. As a matter of historical fact, the intelligibility standards have not remained valid throughout the history of scientific thinking. Also, with regard to contemporary science it is not the case that a single mode of explanation can claim the status of objectively ‘best’ explanation.

[The analysis] should reflect the actual (contemporary and historical) practice of science. It should therefore allow for variation in standards of understanding. (2005: 149)

So there is unity to understanding at one level, for it is an inextricable element of science, but there is a degree of contextuality involved at another level. But in order to say something unifying about explanation in this way, we need to be able to characterise understanding directly, and not just as ‘what explanations provide us’. De Regt & Dieks (2005) propose the following *Intelligibility Criterion*:

Scientific theory T is intelligible for scientists (in context C) if they can recognise qualitatively characteristic consequences of T without performing exact calculations. (2005: 151)

The intuition behind this proposal is that what one wants and needs in science is the ability to grasp how the predictions are brought about by the theory. Only if deriving empirical results from a theory was a matter of algorithmic logical deduction, such understanding would not be required. But as a matter of fact, with regard to a derivation of some result for example, we need to grasp why each derivational move is committed. Basically this just amounts to the rather plausible idea that scientists don’t just try out *arbitrary* moves in a derivation, and follow a qualitatively expressible reasoning as they follow a derivation.

I think this very general characterisation nicely captures the inextricability of understanding, whilst leaving room for variation and contextuality in the ways that understanding can be achieved. Also, it captures well the fact that the causal dimension is so significant in scientific reasoning: in many situations we learn to instinctively follow causal reasoning to see what the consequences of some assumptions are. But it also leaves room for other

‘tools’ of understanding: there can be dependence relations other than the causal one structuring this world, and with practice we learn to follow these other ones as well (even if they are not quite as transparent as the causal one).⁴

Returning now to the problem of diversity of explanations, how is it that the account of explanatory understanding sketched above can help the realist cause? What the De Regt-Dieks intelligibility criterion gives us is a constraint for the kind of abductive reasoning that scientists are engaged in. This is something that the general IBE-template, characterising an abductive inference as an inference to the explanation that provides the most understanding, is silent about. So when it comes to spelling out one’s realist image, and characterising the notion of explanatory success in particular, the intelligibility criterion offers a way to sidestep the diversity of explanations -problem by linking explanatory success of *T* to the scientists’ ability to recognise qualitatively characteristic consequences of *T*. This is what the diverse modes of explanations providing the most understanding have in common.⁵

If the intelligibility seeking in this sense is a general, over-arching feature of the explanatory methodology of science, then perhaps we can frame the realist image directly in terms of it. As a first-order approximation we can say that the realist puts forward the descriptive claim that the understanding-generating abductive practice of theoretical science is truth-conducive. The anti-realist is quick to respond, of course, by pointing out cases in which theoretical understanding was *prima facie completely false*, by the current lights. Clearly this understanding-seeking practice is not tracking the truth since we have all the historical cases which show that we did not *really* understand *the phenomenon* at all—that our best explanations in terms of *ether*, or *caloric*, say, were completely wrong—although we did understand the *theory*. And furthermore, even if appealing to material ether was indispensable for Fresnel’s understanding of light phenomena in his socio-historical context, we can still nowadays follow Fresnel’s derivation and understand, in the sense of De

⁴Cf. Ruben (1993: 10 ff.)

⁵The intelligibility criterion should be thought as *supplementing*, not supplanting, the descriptive IBE model of ampliative scientific reasoning. Perhaps the criterion can be modelled as an invariable background constraint for an interpretation to provide the most understanding.

Regt & Dieks, how one step follows another in his theorising!

But we now know how to begin to eradicate the root of such pessimism, by appealing to explanatory approximate truth. I will next elaborate on this notion by framing it in terms of philosophical reductive explanation. This yields a conception of approximate truth that allows us to drive apart the notion of understanding a *theory* (in De Regt-Dieks sense), on one hand, and the notion of understanding a *phenomenon*, on the other.

9.2 Approximate truth and reductive explanation

The discussion above about the lessons to be drawn from the Fresnel-Maxwell case (chapters §7 and §8) revolved around the two opening questions: (1) What does it take to explain a particular success of science? (2) If appeal to *some* theoretical content is indeed required as the explanans (as the realist argues), then on exactly what *principled grounds* should this content be delineated? These questions are interlinked and different responses result in a variety of realist positions. Answering the second question, in particular, determines a notion of explanatory approximate truth. I argued that a more open-minded characterisation of the realist commitment is desirable in this context. We need to acknowledge that the realist endeavour of providing a *philosophical* explanation of the success of past science is independent of the scientists' endeavour of providing a *scientific* explanation of some phenomenon; in particular, there is no need to insist on spelling out the former, logical explanation in causal terms. In this penultimate section the inchoate suggestions above that the realist commitment should be given in terms of multiply realisable theoretical properties is elaborated in terms of *reductive explanation*.

