THE EPISTEMOLOGICAL STATUS OF SCIENTIFIC THEORIES: AN INVESTIGATION OF THE STRUCTURAL REALIST ACCOUNT

Ioannis Votsis

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

PhD Thesis

To my mother and father

ABSTRACT

In this dissertation, I examine a view called 'Epistemic Structural Realism', which holds that we can, at best, have knowledge of the structure of the physical world. Put crudely, we can know physical objects only to the extent that they are nodes in a structure. In the spirit of Occam's razor, I argue that, given certain minimal assumptions, epistemic structural realism provides a viable and reasonable scientific realist position that is less vulnerable to anti-realist arguments than any of its rivals.

The *first* chapter presents an overview of the scientific realism debate, concentrating on the epistemological dimension. The *second* chapter tracks the development of structural realism, differentiates between several versions, and outlines the objections that have been raised against it. The *third* chapter provides answers to a large subset of these objections, namely those launched by Stathis Psillos, who spearheads the critique of epistemic structural realism. The *fourth* chapter offers an attempted solution to M.H.A. Newman's objection that the epistemic structural realist view, if true, trivialises scientific knowledge. The *fifth* chapter presents a historical case study of the caloric theory of heat. I utilise the study to answer the pessimistic meta-induction argument. The *sixth* chapter addresses the argument from the underdetermination of theory by evidence. I argue that epistemic structural realism can potentially restrict the impact of the argument by imposing structural constraints on the set of all possible theories compatible with the evidence. The *seventh* and final chapter outlines briefly some promising avenues for future research.

CONTENTS

1

THE SCIENTIFIC REALISM DEBATE

1. Introduction

A question in the philosophy of science that has engrossed the minds of many eminent thinkers is the epistemological one of what kind of knowledge, if any, science reveals of the physical world. Answers to this question are typically classified as either realist or anti-realist.[1] Structural Realism, as part of its name suggests, is a position on the realist side of the divide. In very simple terms, its advocates hold that our epistemic access to the world, so far as its non-observable part is concerned, is restricted to its structural features. The position can be traced back at least to the beginning of the twentieth century and has recently been attracting renewed interest.[2]

My main aim in this dissertation is to evaluate the structural realist answer to the aforementioned question. It seems only prudent then to devote the first chapter to an examination of the scientific realism debate. In what follows, I will delineate the boundaries of the debate, articulate the various positions and identify the protagonists. I will also sketch the main arguments and the corresponding objections and counter-objections. Finally, I will set out the main obstacles for realism. In doing so, I hope to set the stage for structural realism, explain its role in this debate as well as reveal more about the conditions of its inception and its reincarnation about a decade and a half ago.

2. The Origins and Boundaries of the Debate[3]

Arguably, the scientific realism debate did not really come into its own, i.e. was not independent from general debates about realism, until the twentieth century. The first quarter of the century was marked by a somewhat unsophisticated general realism,

[1] Unless otherwise noted, the terms 'realism' and 'anti-realism' will denote the more specific viewpoints of scientific realism and scientific anti-realism respectively.
[2] This widely held impression is confirmed by the recent increase in the number of publications dealing with structural realism. Note also that in the latest conference of the American *Philosophy of Science Association* (PSA 2002), structural realism was central to three out of five papers in the realism section.
[3] Detailed overviews of the debate can be found in Boyd (1984: 41-82; 2002) and Psillos (2000b).

most memorably the critical realism of Roy Wood Sellars, formed in reaction to the rampant idealism of the nineteenth century. The logical positivists came to dominate the second quarter of the century. In view of the quantum and relativistic revolutions in physics, they found much support for their instrumentalist version of anti-realism. It was not until the 1960s, after a multifaceted attack on logical positivism, that realism was revived under the guidance of such figures as Karl Popper, Grover Maxwell, and J.J.C. Smart. At around the same time, the historically motivated work of Thomas Kuhn and Paul Feyerabend inspired new converts to, and new versions of, anti-realism. Realist voices were not kept at bay, however, with Hilary Putnam and Richard Boyd, among others, keeping the debate alive in the seventies. In the early eighties, the independent but equally powerful critiques by Bas van Fraassen and Larry Laudan shaped old problems into new challenges for the scientific realist. The debate as it is carried out today owes much to these developments, especially those that emerged after 1960.

The twentieth century gave birth and rebirth to a plethora of realisms and anti-realisms. The current debate is so wonderfully varied that I would be unable to justly review in one chapter, or indeed pursue in the rest of the dissertation. For this reason I will concentrate on one particular corner of the debate, something that will make my task more manageable. Three threads common to the central realist and anti-realist positions in this corner of the current debate are the following:

(CD1) There exists a mind-independent world.
(CD2) Scientific claims/sentences/statements have truth-values.
(CD3) Their truth or falsity is determinable by recourse to the mind-independent world.

These threads help circumscribe the debate. The first thread, CD1, endorses ontological realism thereby excluding positions such as traditional forms of idealism, phenomenalism, and solipsism that deny this view. Idealism holds that the world consists only of minds and/or mental states. Phenomenalism, at least in one form, can be understood in a similar way: namely as the position that the world consists only of experiences/perceptions/phenomena. Solipsism offers a more extreme description, claiming that the only thing in existence is one's own mind and mental states.

The second thread, CD2, endorses 'semantic realism'. This excludes positions such as traditional instrumentalism, the verificationist-based instrumentalism of logical positivists and fictionalism. In more detail: Traditional varieties of instrumentalism view scientific theories as means for the organisation and prediction of the observable aspects of the world and deny that they can have truth-values. Similarly, the verificationist-based instrumentalism of the logical positivists holds that only observational, as opposed to theoretical, statements are meaningful and have truth-values. The later logical positivists, who rejected the verificationist principle, argued that theoretical statements are partially interpreted and can have truth-values, all in virtue of their correspondence with observational statements.[4] Fictionalism can be thought of as a version of instrumentalism, since it holds that theories do not have a truth-value but are instead valued for their reliability or usefulness. It supposedly departs from instrumentalism in that it takes scientific theories and their ontological posits to be reliable *fictions*. How the conception of a theory or posit as a fiction differs from that as a mere tool is not all that clear.

The third and final thread, CD 3, endorses the correspondence theory of truth. It understands the notion of truth as one of correspondence between the mind-independent world and language. This excludes positions such as social constructivism and conventionalism. Social constructivists typically argue that scientific knowledge is the product of theorising, not of discovering facts about the world. Conventionalists consider the claims of science as mere agreements, whose truth is guaranteed by stipulation. While some conventionalists restrict the application of their view to domains like logic, arithmetic, and geometry, others apply it across the board covering, among other things, scientific claims.

I do not presume that the excluded positions are without merit, but rather choose to concentrate on a very specific, and more manageable, problem: Assuming CD1,

[4] Psillos (2000b) calls the instrumentalist positions that deny truth-values altogether 'eliminativist instrumentalism', and the positions that allow for truth-values but claim that the truth and meaning of theoretical statements is parasitic on those of observational statements 'reductive empiricism'.

CD2, and CD3, can science lead us to knowledge about the mind-independent world? Participants in the scientific realism debate have, by and large, sought to answer this type of question, shying away from, or at least sidelining, ontological, semantical, methodological, and ethical questions.[5] This dissertation will be almost exclusively concerned with epistemological questions.

The following two theses will help us in the formulation of realism and anti-realism:

(OT) The observable thesis: We can have knowledge of the observable aspects of the world.

(UT) The unobservable thesis: We can have knowledge of the unobservable aspects of the world.

I have left the meaning of the terms ‘observable’ and ‘unobservable’ undefined for now, since there is disagreement over this issue. In what follows, we take a closer look at each of the two opposing camps.

3. Scientific Realism

First Approximation

As a first approximation, we can represent scientific realism as the conjunction of OT and UT. More precisely, scientific realism states that we *can* have, and *actually do* have some, knowledge of the observable and unobservable aspects of the world. But what exactly do we mean by observable and unobservable? The current consensus amongst realists follows Maxwell’s landmark essay (1962), where he argues that there is a continuum from the observable to the unobservable so that no sharp distinction between them can be drawn. Maxwell also argues that what is unobservable is contingent upon factors such as the physiology of the human eye, and that for this reason we cannot demarcate the observable from the unobservable. Some of Maxwell’s arguments rely on the theory-ladenness of observation, an idea that has been advocated by Pierre Duhem ([1914] 1991), Paul Feyerabend (1962), T.S. Kuhn ([1962]1996) and N.R. Hanson (1958) among others. Though the exact meaning of

[5] For a more detailed treatment of these other dimensions of the debate see Niiniluoto (1999: ch.1).

this notion is contested, most agree that since observation statements are formulated in theory-specific contexts, they are to a certain degree imbued with theoretical prejudices. We shall shortly see that the theory-ladenness of observation is a double-edged sword, employed by both realists and anti-realists in their attempts to defeat one another.