It is useful for starters to call forth some intuitions about approximate truth of an explanation. The idea that an explanation can have *non-explanatory surplus content* is familiar enough at everyday level. Consider explaining why a philosophy student Owen suddenly begins to express optimism about the prospects of conceptual analysis, by suggesting that he has recently read Frank Jackson's book *From metaphysics to ethics*. This suggestion may be false—and hence disqualify as an (actual) explanation as such—yet contain a significant seed of truth: perhaps Owen has recently read another book of Jackson's in which similar ideas were entertained. Or consider explaining why

a crammed elevator refuses to move by suggesting that there is 50 kilograms of excess weight on board. This is likely to be strictly speaking false, and hence disqualified as an (actual) explanation, yet it may contain a significant seed of truth: perhaps there is indeed too much weight on board, but only some 44 kilograms, say. Such everyday examples strengthen the intuition that a false story does not have to be explanatorily *empty*. In both cases above there is a clear sense in which a *less specific* story would have captured what is true in the explanatory proposals, and thus would qualify as an explanation (although an equally specific and fully true explanation also exists, of course). One could have explained Owen's behaviour by the fact that he has recently read *one of the works* by Jackson in which conceptual analysis is defended, for example. It is also clear that such an explanation can be compatible with a multitude of more detailed stories which are incompatible with each other: perhaps Owen has only read his personal copy of the book, or perhaps he consulted a library copy, etc.

Now consider my derivation of Fresnel's equations from the minimal assumptions in section §8.2. The properties that turned out to be minimally required for the derivation—the properties that are common to both Fresnel's and Maxwell's theory—furnish a *minimal explanation* of the corresponding phenomenon. Obviously this is not the explanation of light reflection and refraction that Fresnel (or Maxwell) had in mind, but it is an explanation nevertheless. But such minimal explanation may not be the best, the most transparent and the 'loveliest' explanation of the phenomenon in some context of theorising. The minimal, success-fuelling properties of Fresnel's derivation are 'the less specific story' that captures what is true in Fresnel's explanation, and the minimal scientific explanation in terms of these properties is compatible with a multitude of stories about the lower-level facts.

Corresponding to such minimal explanation of the phenomenon, there is a reductive philosophical explanation of the success of Fresnel's theorising. This can be understood through *multiple realisability of properties*: the explanatory ingredients are properties identified by their causal-nomological roles, and such properties are multiple realisable in the sense that they are instantiated by virtue of having some other lower-order property (or properties) meeting certain specifications, and the higher-order property does not uniquely fix the lower-order one(s). This 'layers of reality' -conceptualisation

of properties is taken to be fully uncontroversial here. It is common place that many macro-realm properties are multiply realisable in this sense—e.g. ‘being a pen’ designates a property instantiated by many types of pen—and multiple realisation arguments form the locus of a central debate between the reductionists and non-reductionists in the philosophy of mind, for example. But multiple realisation is clearly not confined to states of cognition, or the very high-level artefact functional kinds.⁶ Some of the properties are multiply realised in the actual world, whereas others have alternative realisations only across possible worlds. In particular, there is a natural sense in which the properties involved in the minimalist explanation of Fresnel’s successful prediction are unmistakably multiple realisable, even if only in a restricted modal sense.

Consider again the different theoretical accounts of light. I think the following perspective is naturally affiliated to the foregoing. A theory (such as Fresnel’s) is put forward as an account of how *possible* low-level facts (the oscillating mechanical ether molecules) would entail the explanandum phenomenon.⁷ There are typically two parts to such account: there are (1) higher-order multiply realisable explanatory properties, and (2) a set of lower-order properties representing a possible realisation of the latter. Often, of course, these two parts are subtly interwoven, given the theoretical understanding of the time. It is no trivial matter to disassociate the two; it took almost a century for us to begin to see how non-mechanical properties could underlie the explanatory theoretical properties associated with the ether. If the theory is logically consistent⁸ and fully compatible with a given body of evidence,

⁶Consider the higher-order property designated by ‘being a primary colour’, for example. This is realized by the properties of being red, being blue and being yellow. More generally, the relationship between some determinable property and the corresponding set of determinates can be viewed from this perspective; there are many ways of being (spatially) asymmetric, for example, or having the mass of nine grams. (Clapp, 2001; Yablo, 1992) More pertinently, the property designated by *Q* in our reconstruction of Fresnel’s derivation, defined by its continuity over the process of reflection/refraction and direct proportionality to intensity, is multiply realisable in the modal sense of having epistemically possible realisations.

⁷In actual practice the phenomenon to be explained may not be *fully* known at the time of theorising, of course, but become thus acknowledged only after its *prediction* based on some *less* detailed story of the phenomenon. The initial explanandum is then formed by whatever less detailed high-level facts must be accommodated by the theory.