Second Approximation

Another requirement of scientific realism, already pointed out under CD2, is that scientific claims have truth-values. Our rough understanding of the concept of knowledge holds that to know something is to have a justified *true* belief about it. Gettier (1963) famously presented an allegedly devastating counterexample to this analysis of the concept of knowledge. In the current context, one need not get into the details of how best, if at all, to characterise the concept. All that need concern us here is the fact that having a true belief about something is a necessary condition for knowing it. To have knowledge of some aspect of the world involves the true belief that the world is in a certain state. Thus, we can express the scientific realist view that we have knowledge of (the observable and unobservable aspects of) the world by saying that scientific claims about the world are true. As a second approximation then we can represent scientific realism as the position which holds that the scientific claims about the observable and unobservable aspects of the world are true.[6]

Third Approximation

Most, if not all, scientific realists accept that the claims made by our current theories are not typically true but rather *approximately true*. In part, the realisation stems from the simple recognition that even our best theories are invariably, though to different degrees, off the mark when it comes to the production of predictions. The recent interest in this field was initiated by Popper (1963), who used the terms 'truthlikeness' and 'verisimilitude' to express the idea that one theory could stand closer to the truth than another.[7] In Popper's account theories are taken to be sets of sentences closed under deduction. According to him, the truth content of a theory *A*

---

[6] For some realists this holds only of scientific claims from the most successful sciences, i.e. physics and chemistry. Others are more liberal.

[7] Niiniluoto (1999: 65) traces the etymological origin of these terms to the Latin term 'verisimilitudo', which means likeness or similarity to truth and was introduced by the ancient sceptics Carneades and Cicero.

is the intersection between *A* and *T*, i.e. $A\cap T$, where *T* is the set of all true sentences. On the basis of this notion, he defines increased truthlikeness thus: a theory *B* is more truthlike than a theory *A* if and only if one of the following two conditions is met:[8]

(C1) $A\cap T\subseteq B\cap T$ and $B\cap F\subset A\cap F$

(C2) $A\cap T\subset B\cap T$ and $B\cap F\subseteq A\cap F$

Popper's definition of truthlikeness was short-lived, for David Miller (1974) and Pavel Tichý (1974) independently proved that under this definition a false theory could not be more truthlike than any theory whatsoever. This is an unwanted result because one of the demands for a theory of truthlikeness is to be able to compare theories that are strictly speaking false yet approximate the truth to greater or lesser extents. Since the refutation of Popper's definition, a number of different accounts of the notions of truthlikeness, verisimilitude and approximate truth have appeared.[9] The most prevalent of these takes *similarity* or *likeness* as measuring distances from the truth (see, for example, Hiplinen (1976), Niiniluoto (1987), and Oddie (1986)). One of the most serious problems with this approach is that comparative judgments of truthlikeness are not translation-invariant. While in one language a theory A may be more truthlike than a theory B, this relation can be reversed in another language. Various solutions to this problem have been proposed (see, for example, Tichý (1978), Oddie (1986)) but none seems to command a consensus.

Many realists have abandoned the task of trying to give formal treatments to these notions and have instead focused on more informal accounts (see, for example, Aronson, Harré and Way (1994), Newton-Smith (1981), Smith (1998) and Psillos (1999)). Whether any such informal account delivers the goods is a contentious issue. At any rate, it is sufficient for the current purposes to note, as a third approximation, that scientific realism can be represented as the position that the

[8] Obviously, *F* is the set of all false sentences.
[9] Note that some authors (see, for instance, Niiniluoto (1999)) assign different meanings and functions to the concepts of approximate truth and truthlikeness.

scientific claims about the observable and unobservable aspects of the world are at least approximately true.[10]

General Formulation

Before we present a general formulation, we must consider one more element, namely the *aim* that scientific realism ascribes to science. According to the first part of van Fraassen's definition of scientific realism "*Science aims to give us, in its theories, a literally true story of what the world is like*" (1980: 8) [original emphasis]. Most realists are happy with this characterisation. Given the traits we have attributed to scientific realism so far, it seems hardly necessary to state that at least one of the main aims of science is to give us true/ approximately true claims about the world. It is nonetheless worth making this feature explicit in our general formulation of scientific realism:

(SCR) Scientific Realism: Science aims to produce, and has succeeded in producing, true/approximately true claims about both the observable and the unobservable aspects of the world.

This formulation captures the spirit of scientific realism. To present a more complete picture, however, we need to look at the main claims that often accompany scientific realism. In 'A Confutation of Convergent Realism', Laudan provides a list of the central claims advocated by scientific realists, correctly acknowledging that "there is probably no realist who subscribes to all of them [though] most of them have been defended by some self-avowed realist or other" (1981: 20). Here is a no-frills version of that list:

(RC1) Scientific theories in mature sciences are typically approximately true.
(RC2) More recent theories are closer to the truth than earlier ones.
(RC3) All the terms, i.e. observational and theoretical, of theories in mature science genuinely refer.
(RC4) Successive theories in mature science 'preserve' the theoretical relations

[10] It might be objected that this statement needs to be restricted to mature scientific claims. Indeed, most, if not all, scientific realists adopt this restriction. This point is correct and is taken on board in the next few paragraphs. For more on the concept of mature scientific claims, I ask the reader to look at the last few paragraphs of section six of this chapter.

and referents of earlier theories.

(RC5) New theories (do and should) explain the success of their predecessors.

(RC6) Claims (RC1)-(RC5) constitute the best, if not the only, explanation for the success of science, and this success provides empirical confirmation for realism. (1981: 20-21).

Laudan calls the conjunction of all these claims ‘convergent epistemological realism’, the idea being that successive scientific theories steadily converge to an ultimate and final theory that faithfully reflects reality.

Having presented a general formulation of scientific realism plus a list of central accompanying claims, it would now be useful to say a few things about the main varieties of realism. Given the numerous, and usually subtle, disagreements over the claims on the above list, it would prove cumbersome to use the list as a point of departure.[11] However, we can make a rough and ready distinction between *total realism* and *partial realism*.[12] Contra total realism, partial realism imposes a distinction between those kinds of theoretical components that can represent some aspect of the world and those that cannot.[13] By ‘kinds’ I here mean the general classificatory schemes employed to systematise science, i.e. entities, laws, etc. Under the banner of total realism we can place philosophers such as Richard Boyd, Philip Kitcher, Jarrett Leplin, W.H. Newton-Smith, Ilkka Niiniluoto, and Stathis Psillos.[14] Under the banner of partial realism we can cite Nancy Cartwright, Ronald Giere, Ian Hacking, Rom Harré, Ernan McMullin, John Worrall, and Elie Zahar.[15]

[11] Leplin (1984: 1-7) attempts to go down this path with a similar list but the result, though somewhat informative, is rather convoluted.

[12] Similar distinctions have been put forward by others. Ilkka Niiniluoto, for example, distinguishes between critical realism and critical half-realism (1999: 12). Arthur Fine (1998) identifies piecemeal realism in a manner similar to Niiniluoto’s critical half-realism.

[13] Some total realists, like Philip Kitcher and Stathis Psillos, draw their own distinctions between those theoretical components that we should believe in and those that we should not. Their distinction does not make them partial realists in the sense explained above, for it does not discriminate between *kinds* of theoretical components. For example, they do not advocate belief only in laws but not entities, or vice-versa, like partial realists do.

[14] For further reference see Boyd (1990), Kitcher (1993), Leplin (1984), Newton-Smith (1989), Niiniluoto (1999), and Psillos (1999).

[15] See, for example, Cartwright (1983), Giere (1988), Hacking (1982), Harré (1988), McMullin (1984), Worrall (1989), and Zahar (2001).

We can cut deeper into partial realism by asking the question, '*What* is it that the partial realist claims we have knowledge of?' Ian Hacking and Nancy Cartwright, for example, are realists about entities, claiming agnosticism about theories. According to Hacking's influential account, hypothetical entities become real "[w]hen we use them to investigate something else" (1982: 1165). His prime example concerns PEGGY II, a polarising electron gun, built according to our knowledge of the causal properties of electrons. When the gun was successfully used to discover the first known example of parity-violation in a weak neutral current interaction, Hacking maintains, we gained further evidence to believe in the reality of electrons.

Similarly, Cartwright (1980) has argued for a realist attitude towards the causes of phenomena, which, at least in this case, involves realism about the entities that feature in causal accounts. It is the fundamental laws of physics, according to her, that we should be wary about since "to the extent that they are true, [they] don't explain much" (867). In Cartwright's view, the fundamental laws of physics work well, and are considered approximately true, in controlled laboratory experiments. But they do not, and according to her cannot, be taken to be true of or explain what goes on in the world outside the laboratory. Outside, the laws need to be augmented by additional assumptions and auxiliaries to be able to model anything; and even then they under-perform in their predictive and explanatory power when compared to what they can achieve in a laboratory. Worse still, they are often completely inapplicable. Having painted this bleak picture, Cartwight argues against fundamental physical laws and in favour of the reality of entities that feature in more localised causal interactions. Her best-known counter-example to the explanatory power of fundamental laws of physics is the intractable dynamics of a thousand dollar-bill floating around in St. Stephen's Square in Vienna.[16]

The other major type of partial realism is structural realism. John Worrall and Elie Zahar, for instance, are realists about structures, i.e. typically laws of nature represented by mathematical equations, claiming that theoretical posits and non-structural parts of theories alike are suspicious. Since the next chapter is devoted to an explanation and elaboration of structural realism, I will restrict my comments here

[16] See her (1998: 28). Cartwright correctly credits Otto Neurath with the example.

to the prima facie incompatibility between entity realism and structural realism. If entity realists remain agnostic with regard to theories/fundamental laws, which presumably includes structures, and structural realists remain agnostic with regard to theoretical posits, then obviously the two positions can hardly disagree more. Niiniluoto (1999: 139) goes as far as to call them 'diametrically opposite'. I think that his ruling may be a bit premature. Despite their professed aversion towards theory, entity realists make allowances for some, low-level, theory. Hacking, for example, appeals to 'low-level causal properties', which, no matter how much glazing he puts on them, are simply theoretical properties. Similarly, as we shall see in the next chapter, structural realism does not reject knowledge of entities but rather restricts such knowledge to their structural features.[17]

4. Arguments in Support of Realism

Over the years many arguments have been proposed in favour of realism. Of these, few have carried as much weight as the *No Miracle* and *Inference to the Best Explanation* arguments. The following is a brief exposition of the principal claims involved in these arguments as well as objections raised against them.