⁸It was noted above (§8.2) that Fresnel’s theoretical assumptions were not altogether

then one can think of the theoretical description as a description of an *epistemically possible scenario* about which the theory is true—a way the world might be for all we know.⁹ When the description of the explanandum phenomenon is sharpened as new evidence is incorporated it often happens that a particular epistemically possible scenario is no longer such: the low-level properties of the original proposed theory cannot realise the higher-level properties involved in the revised explanandum. A new theory is proposed and accompanied by a new, more fine-grained division of potential epistemically possible scenarios. By successive iteration science proceeds and gets closer to the ideal division of epistemically possible scenarios compatible with all the available evidence, ever.

But how should the realist apprehend such an iterative process of theoretical development in the face of the PMI challenge? How can the understanding generating abductive practice of science be taken to be truth conducive, if many of our best explanations did not yield actual understanding of the phenomena in question, and only relate to our current understanding via a minimal explanation that was not the best explanation of the phenomena (in the context of theorising)?

9.3 Towards a novel formulation of realism

Fresnel's understanding of light as transversal oscillations of ether was misguided. In my account of the approximate truth of Fresnel theory I decidedly pushed the causal image of vibrating ether molecules into heuristics, to ex-

consistent, given the knowledge of elastic forces at the time. What is inconsistent, to be exact, is his appeal to purely transverse fluctuations in the *elastic* ether, where referring to the property of 'elasticity of ether' carries certain conceptions drawn from study of elastic materials. The minimalist core of Fresnel's derivation above is obviously not logically inconsistent as such.

⁹'The way the world might be for all we know' is called an *epistemically possible world*, as opposed to a counterfactually possible world. The theory of ether does not describe a counterfactually possible world if natural kind terms are considered to be rigid designators: there is no counterfactually possible world in which our term 'light' does not denote certain kinds of electromagnetic oscillations, assuming that this is what 'light' denotes in the actual world. This familiar Kripkean point about rigid designation does nothing to undermine the usefulness of the very clear intuitive notion of epistemically possible world in the context of the present argument. More formal presentation would require the use of some form of two-dimensional semantics. (cf. Chalmers, 2004)

tract a description of the minimal explanatory properties that were crucial for the mathematico-logical derivation. But arguably this causal picture of ether was indispensable for Fresnel's understanding, so how can I justify dismissing it just as heuristics?

Appealing to ether was indispensable for Fresnel in the sense that it allowed him to reason qualitatively about the properties that he wrote down in the equations of his derivation. That made his causal ether explanation of the optical phenomenon of reflection and refraction of polarised light the loveliest explanation, in his social and historical context. The inclusion of material ether made the theory intelligible for Fresnel, in the sense of De Regt & Dieks. But these *contextual factors* can be accommodated in the reductive explanation of Fresnel's success in terms of multiply realisable success-fuelling properties. For these are just particular epistemically possible lower-order realisations of the latter, and reasoning about these lower-order properties can help to see the qualitative consequences of the higher-order properties appearing in the actual mathematico-logical derivation. Rather than creating a difficulty for the realist, I maintain that spelling out the notion of explanatory approximate truth in these terms actually helps us to come to grips with how understanding in the De Regt-Dieks sense can function as an inextricable element of IBE-driven science.

The preceding discussion of non-explanatory surplus content and multiply realisable success-fuelling properties, together with the minimalist interpretation of Fresnel's derivation, now manifestly points towards a novel realist position. To an anti-realist who picks out the Fresnel example as a case par excellence of the roots of her pessimism, we must respond by questioning her notion of approximate truth. The notion of approximate truth appropriate for the realist—the notion of *explanatory approximate truth*—can and often does diverge from the naïve notion the anti-realist intuition relies on. Analysing the nature of the theoretical constituents involved in a scientific explanation, and acknowledging the hierarchy of properties appealed to, makes room for a more fine-grained explanation of a theoretical success. Failing to do that has led realists astray, to make huge efforts in the semantics of theoretical terms in order to prove the anti-realist wrong but still remain true to the intuitive sense of approximate truth which demands that 'ether' refers to something.

We need not worry about 'ether' being non-referring *exactly because* it is

actually *not* a central term in the right explanatory sense! If ‘central’ *just means* ‘denoting an entity the existence of which is required for the realist explanation of success’, then this conclusion follows because the existence of the ether is *not* required in that explanation. What is required is that there is a common core of theoretical properties appealed to in both Fresnel’s and the corresponding modern day theorising. The phenomenon of light under theorising is minimally explained by these properties: they are described by their causal-nomological roles in the respective theories via the boundary and symmetry conditions equally present in both derivations, and they entail the phenomenon of reflection/refraction of polarised light. But there being unobservable properties thus described and related to the observations in question is no conceptual truth about light: these properties really are eligible for carrying the explanatory realist commitment. This is all that needs to be said to the anti-realist.

More can be said, however, about the way in which theorising proceeds in iterative fashion through successive falsities towards the truth. Tracking the nature of entities and processes underlying the appearances has turned out to be a challenge stretching human imagination to its limits in trying to explain phenomena by concepts increasingly far removed from the familiar. It is no wonder that in an attempt to conceptualise strange properties underlying some phenomenon more than enough is often said; it is clear that it frequently helps to focus on a possible lower-order realisation of the higher-order properties employed in a derivation. The most intelligible theory, in the sense of allowing a scientist equipped with particular abilities to draw qualitative consequences from it, can well feature such lower-order properties and often, consequently, non-explanatory surplus content. But these lower-order properties typically entail derivationally crucial properties that are multiply realisable. The art of finding the crucial properties and the actual explanatory content, and retaining it whilst discarding or replacing the rest, is to be found in the science itself, not philosophy of science. For the latter, and for realism in particular, it is enough to describe a principled discriminatory *framework* to describe and explain the successfulness of this scientific practice. This last Part has put forward the beginnings of a novel description of such a framework.

In celebration of the aforementioned art of discerning the explanatory,

success-fuelling constituents, such a framework could perhaps be called *eclectic realism*. The true measure of it, of course, only comes through a great variety of case studies and only prudent optimism can be expressed in the meanwhile. But when it comes to the highly frequented Fresnel-Maxwell case—the adopted battle ground for Psillos’s standard realism and Worrall’s structural realism, and widely commented on by others—it can be safely concluded that some *philosophical eclecticism* is rightfully needed to reach a defensible middle ground. Explicating the notion of explanatory approximate truth, and working explicitly at the level of properties as suggested by Chakravartty, and using this to take a closer-look at the details of Fresnel’s derivation, shows how there is indeed more than structural continuity in the theory change from Fresnel to Maxwell, yet not enough to warrant the standard formulation of realism with its insistence on the successful reference of ‘ether’.

BIBLIOGRAPHY

- Achinstein, Peter (1992). Inference to the best explanation: Or, who won the Mill-Whewell debate? *Studies in History and Philosophy of Science*, **23**, 349–362.
- Achinstein, Peter (2001). *The Book of Evidence*. Oxford University Press, New York.
- Achinstein, Peter (2002). Is there a valid experimental argument for scientific realism? *Journal of Philosophy*, **99**(9), 470–495.
- Barnes, Eric (1995). Inference to the loveliest explanation. *Synthese*, **103**, 251–278.
- Bishop, Michael (2003). The pessimistic induction, the flight to reference and the metaphysical zoo. *International Studies in the Philosophy of Science*, **17**(2), 161–178.
- Bishop, Michael A. and Stich, Stephen P. (1998). The flight to reference, or how "not" to make progress in the philosophy of science. *Philosophy of Science*, **65**(1), 33–49.
- Boyd, Richard (1981). Scientific realism and naturalistic epistemology. In *PSA 1980* (ed. P. Asquith and R. Giere), Volume 2, pp. 613–662. East Lansing, MI: Philosophy of Science Association.
- Boyd, Richard (1990). Realism, approximate truth, and philosophical method. In *Minnesota Studies in the Philosophy of Science* (ed. C. W. Savage), Volume 14, pp. 355–91. Minnesota University Press, Minneapolis.

- Buchwald, Jed Z. (1989). *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century*. The University of Chicago Press, Chicago.
- Bueno, Otávio (1999). What is structural empiricism? Scientific change in an empiricist setting. *Erkenntnis*, **50**(1), 59–85.
- Bueno, Otávio (2006). Representation at the nanoscale. *Philosophy of Science* (5).
- Carnap, Rudolph (1966). *The Philosophical Foundations of Physics*. Basic Books, New York.
- Cartwright, Nancy (1983). *How the Laws of Physics Lie*. Clarendon Press, Oxford.
- Chakravartty, Anjan (1998). Semirealism. *Studies in History and Philosophy of Science*, **29A**(3), 391–408.
- Chakravartty, Anjan (2004). The structuralist conception of objects. *Philosophy of Science*, **70**(5), 867–878.
- Chalmers, David (2004). Epistemic two-dimensional semantics. *Philosophical Studies*, **118**, 153–226.
- Churchland, P.M. (1985). The ontological status of observables: In praise of the superempirical virtues. In *Images of Science* (ed. P. Churchland and C. Hooker), pp. 35–47. Chicago University Press, Chicago.
- Churchland, P.M. and Hooker, C.A. (ed.) (1985). *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen*. Chicago University Press, Chicago.
- Clapp, Leonard (2001). Disjunctive properties: multiple realizations. *Journal of Philosophy*, **98**, 111–136.
- Clarke, Steve (2001). Defensible territory for entity realism. *The British Journal for the Philosophy of Science*, **52**, 701–722.