The argument that came to be known as the 'no miracle argument' (NMA) was independently proposed by J.J.C. Smart (1963) and Hilary Putnam (1975). According to the NMA, scientific realism is the *only* view that does not make the success of science a miracle. Given the empirical success of scientific theories, it would be a coincidence of almost cosmic proportions or a miracle if they were not at least approximately true. The tacit assumption underlying the NMA is that most of us are unwilling to accept that the success of science is a miracle. We thus opt for the purportedly only alternative, scientific realism.

It could be objected that the NMA poses an unfair dilemma: either uphold scientific realism or consider the success of science a miracle. The second disjunct is generally accepted as not really an option. Indeed, van Fraassen (1980: 39-40) concedes that we might need to account for the success of science, but denies that the

[17] Chakravartty (1998) goes as far as to say that the two views, when properly construed, entail each other. Though I do not find this claim convincing, I agree with the general idea that entity and structural realism can be harmonised with one another without much difficulty. For more on this see my review (2003a) of Niiniluoto's book.

only or best account is scientific realism. To support his point he makes an analogy between the practice of science and the theory of evolution. Scientific theories also struggle for survival with only the 'fittest', i.e. most successful, surviving. These, van Fraassen says, need not be true or approximately true but they need to be *empirically* successful.[18]

In order to strengthen arguments like the NMA, realists oftentimes emphasise the importance of novel predictions.[19] It is argued, for example, that scientific realism best accounts for the *novel* success of science. A prediction is novel, according to the most basic notion of novelty, if the phenomenon predicted was not known to have existed prior to the theory's prediction of it. This is often called *temporal novelty*. More sophisticated notions have been proposed over the years. Elie Zahar (1973), for example, has proposed the notion of *heuristic novelty*, also called *design novelty*, to convey the idea even if a phenomenon *P* is known prior to the inception of a theory *X,* its prediction by *X* will be novel provided that *P* was not used in the construction of *X*. More generally, so long as a body of evidence was not used in a theory's construction it counts as heuristic-novel.[20] Newton's gravitational theory is a case in point. Although the precession of the equinoxes was known to Newton, his theory was not constructed using this phenomenon. According to the notion of heuristic novelty, the subsequent prediction of the phenomenon using Newton's theory counts as novel.

Many philosophers believe that the concept of *inference to the best explanation* (IBE) is due to C.S. Peirce, who introduced it under the name of 'abduction'. What certainly is uncontestable is that a century later Gilbert Harman (1965) branded this type of reasoning 'inference to the best explanation'.[21] The idea behind IBE is simple and intuitive, its use abundant in scientific practice. If a theory X *explains* some

---

[18] To understand this argument properly one needs to know more about van Fraassen's take on the epistemological status of scientific theories. This task will be taken up in the coming sections. In the meantime it is important to note that van Fraassen's evolution analogy is criticised in, among other places, Brown (1994: 6-7).
[19] There is a thriving literature on this topic. Some notable articles include Worrall (1985), Mayo (1991), and Achinstein (1994).
[20] John Worrall offers a notion of *use-novelty* that is a development of Zahar's notion (see the former's (1985) and especially his (2002)).
[21] In Peirce's work, abduction is more general than inference to the best explanation; it is inference to *some* explanation.

evidence better than any of its rivals, then it is reasonable to choose X over the others. IBE is thus essentially comparative in nature, with explanatory merits as the adjudicating force.[22] This much seems trivial. More contentiously, many realist supporters of IBE have argued that we should not merely choose X over its rivals but that we should believe in the *truth* or *approximate truth* of X.

It is not hard to see how this largely methodological concern has been hijacked for the epistemological concerns of the scientific realism debate. Boyd and Putnam, in particular, are credited with developing an IBE-based explanationist defence of realism that has come to dominate the realists' arsenal.[23] Their argument is that the empirical success of science, not just a body of evidence, requires explaining. The best, indeed the only, explanation for this success, according to them, is realism.[24] They thus see the NMA as an instance of IBE. That is, it is inferred that the success of science is not due to a miracle but rather to the truth/approximate truth of the theories employed. In fact, Boyd, Putnam, and more recently Psillos, treat scientific realism as a scientific hypothesis, whose support comes from the view that it is the only viable explanation of the methodological success of science.[25]

The most thorough study of IBE thus far has been that of Peter Lipton (1991). He compares IBE to various traditional inferential devices, such as the 'instantial model of inductive confirmation' and the hypothetico-deductive model, arguing that IBE overcomes some of their shortfalls. Not only is IBE better than the competition, according to Lipton, but it also "gives a natural description of familiar aspects of our inferential procedures" and "has a number of distinctively philosophical applications" (66,70). IBE is not a monolithic concept. Lipton identifies a range of IBEs, of which he singles out inference to the loveliest potential explanation (see ch.4). This is contrasted to inference to the likeliest potential explanation, the

[22] Having said that, I don't think that proponents of this view would be alarmed if someone pointed out the fact that a lot of theories have no extant rivals. In reply, they would probably say that if it is a theory of a mature science and it explains the data, it should still be considered as true or at least approximately true. For them, being the sole contestant just means that it is the only one that explains the data, and in that respect the best available theory.

[23] Not all scientific realists in fact accept the explanationist defence (see Newton-Smith (1989)).

[24] Psillos (1999: 71) argues that realism is the *best* rather than the *only* explanation of science.

[25] Boyd calls this the 'abductive strategy' which he contrasts with a similar approach that he calls 'local explanationism'. For more see his (2002: 7-9). Notice that the so-called abductive strategy is similar to Laudan's claim RC6.

loveliest explanation offering the most understanding while the likeliest being best supported by the evidence. Lipton prefers inference to the loveliest potential explanation because he thinks that explanatory loveliness can be a guide to likeliness and that our inference making becomes less interesting the more we restrict the role of explanatory virtues.

Both IBE and the explanationist defence of realism have been criticised on numerous grounds. The most common objection is that using IBE to choose one theory over existing rivals guarantees neither the theory's truth nor its approximate truth. After all, the pool may contain only false theories. This objection has, in turn, been used to argue that the explanationist defence of realism is question-begging (see, for example, van Fraassen (1985)). Given that non-realists do not accept IBE in science, the argument goes, there is no basis to accept the (meta-level) explanationist defence of realism. Van Fraassen, in particular, offers an alternative account according to which "we are always willing to believe that the theory which best explains the evidence, is *empirically adequate*" (1980: 20). He thus uses IBE, originally brought into the debate to support the realist, to make an anti-realist inference, namely that an explanatory better theory is empirically adequate.

5. Scientific Anti-Realism

All anti-realists, not surprisingly, share a distrust of, or scepticism towards, realist claims. Just like realism, anti-realism can be found in various forms and guises. With regard to scientific knowledge, the general anti-realist intuition is that we cannot know whether any of the claims made by scientific theories about the mind-independent world are true or approximately true. As a consequence, anti-realists consider the realist claims RC1-RC6 unwarranted. In particular, they denounce the realists' principal claims that theories are increasingly approximating the truth and that the theoretical terms in currently successful theories refer, i.e. the entities alleged to exist by these theories really do exist.

As previously indicated, given the assumptions that I set out in the beginning of this chapter, i.e. CD1, CD2 and CD3, only one anti-realist position qualifies as an alternative to realism. I am referring to constructive empiricism, the position that is widely thought of as the main anti-realist competitor in this debate. In what follows I take a look at the main tenets of constructive empiricism.

Constructive Empiricism

The view identified as 'constructive empiricism' is the brainchild of Bas van Fraassen. It shares some features of the older instrumentalism, but it diverges from it in at least one important respect. As van Fraassen is at pains to point out, constructive empiricism insists on a literal construal of the language of science. In short, theoretical statements are understood as having truth-values. The catch, however, is that we cannot find out what truth-values theoretical statements have. We can only assign truth-values to observational statements. That, according to him, is enough to present science as a rational process.

In line with logical positivists, but against realists, van Fraassen supports a distinction between observables and unobservables. To be precise, he lambasts the use of expressions such as 'observational vs theoretical dichotomy' and 'theoretical entity', saying that these are examples of category errors. Entities are observable or unobservable, while terms and concepts are theoretical. This clarification, argues van Fraassen, leads to two important questions: 1) Is language divisible into theoretical and non-theoretical parts? 2) Are objects and events divisible into observable and unobservable ones? He answers the first negatively by appeal to the idea that our language is thoroughly theory-laden. He answers the second affirmatively in saying that though the term 'observable' is a vague predicate, just like most predicates in natural language, "it is usable provided it has clear cases and clear counter-cases" (1980: 16). He goes on to say that seeing with the naked, i.e. unaided, eye is a clear case of observation whereas 'seeing' particles in a cloud chamber is a clear counter-case.

Constructive empiricism is offered as an epistemologically frugal view that can nonetheless make sense of science. More precisely, constructive empiricism is the view that "*science aims to give us theories which are empirically adequate; and acceptance of a theory involves a belief only that it is empirically adequate*" (12) [original emphasis]. One evident difference between van Fraassen's position and realism is the replacement of the criterion of *truth* with that of *empirical adequacy*.

What exactly is empirical adequacy and why should we prefer it to truth? The answer to the first part of the question is that a theory is empirically adequate when everything it asserts about the *observable* world is true. Echoing Duhem's phrase 'saving the phenomena', van Fraassen argues that a theory is empirically adequate if it saves the phenomena. The answer to the second part of the question is that the criterion of empirical adequacy is less demanding (and presumably more warranted) than the criterion of truth, for it requires theories to make true assertions *only* about the observable aspects of the world. In other words, Van Fraassen rejects UT.

Discussions of the merits and drawbacks of constructive empiricism can be found in abundance.[26] Many of the objections raised against it are directed at the notion of empirical adequacy. John Worrall (1984) and Alan Musgrave (1985), for example, have independently argued that if a theory is to be empirically adequate in van Fraassen's sense, then it must save all the phenomena, not just those actually observed so far. But since we can never have access to all the phenomena, we will never be warranted in accepting a theory as empirically adequate. Many other objections are directed at the observable-unobservable distinction. It has been argued, for example, that the selective scepticism that van Fraassen advocates cannot really be upheld since it is presumably based on an arbitrarily drawn distinction (see Paul Churchland (1982) and Gary Gutting (1983)).

6. Arguments in Support of Anti-Realism

Two arguments that have a venerable history supporting anti-realism are: 1) the underdetermination of theory by evidence and 2) the damning historical record of science.

Underdetermination of Theory by Evidence

Though currently found in various formulations, the main idea behind the underdetermination of theories by evidence (UTE) is, roughly speaking, that for any given body of evidence there are infinitely many competing theories that can 'accommodate' it, so that the evidence cannot uniquely determine a scientific theory.

[26] Churchland and Hooker (1985) contains a collection of essays on constructive empiricism including a reply from van Fraassen.

That the inference from the evidence to the theory is not deductively valid is an age-old idea. One prominent advocate is David Hume. Hume famously argued that no matter how many occurrences of an event we observe, we cannot derive a universal statement from them. This has come to be known as the problem of induction.[27] A similar idea that has been around since (at least) the late nineteenth century concerns the fitting of curves. It is a matter of fact that infinitely many curves pass through any finite number of points. The analogy with UTE should be obvious, i.e. infinitely many theories can accommodate the same (inevitably finite) body of evidence.

A related, though distinct, idea was put forth by Pierre Duhem ([1914] 1991). He argues that confirmation is a holistic affair. More precisely, he argues that a hypothesis can never be tested in isolation, since it cannot produce testable predictions without auxiliary assumptions. Put differently, a counterinstance falsifies the whole conjunction (i.e. hypothesis plus auxiliaries), leaving us uncertain about which of the conjuncts are to blame. Duhem's thesis was subsequently revived, though arguably in a different guise, by W.V. Quine (1951). He has proposed the stronger argument that any hypothesis in our web of beliefs can always be saved by adjusting the web to accommodate evidence that was previously thought of as negative.[28]

UTE supports anti-realist accounts in that it holds that no matter how much evidence we amass we will always have infinitely many theories to choose from, i.e. we will never be able to uphold any one theory as the true one. We can formulate a constructive empiricist version of UTE:

(UTE-CE): For any given body of observational evidence there are infinitely many *empirically equivalent* theories that diverge on their theoretical claims.

Though it is not uniquely associated with constructive empiricism, the concept of empirical equivalence features centrally in it. We say that two or more theories are

---

[27] Nelson Goodman (1965) presents concrete examples of how induction can fail to pick the right theory. See Colin Howson (2001) for one of many alleged solutions of the problem.

[28] Donald Gillies (1993) argues that Duhem's thesis differs from Quine's thesis. Carl Hoefer and Alexander Rosenberg (1994) point out the differences between underdetermination and what has come to be known as the 'Duhem-Quine thesis'.

empirically equivalent when they entail the same observational consequences.[29] To remind the reader, constructive empiricism urges belief in a theory's empirical adequacy, i.e. roughly speaking belief that only the observational consequences of the theory are true. UTE-CE supports constructive empiricism for it holds that no observational evidence will ever allow us to find out which theoretical claims are true or approximately true. Consequently, UTE-CE upholds the belief that only the observational consequences of the theory can be shown to be true.

Given the gravity of these allegations, it is not surprising that the many UTE variants have come under heavy fire (see, for example, Clark Glymour (1980)). In a landmark article, Laudan and Leplin (1991) have objected, among other things, that the notion of empirical equivalence is not well defined. Even if we ignore this, they argue, we can still choose between empirically equivalent theories because: (1) a theory is not necessarily supported by the empirical consequences it entails and (2) a theory can be supported by evidence that it does not itself entail. The second point can be interpreted in one of two ways: (2a) a theory can be supported by empirical evidence over and above the evidence it entails and (2b) a theory can be supported by extra-empirical evidence, namely by considerations of economy, simplicity, unity, explanatory worth, etc. Whether such considerations are epistemically relevant is the object of debate. Moreover, what counts as evidence for a theory can have a tremendous impact on the efficaciousness of the above claims and, by extension, on UTE and the debate as a whole. In all, the realists hope to show that there are justifiable methods through which we can choose between empirically equivalent theories.

The Damning Historical Record of Science

At the beginning of the twentieth century, Pierre Duhem and Henri Poincaré made a compelling case that the history of science is punctuated by the overthrow of hitherto successful theories.[30] The logical positivists, who inherited much from both Duhem and Poincaré, largely ignored historical considerations. The result was a pervasive, though tacit, assumption that scientific knowledge was at once both cumulative and progressive.

---

[29] For a somewhat different notion of empirical equivalence see Quine (1975).
[30] See the next chapter for details.

It was not until the 1960s that this assumption was genuinely brought into question. Thomas Kuhn ([1962]1996), Paul Feyerabend (1962; 1965), and many others reinstated the point made earlier by Duhem and Poincaré and reinforced it with historical case studies. Kuhn, in particular, argued that defining moments such as the Copernican, Newtonian and Einsteinian revolutions, bring about a shift in paradigm that replaces old concepts and theories by radically new ones.[31] The meanings of theoretical concepts belonging to competing paradigms are so radically different, Kuhn argues, that it is impossible to compare either the paradigms or the concepts, let alone support the view that there is some continuity between them.[32] This has come to be known as the 'incommensurability thesis'.

Indeed, Kuhn avoids the notions of truth and approximate truth altogether, opting instead for an account of progress that views science as a problem-solving endeavour. Given incommensurability, the argument goes, there is no common ground from which to judge the goals of the competing theories and, therefore, scientific theories cannot be said to be increasingly approaching the truth. The notion of incommensurability is often intertwined with that of the theory-ladenness of observation. Since observation is theory-laden, the anti-realist argues, it cannot serve as independent ground upon which rival theories can be judged. In sum, Kuhn claims that theory change involves radical shifts in which essential theoretical components including central theoretical terms are thrown away and thus that scientific knowledge is neither cumulative nor progressive towards the truth.

Even though arguments based on the historical record of science were originally launched against logical positivist instrumentalism, an anti-realist position, they have since become the staple of anti-realists in their attempts to bring down realism. At stake are the realist claims on the above list. RC4, for example, is in direct conflict with the historical arguments, for the latter undermine the claim that successive theories in mature sciences preserve at least some of the theoretical relations and referents of earlier theories – notably the central ones.

---

[31] Put simply, a paradigm consists of one or more theories, auxiliary hypotheses, heuristic models, ontological assumptions and methodological principles.
[32] The implicit assumption here is a descriptive theory of reference according to which a theoretical term *t* refers to an entity *a* if and only if *a* satisfies the theoretical (i.e. descriptive) claims made by the scientific theory employing *t*.

The realist reaction to these early historical arguments has followed one of two strategies. On one strategy the realists have launched an offensive against the notions of scientific revolution, paradigm, and incommensurability, claiming that they suffer from vagueness (see, for example, Dudley Shapere (1964) and Lakatos (1970)). Lakatos' 'methodology of scientific research programmes', in particular, replaced the concept of paradigm with that of scientific research programme, characterising the latter in ways that would support a more rationalist outlook towards theory change in the history of science.[33] On the other strategy realists have contested the anti-realist points on historical grounds (see, for example, Richard Purtill (1967)).