- Cruse, Pierre (2004). Scientific realism, Ramsey-sentences, and the reference of theoretical terms. *International Studies in the Philosophy of Science*, **18**, 133–149.
- daCosta, N. and French, S. (2003). *Science and Partial Truth*. Oxford University Press, Oxford.
- Day, T. and Kincaid, H. (1994). Putting inference to the best explanation in its place. *Synthese*, **98**, 271–295.
- De Regt, Henk W. and Dieks, Dennis (2005). A contextual approach to scientific understanding. *Synthese*, **144**, 137–170.
- Demopoulos, William (2003a). On the rational reconstruction of our theoretical knowledge. *British Journal for the Philosophy of Science*, **54**(3), 371–403.
- Demopoulos, William (2003b). Russell’s structuralism and the absolute description of the world. In *The Cambridge Companion to Russell* (ed. N. Griffin), pp. 392–419. Cambridge University Press, Cambridge.
- Demopoulos, William and Friedman, Michael (1985). Russell’s analysis of matter: Its historical context and contemporary interest. *Philosophy of Science*, **52**, 621–639.
- Devitt, Michael (1991). *Realism and Truth* (2nd edn). Oxford University Press, Oxford.
- Devitt, Michael (2002). Underdetermination and realism. *Nôus*, **36**(1 Supplement), 26–50.
- Doppelt, Gerald (2005). Empirical success or explanatory success: What does current scientific realism need to explain. *Philosophy of Science*, **72**(5). Forthcoming.
- Dorato, Mauro and Pauri, Massimo (2006). Holism and structuralism in classical and quantum gr. In *Structural Foundations of Quantum Gravity*. Oxford University Press, Oxford.

- Dorling, Jon (1992). Bayesian conditionalization resolves Positivist/Realist disputes. *Journal of Philosophy*, **92**, 162–182.
- Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Princeton University Press, Princeton, NJ.
- Earman, John (1993). Underdetermination, realism and reason. *Midwest Studies in Philosophy*, **18**, 19–38.
- Fine, Arthur (1984). The natural ontological attitude. In *Scientific Realism* (ed. J. Leplin), pp. 83–107. University of California Press, Berkeley.
- Fine, Arthur (1986). Unnatural attitudes: Realist and instrumentalist attachments to science. *Mind*, **95**, 149–179.
- Fine, Arthur (1991). Piecemeal realism. *Philosophical Studies*, **61**(1-2), 79–96.
- Forrest, Peter (1994). Why most of us should be scientific realists: A reply to van Fraassen. *Monist*, **77**(1), 47–70.
- French, Steven (1989). Identity and individuality in classical and quantum physics. *Australasian Journal of Philosophy*, **67**, 432–446.
- French, Steven (1998). On the withering away of physical objects. In *Interpreting Bodies: Classical and Quantum Objects in Modern Physics* (ed. E. Castellani), pp. 93–113. Princeton University Press, Princeton.
- French, Steven and Krause, Décio (2006). *Identity and Individuality in Modern Physics*.
- French, Steven and Ladyman, James (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science*, **13**, 103–121.
- French, Steven and Ladyman, James (2003). Remodelling structural realism: Quantum physics and the metaphysics of structure. *Synthese*, **136**(1), 31–56.
- French, Steven and Rickles, Dean (2003). Understanding permutation symmetry. In *Symmetries in Physics: Philosophical Reflections* (ed. K. Brading and E. Castellani), pp. 212–238. Cambridge University Press, Cambridge.

- French, Steven and Rickles, Dean (2006). Quantum gravity meets structuralism: Interweaving relations in the foundations of physics. In *Structural Foundations of Quantum Gravity*. Oxford University Press, Oxford.
- French, Steven and Saatsi, Juha T. (2006). Realism about structure: Semantic view and non-linguistic representation. *Philosophy of Science*, **72**.
- Fresnel, A. (1923). Mémoire sur la loi des modifications que la réflexion imprime à la lumière polarisée. In *Oeuvres Complètes D'Augustin Fresnel* (ed. H. de Senarmont, E. Verdet, and L. Fresnel), Volume 1, pp. 767–799. Imprimerie Impériale, Paris.
- Gettier, Edmund L. (1963). Is justified true belief knowledge? *Analysis*, **23**, 121–123.
- Hacking, Ian (1981). Do we see through a microscope? *Pacific Philosophical Quarterly*, **62**, 305–322. Reprinted in *Images of Science*. (Eds) P.M. Churchland and C.A. Hooker.
- Hacking, Ian (1982). Experimentation and scientific realism. In *Philosophy of Science: The Central Issues* (ed. M. Curd and J. Cover), pp. 1153–1168. W.V. Norton and Company, 1998, New York.
- Hacking, Ian (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, Cambridge.
- Hardin, C. and Rosenberg, A. (1982). In defense of convergent realism. *Philosophy of Science*, **49**, 604–615.
- Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, **74**, 88–95.
- Hartmann, Stephan (2002). Essay review: On correspondence. *Studies in History and Philosophy of Modern Physics*, **33**, 79–94.
- Hempel, C.G. (1965). *Aspects of Scientific Explanation*. Free Press, New York.
- Hitchcock, Christopher Read (1992). Causal explanation and scientific realism. *Erkenntnis*, **37**(2), 151–178.