A more sophisticated version of the historical argument has been put forward by Laudan (1977; 1981). Laudan criticises the use of connections between reference, approximate truth, and success in support of the explanationist defence of realism as tenuous. More precisely, he argues that the predictive and explanatory success of a theory guarantees neither its approximate truth nor that its central theoretical terms genuinely refer. The available historical evidence, according to him, clearly shows a repeated overthrow of scientific theories as false and their referents as not genuinely referential, despite explanatory and predictive success. Laudan cites the following long list of theories as evidence for his claim: the crystalline spheres of ancient and medieval astronomy, the humoral theory of medicine, the effluvial theory of static electricity, the 'catastrophist geology', the phlogiston theory of combustion, the caloric theory of heat, the vital forces theories of physiology, the electromagnetic ether, the optical ether, the theory of circular inertia, and the theories of spontaneous generation (1981: 33). This argument, thus, challenges the realist claims RC3 and RC4.[34]

Implicit in Laudan's argument is the so-called 'pessimistic induction' (PI).[35] Laudan argues that, given the historical evidence, the inference from explanatory and

---

[33] Lakatos presented his work as a synthesis of some of Kuhn's and Popper's ideas.

[34] Laudan does not stop there. Like many others, he accuses the realists of failing to provide a semantical and epistemological characterisation of the notion of approximate truth, holding that this makes RC1 and RC2 'so much mumbo jumbo' (1981: 32). He also questions RC5 saying that a theory might be better supported than its rivals yet not be able to explain why its rivals were successful (47). Given his distrust of all these claims, i.e. RC1-RC5, he thinks that RC6, which relies upon them, cannot be upheld.

[35] Though Laudan's (1981) argument is now widely known as the 'pessimistic induction' argument, it has been pointed out (see Timothy D. Lyons (2002)), that this argument is only present in his (1977).

predictive success to approximate truth and successful reference is unwarranted. Thus construed, the argument is a modus tollens, not an induction (see Lyons (2002)). However, one can read this argument as an induction. That is, given the historical evidence that past successful theories were abandoned as false and referentially unsuccessful, we can inductively argue that current or even future theories will also succumb to the same fate. This reasoning employs historical evidence to argue, inductively, for pessimism with regard to the approximate truth and referential success of our theories.

Though some realists have largely ignored the pessimistic induction, many more have taken it seriously. Some of these have attacked the argument itself (see, for example, Hardin and Rosenberg (1982), Psillos (1996), and Devitt (1984: 143-9)). Others have engaged in historical case studies in an attempt to show that the historical record can be reconciled with scientific realism (see, for example, Worrall (1989; 1994), Kitcher (1993), Psillos (1999: ch.6)). This last move usually involves showing that abandoned theoretical components are not essential for the explanatory and predictive success enjoyed by the theories they were embedded in. In other words, the theoretical components that survive theory change are those that are responsible for the abandoned theories' successes.

In their fight against historical arguments the realists have appealed to the notion of *mature science*. By categorising those theories that have been abandoned in their entirety as belonging to an insufficiently developed or immature science, the realists hope to restore the cumulativity of scientific knowledge. The distinction between mature and immature science is appealing on independent grounds because many would like to draw a line between the early primitive/undeveloped stages of a given science and the latter stages where the science presumably begins to blossom.[36] Many, for instance, would find Aristotelian physics or the Ptolemaic systems of

The argument has also been put forward, independently on the face of it, by Putnam, who says that "…eventually the meta-induction becomes compelling: *just as no term used in the science of more than fifty* (or whatever) *years ago referred, so it will turn out that no term used now* (except maybe observation terms, if there are such) *refers*" (1978: 25) [original emphasis]. It is worth noting that the argument is also called 'pessimistic meta-induction'. Obviously the 'meta' refers to the fact that it is *about* science and its inductive methods, rather than *within* science.

[36] A similar distinction is utilized to demarcate science from religion.

astronomy unworthy of even being called proto-science.[37] Boyd (1984) and Putnam (1978) cite the phlogiston theory of combustion as another example of an immature science – in this case chemistry.

The concept of maturity is notoriously elusive. Laudan complains that the vagueness besetting the concept risks making the realist claims RC4 and RC5 vacuously true because theories that have not bequeathed anything to their successors can always be branded 'immature'. One way to anchor the concept is by attaching a condition of genuine predictive success to it. That is, unless a theory is explanatorily and predictively successful, it will not count as mature. Yet, even this is not enough to save the realists from the clutches of history. This is made obvious by Laudan's list, which specifically targets theories with presumably genuine empirical success that were subsequently abandoned nonetheless.

Worrall has pressed for a more refined notion of mature science arguing that "[t]his must mean more than simply having correct empirical consequences" (1989: 153). His suggestion is that a science reaches maturity only when its theories can predict entirely novel types of phenomena. Chief amongst his examples is Fresnel's theory of light. The theory unexpectedly and correctly predicted a bright spot at the centre of the shadow of an opaque disc that was lit from a single slit. Though this theory appears on Laudan's list, Worrall argues, the essential part of the theory, namely Fresnel's equations, were preserved through theory change.

Whether Worrall's notion of maturity saves the realist from the allegedly embarrassing historical record is an issue that has yet to be taken up. Prima facie, it seems to me that his criterion is too strict in that it could eliminate approximately true theories that do not make any predictions of novel types of phenomena. One need only consider that a successor to a given theory may be closer to the truth simply on account of accuracy, and not by predicting new types of phenomena.

[37] Michael Friedman, however, suggests that even Aristotelian physics has handed down some approximately correct theoretical components (see his (2001)).

7. The Main Realist Obstacles

Given the current state of the debate, we can easily identify the main obstacles realists have to overcome if they are to make any progress. With little or no doubt, the following four are the most talked about and presumably most important obstacles for the realist in the current debate:

(RP1) We must be able to choose between empirically equivalent theories. That is, we must be able to show that from a pool of empirically equivalent theories at least some are more epistemically warranted than others.

(RP2) The historical record of science must be accounted for somehow. It must be shown that at least some components of theories, other than observational consequences, survive theory change, and that only those that survive were responsible for the success of a given theory.

(RP3) It must be shown, or at least it is preferable to show, why the success of science needs explaining and, furthermore, why scientific realism provides a better explanation than any alternative position.[38]

(RP4) The notions of approximate truth, truthlikeness and verisimilitude need to be given rigorous characterisations. If no adequate formal treatments can be given, as indeed conceded by some realists, more robust informal accounts as well as the reasons why such accounts would work need to be clearly explained.

This dissertation will investigate whether structural realism can overcome RP1 and RP2. RP3 and RP4 are thus purposely bracketed. To include RP3 and RP4 into my investigation would mean either to excessively expand the dissertation or to cover one or more of the four obstacles only superficially. I think that RP1 and RP2 are sufficiently independent to be able to be addressed without first addressing RP3 and RP4. Regarding RP3, I will assume that the success of science needs some explanation, or, at least, some accounting for. Though I do not aim to provide a thorough answer to the question why scientific realism and in particular structural

---

[38] Not all realists think that the success of science needs explaining. Worrall is one such realist.

realism offers a better explanation/account of this success than anti-realism, some of my arguments will be supportive of this view. Regarding RP4, I will rely, like so many other philosophers I mentioned earlier, on some informal understanding of the notion of approximate truth. I do not assume that this informal understanding is sufficient for a realist programme. That is an issue that needs to be investigated thoroughly but not in this dissertation.

8. Conclusion

Arthur Fine (1984) has suggested an alternative to both realism and anti-realism, which he has called the ‘natural ontological attitude’ (NOA) and which he classifies as non-realism. According to him, NOA is a deflationary attitude that does not seek to impose a ‘general interpretive scheme’ on science. Unlike realism and anti-realism, for example, NOA does not set any aims, like truth or empirical adequacy, for science. So much for what NOA is not. What about its positive dimensions? Fine claims that NOA is to be equated with what he calls the ‘core position’, i.e. that which is common to both realists and anti-realists. The ‘core position’ is simply the view that the results of scientific enterprise are true. The realist and anti-realist positions are ‘unnatural’, according to Fine, because they add metaphysical theses about the character of truth and reality to the core position.

On the basis of the above, Fine has called for a dismissal of the whole debate pronouncing realism dead. His call has not been heeded however. What is more, NOA has been rightly criticised for its failure to adequately distinguish itself from realism. It has been argued, for example, that NOA is just realism in disguise, for it accepts something that anti-realists like van Fraassen reject, namely the truth of scientific claims about the unobservable world. That is, against Fine’s claims, NOA cannot be equated with the common core.

Despite Fine’s dismissal, many philosophers believe that at least some headway can be made in the debate. What seems evident from the elaboration of the debate offered in this chapter is that the central arguments are now more sophisticated than fifty or a hundred years ago. That, of course, does not necessarily mean that we are progressing towards a resolution of the debate. Rather, it, at least, means that a lot of interesting tools have been discovered or invented in the process. Indeed, some of these contributions have been made by structuralists, and, in particular, structural realists. It is to the historical and conceptual development of structural realism that I turn to next.

2

TRACING THE DEVELOPMENT OF STRUCTURAL REALISM

1. Introduction

This chapter traces the development of structural realism within the scientific realism debate and the wider current of structuralism that has swept the philosophy of the natural sciences in the twentieth century.[39] The primary aim is to make perspicuous the many manifestations of structural realism and their underlying claims. Among other things, I will compare structural realism's various manifestations in order to throw more light onto the relations between them. At the end of the chapter, I will identify the main objections raised against the epistemic form of structural realism. This last task will pave the way for the evaluation of the structural realist answer to the main epistemological question, an evaluation that will be central to the rest of this dissertation.