- Howson, Colin (2000). *Hume's Problem: Induction and Justification of Belief*. Oxford University Press, Oxford.
- Huggett, Nick (1997). Identity, quantum mechanics and common sense. *The Monist*, **80**, 118–130.
- Jackson, J. (1999). *Classical Electrodynamics* (3rd edn). Wiley, New York.
- Josephson, John R. (2000). Smart inductive generalisations are abductions. In *Abduction and Induction: Essays on their Relations and Integration* (ed. P. Flach and A. Kakas). Kluwer Academic Publishers, Dordrecht.
- Josephson, John R. and Josephson, Susan G. (1996). *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press, Cambridge.
- Ketland, Jeffrey (2004). Empirical adequacy and ramsification. *British Journal for the Philosophy of Science*, **55**, 287–300.
- Kitcher, Philip (1993). *The Advancement of Science*. Oxford University Press, Oxford.
- Kitcher, Philip (2001). Real realism: the Galilean strategy. *The Philosophical Review*, **110**(2), 151–197.
- Kitcher, Philip and Salmon, Wesley C. (1987). van Fraassen on explanation. *Journal of Philosophy*, **84**, 315–330.
- Kukla, André (1998). *Studies in Scientific Realism*. Oxford University Press, Oxford.
- Kukla, André and Walmsley, Joel (2004). A theory's predictive success does not warrant belief in the unobservable entities it postulates. In *Contemporary Debates in Philosophy of Science* (ed. C. Hitchcock), pp. 133–148. Blackwell, Oxford.
- Ladyman, James (1998). What is structural realism? *Studies in History and Philosophy of Science*, **29A**(3), 409–424.

- Ladyman, James (2000). What's really wrong with constructive empiricism? van Fraassen and the metaphysics of modality. *British Journal for the Philosophy of Science*, **51**(4), 837–856.
- Ladyman, James (2004). Constructive empiricism and modal metaphysics: A Reply to Monton and van Fraassen. *The British Journal for the Philosophy of Science*, **55**(4), 755–765.
- Ladyman, J., Douven, I., Horsten, L., and van Fraassen, B. (1997). In defence of van Fraassen's critique of abductive reasoning: A reply to Psillos. *Philosophical Quarterly*, **47**, 305–321.
- Lange, Marc (2002a). Baseball, pessimistic inductions and the turnover fallacy. *Analysis*, **62**(4), 281–285.
- Lange, Marc (2002b). *An Introduction to the Philosophy of Physics*. Blackwell, Oxford.
- Laudan, Larry (1981). A confutation of convergent realism. *Philosophy of Science*, **48**, 19–49.
- Laudan, Larry (1996). *Beyond Positivism and Relativism: Theory, Method and Evidence*. Westview Press, Boulder, CO.
- Laudan, Larry and Leplin, Jarrett (1991). Empirical equivalence and underdetermination. *Journal of Philosophy*, **88**, 449–472.
- Leplin, Jarrett (1997). *A novel defence of scientific realism*. Oxford University Press, Oxford.
- Leplin, Jarrett (2000). Realism and instrumentalism. In *A Companion to the Philosophy of Science* (ed. W. H. Newton-Smith), pp. 393–401. Blackwell, Oxford.
- Lewis, David (1970). How to define theoretical terms. *Journal of Philosophy*, **67**, 427–446.
- Lewis, David (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, **50**, 249–58.

- Lewis, Peter J. (2001). Why the pessimistic induction is a fallacy. *Synthese*, **129**(3), 371–380.
- Lipton, Peter (1990). Contrastive explanation. In *Explanation and its Limits* (ed. D. Knowles), pp. 247–266. Cambridge University Press, Cambridge.
- Lipton, Peter (1993). Is the best good enough? *Proceedings of the Aristotelian Society*, **93**(2), 89–104.
- Lipton, Peter (2004). *Inference to the best explanation* (2nd edn). Routledge, London.
- Lorentz, H. (1875). Sur la théorie de la réflexion et de la refraction de la lumière. In *Collected Papers* (ed. P. Zeeman and A. Fokker), Volume 1, pp. 193–383. Nijhoff, The Hague.
- Lyre, Holger (2004). Holism and structuralism in U(1) gauge theory. *Studies in History and Philosophy of Modern Physics*, **35**, 597–624.
- Magnus, P.D. (2003). Success, truth, and the Galilean strategy. *British Journal for the Philosophy of Science*, **54**(3), 465–474.
- Magnus, P.D. and Callender, Craig (2004). Realist ennui and the base rate fallacy. *Philosophy of Science*, **71**(3), 320–338.
- Maxwell, Grover (1966). Scientific methodology and the causal theory of perception. In *Problems in the Philosophy of Science* (ed. I. Lakatos and A. Musgrave). North-Holland Publishing Co., Amsterdam.
- Maxwell, Grover (1970a). Structural realism and the meaning of theoretical terms. In *Analysis of Theories and Methods of Physics and Psychology: Minnesota Studies in the Philosophy of Science* (ed. S. Winokur and M. Radker), Volume IV, Minneapolis, pp. 181–192. University of Minnesota Press.
- Maxwell, Grover (1970b). Theories, perception and structural realism. In *Nature and Function of Scientific Theories* (ed. R. Colodny), pp. 3–34. University of Pittsburgh Press, Pittsburgh.

- McMullin, Ernan (1984). A case for scientific realism. In *Scientific Realism* (ed. J. Lepplin). University of California Press, Berkeley.
- McMullin, Ernan (1987). Explanatory success and the truth of theory. In *Scientific Inquiry in Philosophical Perspective* (ed. N. Rescher). University Press of America, Lanham.
- McMullin, Ernan (1994). Enlarging the known world. In *Physics and Our View of the World* (ed. J. Hilgevoord). Cambridge University Press, Cambridge.
- McMullin, Ernan (2003). Van Fraassen's unappreciated realism. *Philosophy of Science*, **70**(3), 455–478.
- Melia, Joseph W. and Saatsi, Juha T. (2006). Ramseyfication and theoretical content. *British Journal for the Philosophy of Science*. Forthcoming.
- Menuge, A. (1995). The scope of observation. *Philosophical Quarterly*, **45**, 60–69.
- Miller, Richard W. (1987). *Fact and Method : Explanation, Confirmation and Reality in the Natural and the Social Sciences*. Princeton University Press, Princeton.
- Müller, F.A. (2004). Can a constructive empiricist adopt the concept of observability? *Philosophy of Science*, **71**(1), 80–97.
- Müller, F.A. (2005). The deep black sea: Observability and modality afloat. *British Journal for the Philosophy of Science*, **56**, 61–99.
- Musgrave, Alan (1988). The ultimate argument for scientific realism. In *Relativism and Realism in Science* (ed. R. Nola), pp. 229–252. Kluwer Academic Publishers, Dordrecht.
- Newman, M.H.A. (1928). Mr Russell's causal theory of perception. *Mind*, **37**, 137–148.
- Nola, Robert (1980). Fixing the reference of theoretical terms. *Philosophy of Science*, **47**, 505–531.

- Norton, John (2003). A material theory of induction. *Philosophy of Science*, **70**(4), 647–670.
- Okasha, Samir (2002). Underdetermination, holism and the theory/data distinction. *Philosophical Quarterly*, **52**(208), 302–319.
- Papineau, David (1996). Theory-dependent terms. *Philosophy of Science*, **63**, 1–20.
- Poincaré, Henry (1952). *Science and Hypothesis*. Dover, New York. La Science et l'hypothèse (Paris: Flammarion, 1902).
- Post, Heinz (1971). Correspondence, invariance and heuristics. *Studies in History and Philosophy of Science*, **2**, 213–255. Reprinted in S. French and H. Kamminga (1993).
- Psillos, Stathis (1996a). On van Fraassen's critique of abductive reasoning. *Philosophical Quarterly*, **46**, 31–47.
- Psillos, Stathis (1996b). Scientific realism and the 'pessimistic induction'. *Philosophy of Science*, **63**(Proceedings), S306–S314.
- Psillos, Stathis (1999). *Scientific Realism: How science tracks the truth*. Routledge, London.
- Psillos, Stathis (2001). Is structural realism possible? *Philosophy of Science*, **68**(Proceedings), S13–S24.
- Psillos, Stathis (2002). Simply the best: A case for abduction. In *Computational Logic: From Logic Programming Into the Future* (ed. A. Kakas and F. Sadri), Berlin, pp. 605–625. LNAI 2408: Springer-Verlag.
- Psillos, Stathis (2005a). Putting a bridle on irrationality: An appraisal of van Fraassen's new epistemology. In *Images of Empiricism* (ed. B. Monton). Oxford University Press, Oxford.
- Psillos, Stathis (2005b). Ramsey's Ramsey-sentences. In *Cambridge and Vienna* (ed. M. Galavotti), Volume 12 of *Vienna Circle Institute Yearbook*. Springer.