Generally construed, structuralism is a point of view that emphasises the importance of relations. It takes the structure, i.e. the nexus of relations, of a given domain of interest to be the foremost goal of research and holds that an understanding of the subject matter has to be, and most successfully is, achieved in structural terms. The following quote from Redhead (2001a) nicely conveys this intuition: "Informally a structure is a system of related elements, and structuralism is a point of view which focuses attention on the relations between the elements as distinct from the elements themselves"(74). This vision has shaped research programmes in fields as diverse as mathematics, linguistics, literary criticism, aesthetics, anthropology, psychology, and philosophy of science. It is the last-mentioned that I am concerned with in this chapter.

The first explicit statements of a structuralist programme in the philosophy of science can be traced back to Henri Poincaré, Pierre Duhem, and Bertrand Russell.

[39] To the best of my knowledge, the only other attempt to trace the historical and ideological development of structural realism is to be found in Barry Gower (2000). Gower's article is rather narrowly construed, however, for he focuses mainly on Ernst Cassirer and Moritz Schlick.

Other structuralists or structuralist-oriented philosophers followed, notably Arthur Eddington, Ernst Cassirer, Rudolph Carnap, Moritz Schlick, W.V.Quine and Grover Maxwell. During the last decade and a half, the position has been revived, reformulated, and vigorously defended, by Otávio Bueno, Anjan Chakravartty, Tian Yu Cao, Bas van Fraassen, Steven French, James Ladyman, Michael Redhead, John Worrall, and Elie Zahar, to a name a few. Given the numerous differences between many of these authors' philosophy it is not surprising then that there are almost as many structuralisms as there are structuralists.

A terminological remark is required at this point to ward off misapprehensions. 'Structuralism' will refer to the general intuition that the focus is on the relations and not the relata. To identify each individual position I will employ variant terms like 'epistemic structural realism', 'ontic structural realism', 'structural empiricism', etc. Often, these names are already available but, where needed, I will provide my own names so as to keep track of who is arguing for what.

Before we delve into the different types of structuralism, I must present a definition of the notion of structure that is precise enough to help disambiguate some of the discussions. Although, as we shall see later, matters are more complex, we can begin the elucidation of the various forms of structuralism by presupposing the standard definition of structure:[40] A *structure* S = (U, R) is specified by two things: i) a non-empty set U of objects (the domain of S) and ii) a non-empty set of relations R on U.[41] A structure may also specify one-place properties but these are not essential. In other words, a minimum requirement for setting up a structure is to have a set-theoretically specifiable (i.e. extensionally defined) relation between objects. Notice that many of the mathematical statements central to science, i.e. functions, equations, laws, symmetries, principles, covariance statements, etc., postulate relations between terms that can usually be expressed set-theoretically in the above-mentioned way.

[40] As we shall see by the end of this chapter, John Worrall and Elie Zahar argue against such a view of structure because individuals are taken as more basic than relations, i.e. relations are defined as sets of ordered n-tuples of individuals. They instead call for a new semantics that takes structures and, by extension, relations as more primitive than individuals.
[41] The definition of structure sometimes includes a third condition, i.e. a set O of operations on U (which may be empty). This condition is optional because operations are functions and thus can be regarded as special kinds of relations capturable by condition two.

2. The Prehistory of Structuralism

As mentioned above, the history of structuralism starts with Poincaré, Duhem and Russell. Van Fraassen (1997; 1999), however, has recently added an interesting pre-history to the topic that deserves consideration. Drawing from 19th century discussions of how science represents natural phenomena, van Fraassen (1997) traces the beginnings of structuralism to the emergence of non-Euclidean systems of geometry. The discovery of such systems led to the realisation that no system is privileged, i.e. to a 'relativisation of representation'. The applicability of these systems to physics, van Fraassen claims, resulted in a parallel relativisation. For obvious reasons, this result challenged the naïve realist view that there is a unique way to represent physical space and, more generally, the physical world. In light of these developments, van Fraassen argues, Russell was led from naïve realism to structuralism. Though van Fraassen is not very informative about the reasons behind Russell's change of heart, the implication seems to be that because structuralism necessitates the non-uniqueness of descriptions, through the idea that things can be described only up to isomorphism, it supports a kind of 'relativisation of representation'.

In his more recent paper (1999), van Fraassen stretches our imagination even further by attempting to extend the prehistory of structuralism. He entertains the idea that structuralism could have gained support as far back as the 17th century. It is the increasing mathematisation of science, van Fraassen argues, that paves the way for structuralism. He sees Isaac Newton's introduction of non-mechanical, highly abstract and mathematical descriptions of nature as the end of one era and the beginning of a new one. At the same time, he sees in Newton a disdain for too much mathematisation for fear that it may lead to the Aristotelian occult properties he so desperately tried to avoid. Indeed, he sees the same misgivings in James Clark Maxwell. Both Maxwell and Newton, according to van Fraassen, oscillated and agonised between two extreme positions: 'reification' and 'structuralism'. These positions, he argues, "emerge very naturally when science proves itself too complex for philosophical naiveté. We see a clear tendency to reify whatever theories invoke in their representation of nature. But conceptual difficulties and the increasingly mathematical character of science foster the structuralist impulse" (1999: 7). And, he continues, "[t]his is one of the main reasons why, I think, we see the structuralist reaction emerging in the 19th century. As so often happens, what is earlier seen as a failure or shortcoming becomes the glory of a new generation" (12).

I think van Fraassen's claim, that Newton and Maxwell were wavering between reification and structuralism, is reading too much into history. His examples can only establish that these scientists were sceptical about too much mathematisation. Similarly, his colourful and somewhat cryptic remarks about the emergence of reification and structuralism are in need of further elaboration if they are to be taken seriously. Even so, it is certainly plausible that the mathematisation of nature in general and the rise of non-Euclidean geometries in particular, facilitated structuralist inclinations.

3. The Early Years

Poincaré

Poincaré is often thought of as a conventionalist, not only with regard to geometry but also physics, and as such not a realist. However, Grover Maxwell (1968), Jerzy Giedymin (1982), Worrall (1982; 1989; 1994), Zahar (1989; 1996; 2001), David Stump (1989), Stathis Psillos (1995; 1999), Barry Gower (2000), and Redhead (2001a) are all in agreement that Poincaré was an epistemic structural realist.[42] Epistemic structural realism (ESR) is, simply put, the view that our knowledge of the physical world is restricted to structure. I agree that Poincaré was an ESR-ist and, in what follows, present the reasons why I think this is the case.

Poincaré was heavily influenced by German idealism, a philosophical school that, as is well known, considers Kant as its progenitor. More precisely, Poincaré subscribed to the view that the non-phenomenal entities postulated by scientific theories are the Kantian things-in-themselves. Unlike Kant, however, he thought that it is possible to gain indirect knowledge of the things-in-themselves. What is it exactly that he thought we could know about them? Poincaré is unequivocal: "[T]he aim of science is not things themselves, as the dogmatists in their simplicity imagine, but the

[42] The term 'structural realism' was coined by Grover Maxwell (1968) with reference to Russell's position. Stump does not use the term 'structural realism' but nonetheless understands Poincaré as a structural realist.

relations between things; outside those relations there is no reality knowable" ([1905]1952: xxiv). And again later on in the same book: "The true relations between these real objects are the only reality we can attain" (161). Despite the fact that the term 'structure' does not appear in these or other relevant passages, we are entitled to call Poincaré an epistemic structural realist for, after all, structures in their simplest form are just collections of one or more relations.

As many authors have pointed out, the motivation for Poincaré's structural realism is largely historical.[43] More precisely, he takes the survival of theoretical relations through theory change as indicative of their having latched onto the world. Here's an illuminating passage from *The Value of Science*:

> …science has already lived long enough for us to be able to find out by asking its history whether the edifices it builds stand the test of time, or whether they are only ephemeral constructions.
>
> Now what do we see? At the first blush it seems to us that the theories last only a day and that ruins upon ruins accumulate… But if we look more closely, we see that what thus succumb are the theories properly so called, those that pretend to teach us what things are. But there is in them something which usually survives. If one of them taught us a true relation, this relation is definitively acquired, and it will be found again under a new disguise in the other theories which will successively come to reign in place of the old ([1913]1946: 351).

To support his argument, Poincaré draws examples from the history of science that exemplify precisely the survival/preservation of relations. Two main examples are worth citing here:

> This [i.e. the prediction of optical phenomena] Fresnel's theory enables us to do today as well as it did before Maxwell's time. The differential equations are always true… [they] 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 current*. But these are merely names of the images we substituted for the real objects which Nature will hide for ever from our eyes ([1905]1952: 160-1).

> In its primitive form, Carnot's theory expressed in addition to true relations, other inexact relations, the *débris* of old ideas; but the presence of the latter did not alter the reality of the others. Clausius had only to separate them, just as one lops off dead branches.

[43] See the section on Russell for additional insight into Poincaré's motivation.

> The result was the second fundamental law of thermodynamics. The relations were always the same, although they did not hold, at least to all appearance, between the same objects. This was sufficient for the principle to retain its value (165).