-
- Psillos, Stathis (2006a). Cartwright's realist toil: From entities to capacities. In *Nancy Cartwright's Philosophy of Science* (ed. L. Bovens and S. Hartmann). Routledge, London.
- Psillos, Stathis (2006b). The structure, the whole structure and nothing but the structure? *Philosophy of Science*, **XX**(Proceedings).
- Putnam, Hilary (1978). *Meaning and the Moral Sciences*. Routledge, London.
- Ramsey, Frank Plumpton (1929). Theories. In *Foundations of Mathematics* (ed. R. Braithwaite). Routledge and Kegan Paul, London.
- Redhead, Michael (2001). The quest of a realist. *Metascience*, **10**, 341–347.
- Redhead, Michael and Teller, Paul (1992). Particle labels and the theory of indistinguishable particles in quantum mechanics. *British Journal for the Philosophy of Science*, **43**, 201–218.
- Reiner, Richard and Pierson, Robert (1995). Hacking's experimental realism: An untenable middle ground. *Philosophy of Science*, **62**, 60–69.
- Resnik, David B. (1994). Hacking's experimental realism. *Canadian Journal of Philosophy*, **24**, 395–412.
- Ruben, David-Hillel (ed.) (1993). *Explanation*. Oxford University Press, Oxford.
- Russell, Bertrand (1927). *The Analysis of Matter*. George Allen et Unwin, London.
- Salmon, Wesley C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton.
- Saunders, Simon (2003a). Critical notice: Tian Yo Cao's "The conceptual development of 20th century field theories". *Synthese*, **136**(1), 79–105.
- Saunders, Simon (2003b). Physics and Leibniz's principles. In *Symmetries in Physics: Philosophical Reflections* (ed. K. Brading and E. Castellani), pp. 289–308. Cambridge University Press, Cambridge.

- Saunders, Simon (2003c). Structural realism, again. *Synthese*, **136**(1), 127–133.
- Shapiro, Stewart (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, Oxford.
- Stein, H. (1982). ‘Subtler forms of matter’ in the period following Maxwell. In *Conceptions of Ether: Studies in the History of Ether Theories, 1740-1900* (ed. G. N. Cantor and M. J. S. Hodge), pp. 309–340. Cambridge University Press.
- Suarez, M. (2006). Experimental realism defended: How inference to the most likely cause might be sound. In *Nancy Cartwright’s Philosophy of Science* (ed. L. Bovens and S. Hartmann). Routledge, London.
- van Fraassen, Bas C. (1980). *Scientific Image*. Oxford University Press, Oxford.
- van Fraassen, Bas C. (1985). Empiricism in philosophy of science. In *Images of Science* (ed. P. Churchland and C. Hooker). University of Chicago Press, Chicago.
- van Fraassen, Bas C. (1989). *Laws and Symmetry*. Oxford University Press, Oxford.
- van Fraassen, Bas C. (1991). *Quantum Mechanics: an Empiricist View*. Oxford University Press, Oxford.
- van Fraassen, Bas C. (2001). Constructive empiricism now. *Philosophical Studies*, **106**, 151–170.
- van Fraassen, Bas C. (2003). On McMullin’s appreciation of realism concerning the sciences. *Philosophy of Science*, **70**(3), 479–492.
- van Fraassen, Bas C. (2005). Wouldn’t it be lovely? Review of Peter Lipton’s *Inference to the Best Explanation*. *Metascience*. Forthcoming.
- Votsis, Ioannis (2003). Is structure not enough? *Philosophy of Science*, **70**(5), 879–890.

- Winnie, John (1967). The implicit definition of theoretical terms. *British Journal for the Philosophy of Science*, **18**, 223–229.
- Woodward, James (2004). *Making Things Happen*. Oxford University Press, Oxford.
- Worrall, John (1989). Structural realism: The best of both worlds? *Dialectica*, **43**, 99–124.
- Worrall, John (1994). How to remain (reasonably) optimistic: Scientific realism and the "luminiferous ether". In *PSA 1994* (ed. D. Hull and M. Forbes), Volume 1, pp. 334–342. Philosophy of Science Association, East Lansing.
- Worrall, John and Zahar, Elie (2001). Ramseyfication and structural realism. In *Poincare's Philosophy: From Conventionalism to Phenomenology* (ed. E. Zahar), Chapter Appendix, pp. 236–251. Open Court Publishing Co., Chicago.
- Wright, Cory and Bechtel, William (Forthcoming). Mechanism. In *Philosophy of Psychology and Cognitive Science* (ed. P. Thagard), Handbook of the Philosophy of Science, Vol.4. Elsevier, New York.
- Wylie, Alison (1986). Arguments for scientific realism: The ascending spiral. *American Philosophical Quarterly*, **23**, 287–298.
- Yablo, Steven (1992). Mental causation. *Philosophical Review*, **101**, 245–380.