The first passage draws attention to the fact that Fresnel's equations survive the shift from the ethereal theory of light to the non-ethereal electromagnetic theory. The reason for this, according to Poincaré, is that they express real relations (and hence structures) between physical objects. By contrast, the elastic solid ether itself and the conception of light as consisting of disturbances transmitted through the ether are abandoned. The second passage draws attention to the fact that some of Carnot's postulated relations in his ideal theory of heat engines, such as the so-called 'Carnot cycle', survive the transition from the caloric conception of heat to thermodynamics. In this case, it is the caloric, i.e. the conception of heat as a material fluid, which gets abandoned.

In sum, Poincaré's point is that the history of science indicates a preservation of these relations (but not of their relata) from theory to theory. This, he takes to be a good reason why we should be epistemological realists about the relations between which the objects hold, but not the objects themselves. As we shall see in the following sections, the historical evidence for structural realism becomes less clear-cut as we move from the classical framework to the relativistic and quantum revolutions of the twentieth century.

Duhem

Like Poincaré, Pierre Duhem is often seen as a conventionalist. Recently, some authors (see, for example, Worrall (1989), Chakravarrty (1998), Gower (2000), and Zahar (2001)) have argued that he is either an epistemic structural realist or, at least, has close affinities to the position. I agree that there is a structuralist vein to Duhem's work but do not think the evidence unequivocally warrants his classification as an epistemic structural realist. In what follows I present a short exposition of his views.

A central distinction in Duhem's work is that between the explanatory and the representative parts of a theory. According to Duhem, the *explanatory* part of a theory is that 'which proposes to take hold of the reality underlying the phenomena' whereas the *representative* part is that ‘which proposes to classify laws’. Duhem likens the explanatory part to a parasite saying that:

> It is not to this explanatory part that theory owes its power and fertility; far from it. Everything good in the theory, by virtue of which it appears as a natural classification and confers on it the power to anticipate experience, is found in the representative part… On the other hand, whatever is false in the theory and contradicted by the facts is found above all in the explanatory part; the physicist has brought error into it, led by his desire to take hold of realities ([1914]1991: 32).

It is, thus, only the representative part of the theory that is doing the real work, i.e. that is producing the predictions. What, in Duhem’s mind, is the epistemological status of the representative part?

There are certainly several passages where Duhem ascribes an epistemic structural realist status to the representative part of theories. For example, when he says that physical theory “never reveals realities hiding under the sensible appearances; but the more complete it becomes... the more we suspect that the relations it establishes among the data of observation correspond to real relations among things” (26-27). And also a few pages later when he says “…we are convinced that they [i.e. the relations postulated by theories] correspond to kindred relations among substances themselves, whose nature remains deeply hidden but whose reality does not seem doubtful” (29). In another remarkable similarity to Poincaré’s position, Duhem claims that science’s historical record reveals a preservation of relations through theory change:

> When the progress of experimental physics goes counter to a theory and compels it to be modified or transformed, the purely representative part enters nearly whole in the new theory, bringing to it the inheritance of all the valuable possessions of the old theory, whereas the explanatory part falls out in order to give way to another explanation (32).

Given the context set up by the earlier passages, it seems safe to assume that the relations preserved through theory change reflect relations between physical objects.

Despite these striking examples, we need to take note of some important qualifications that Duhem makes in the same passages. Although he acknowledges the existence of a strongly felt *intuition* that our theories correspond to reality, he holds that the data of observation "cannot prove that the order established among experimental laws reflects an order transcending experience" (27). The belief in this correspondence is merely "an act of faith", says Duhem, which "assures us that these theories are not a purely artificial system, but a natural classification" (27). Thus, perhaps Duhem was an anti-realist after all.

Critics of this view will undoubtedly point out that no realist holds that we can *prove* the correspondence between theories and reality. That is, realists only claim that there are good reasons for holding such a belief. Moreover, given the centrality of faith to Duhem's thinking, the ascription of the phrase 'act of faith' to the belief that there is a structural correspondence between observation and the world does not seem as threatening. It could even be an indication of Duhem's strong support for the idea that the representative part of our theories corresponds to reality.[44]

Though the last comment is admittedly speculative, the plausibility of interpreting Duhem as an epistemic structural realist does not seem to be severely undermined, given his unequivocal claim about the preservation of relations through theory change. At any rate, Duhem is at least a structuralist of sorts. Depending on how much weight one assigns to the above qualifications, his position can be seen as a precursor to van Fraassen's latest position, viz. empiricist structuralism, according to which even the preservation of structure through theory change can be given an anti-realist explanation (see section 8 of this chapter).

Russell

It is quite unsurprising that Russell has a substantial role in the history of structuralism, given that he initiated, developed, and significantly contributed to most important debates in analytic philosophy. What is not widely realised is how strongly the concept of structure permeated his philosophical work.[45] One of his first

[44] Duhem was a devout Catholic who placed great importance on faith. Louis de Broglie, in the foreword to *The Aim and Structure of Physical Theory*, suggests that aspects of Duhem's faith were also extended to his philosophical concerns. For example he says "It was not that Pierre Duhem, a convinced Catholic, rejected the idea of metaphysics; he wished to separate it completely from physics and to give it a very different basis, the religious basis of revelation"(ix).

[45] It is worth quoting a comment from Hiram McLendon, one of Russell's students, who said of Russell's preoccupation with the concept of structure: "In fact, so fundamental and pervasive is Russell's use of this concept in *all his periods* of philosophizing and throughout each of his systems developed in *each of his major periods* that one might well survey most of his philosophy since 1912,

steps towards structuralism can be found in *The Problems of Philosophy*. Having recently read and been influenced by the British Empiricists, Russell regards the items of perception, which at the time take the form of ‘sense-data’, as the foundation of all knowledge. He argues that we have good reasons to believe that the causes of the sense-data we perceive are physical objects. But what can science tell us about physical objects? Russell’s answer is unmistakably clear:

> Assuming that there is physical space, and that it does thus correspond to private spaces, what can we know about it? We can know *only* what is required in order to secure the correspondence. That is to say, we can know nothing of what it is like in itself, but we can know the sort of arrangement of physical objects which results from their spatial relations… We can know the properties of the relations required to preserve the correspondence with sense-data, but we cannot know the nature of the terms between which the relations hold (1912: 15-16) [original emphasis].

And again a page later:

> Thus we find that, although the *relations* of physical objects have all sorts of knowable properties, derived from their correspondence with the relations of sense-data, the physical objects themselves remain unknown in their intrinsic nature, so far at least as can be discovered by means of the senses (17) [original emphasis].

Thus, Russell argues, we can know only the properties of the relations physical objects stand in, and not, as common-sense realism tells us, their intrinsic nature. This is patently an epistemic structural realist position.[46]

It is worth pausing here and comparing Russell to Poincaré. Russell’s Kantian remarks that we can know nothing about what space is ‘in itself’ and that the physical objects ‘themselves remain unknown in their nature’ share much with Poincaré’s own Kantian undertones. Unlike Poincaré, Russell holds that we only have access to the properties of relations between physical objects, not the relations themselves. This does not seem to amount to a real difference since knowing the relations without knowing the relata simply means knowing the properties of the

---

when he published *The Problems of Philosophy*, from the standpoint of his uses of the concept of similarity of structure [i.e. structural isomorphism]” (1955: 88). See also Michael Bradie (1977) where the development of Russell’s use of the concept of structure is traced from *The Analysis of Matter* to *Human Knowledge*.

[46] Even Russell’s theory of truth and belief, appropriately named ‘correspondence by congruence theory’, is structuralist (see (1912: ch.12)). The nature of the correspondence relation is one of congruence, i.e. isomorphism. The truth bearer is assumed to be structurally isomorphic to the physical state of affairs. For more on this see Kirkham (1992).

relations. What does seem, at first glance, different between the two philosophers is their motivation. Russell does not appeal to the history of science but rather to foundational considerations. A closer inspection of Poincaré's work, though, reveals that his motivation too was not *merely* historical but also foundational. In *The Value of Science*, Poincaré stresses that "nothing is objective which is not transmissible, and consequently that the relations between the sensations can alone have an objective value" (348). This idea follows from Poincaré's foundational concern that sensations are private and therefore intransmissible. Interestingly, Russell (1948: 485-6) makes similar remarks about the intransmissibility of everything but structure (see also Carnap (1928: §16) and Quine (1968: 161)). Conversely, it is not difficult to imagine Russell motivating his position with historical considerations. After all, if science identifies the properties of relations between physical objects, we should expect their preservation through theory change.

Russell's version of epistemic structural realism reached maturity in *The Analysis of Matter* (1927). There he argued that there are external causes of our perceptions, even though we should "not expect to find a *demonstration* that perceptions have external causes" (198) [my emphasis]. In fact, he devoted the twentieth chapter of this book to a causal theory of perception, rejecting "the view that perception gives direct knowledge of external objects" (197). We only have direct knowledge of the 'intrinsic character', 'nature', or 'quality' of percepts, i.e. the items of our perception. The only way to attain knowledge of the external world is by drawing inferences from our perceptions. To underwrite such inferences Russell employed a number of assumptions. The most important of these are:

Helmholtz-Weyl Principle (H-W): "…we assume that differing percepts have differing stimuli" (255). In short, *different effects* (*i.e. percepts*) *imply different causes* (*i.e. stimuli/physical objects*).[47, 48]

[47] Psillos (2001a) suggested this name for the principle on the basis of Helmholtz's and Weyl's appeal to it. It is worth noting that Russell sometimes uses the principle in its contrapositive (but equivalent) form, namely same causes imply same effects. Even Hume seems to endorse this principle as he advertises in the *Treatise* that "Like causes still produce like effects" (Book II, Part III, §1).

[48] Stimuli, according to Russell, are "the events just outside the sense-organ" (1927: 227). They are thus classified as physical events.

Mirroring Relations Principle (MR): "My point is that the relations which physics assumes… are not identical with those which we perceive… but merely correspond with them in a manner which preserves their logical (mathematical) properties" (252). In short, *relations between percepts mirror* (*i.e. have the same mathematical properties as*) *relations between their non-perceptual causes*.

For a closer examination of H-W and MR, I must ask the reader to wait until chapter three. For now suffice it to say that armed with these assumptions Russell argued that from the structure of our perceptions we can "infer a great deal as to the structure of the physical world, but not as to its intrinsic character" (400). More precisely, he argued that all that we can guarantee is that the structure of our perceptions is at most isomorphic to the structure of the physical world.

The notion of structure received a formal treatment from Russell. According to him, "[t]he 'relation-number' of a relation is the same as its 'structure', and is defined as the class of relations similar [i.e. isomorphic] to the given relation" (250).[49] The concept of isomorphic relations is employed here to convey the idea that the domain of interest is solely that of the properties isomorphic relations share. The motivation behind this idea arises from Russell's view that our epistemic access to the external world is indirect and, hence, cannot involve the unique identification of properties of, and relations between, physical objects.

Redhead (2001a) has called the notion of structure employed by Russell 'abstract structure'. To understand the notion of abstract structure we must first understand what it means for two structures to be isomorphic. A structure $S = (U, R)$ is *isomorphic* to a structure $T = (U', R')$ just in case there is a bijection $\phi: U \to U'$ such that for all $x_1, \ldots, x_n$ in U, $(x_1,\ldots,x_n)$ satisfies the relation R_i in U iff $(\phi(x_1),\ldots,\phi(x_n))$ satisfies the corresponding relation R_i' in U'. If, like Russell, one wants to talk about a particular relation being isomorphic to some other relation, one need not go further than the definition of isomorphism between structures, for any particular relation specifies a structure, namely a structure whose set of relations contains one, and only one, member. We can now define the notion of abstract structure: An *abstract*

[49] For more on the reason why Russell's notions of structure and relation-number are co-extensive see Solomon (1989).

structure Σ is an isomorphism class (or “isomorphism type”) whose members are all, and only those, structures that are isomorphic to some given structure (U, R). Qua isomorphism class, it can only identify the logico-mathematical properties of its members.

The notion of abstract structure is contrasted with what Redhead calls ‘concrete structure’. The former makes explicit that the domain of objects and the relations defined on these objects are not uniquely specified but only up to isomorphism. That is, whereas a concrete structure specifies one domain of objects that comes with a set of relations, an abstract structure just specifies a constraint as to which domains and relations qualify, namely those domains equinumerous to some given number and those relations that share the same properties.[50]

On the basis of these definitions we can now summarise Russell’s epistemic commitments as follows:

(REC1) Concrete observational structures.
(REC2) Abstract structures whose members are the concrete observational structures referred to in REC1.
(REC3) The existence of concrete physical structures that 1) have as domain members the causes of the concrete observational structures’ domain members referred to in REC1 and 2) are members of the isomorphism classes referred to in REC2.

Russell’s view can be presented as follows: Observational data falls into certain patterns allowing us to discover/postulate relations between observables.[51] Taking observables as our domain and collecting these relations into a set gives us the so-called ‘concrete observational structures’. They are concrete because their domain is specified uniquely. The abstract structures corresponding to these concrete observational structures can then be deduced in a straightforward manner by a *process of abstraction*. To do that, all one needs to do is to write down the

[50] The equinumerocity requirement simply reflects the fact that for there to be a bijection between two sets, the sets must have the same number of objects.
[51] This involves some sort of inference to the best explanation.

isomorphism class that the given concrete observational structure is a member of. By appeal to principles H-W and MR, we can then infer that to each concrete observational structure corresponds one, and only one, concrete physical structure such that: 1) the two are isomorphic, and 2) the domain members of the concrete physical structure, i.e. the physical objects, are causally responsible for the domain members of the concrete observational structure, i.e. the observables. Being isomorphic just means that the two concrete structures, i.e. the observational and the physical, are members of the same abstract structure, i.e. the same isomorphism class. The figure below illuminates the relationships between concrete observational, concrete physical, and abstract structures.

It is extremely important to note here that Russell's programme leans more towards an epistemological *reconstruction* of scientific knowledge rather than a description of what goes on in science. He does not claim that scientists actually observe first, and, solely on the basis of their observations, posit concrete observational structures that are then abstracted to a higher level, thereby allowing them to posit the existence of concrete physical structures instantiating the same abstract structure. The whole purpose of epistemological reconstruction is to offer a system through which knowledge claims can be evaluated, oftentimes ignoring the actual methods employed in science. In any case, the question whether reconstruction is a desirable enterprise, though interesting in its own right, will not be addressed in this dissertation.

Another important qualification present in Russell's work is that the relations postulated between observables might not always be exact.

> Hence we conclude that we have to do with a correlation which is usual but not invariable, and that, if we wish to construct an exact science, we must be sceptical of the associations which experience has led us to form, connecting sensible qualities with others with which they are often but not always combined (1927:182).

A consequence of this view is that the relations postulated to exist between the observables' probable causes inherit the inexactness. This qualification should be kept in mind when we are evaluating Russell's view in subsequent chapters.

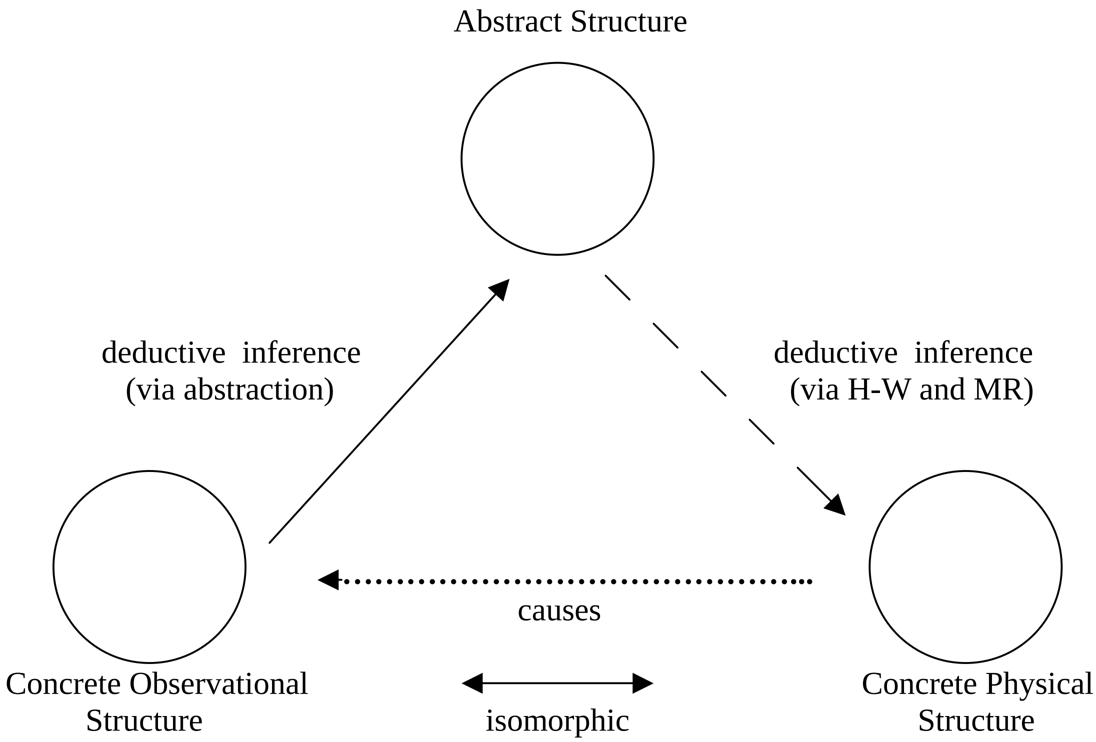

Figure 1: Russellian ESR

As indicated earlier, though Russell's epistemic commitments involve the properties of relations and Poincaré's involve the relations themselves, no real difference seems evident between them. Knowing the relations without knowing the relata simply reduces to the view that we can only know the properties of these relations. That is why we appeal to the notion of abstract structure. According to standard semantics, the interpreted terms of first order structures uniquely pick out individuals. This is something that advocates of ESR cannot sign up to, since they hold that we cannot uniquely pick out individuals. They thus resort to notions such as that of abstract structure. There is, however, another option for the ESR-ist, namely to change our understanding of standard semantics in order to accommodate the non-uniqueness of representation. As we shall see in section five below, Zahar advocates precisely such a change.

The Newman Objection

According to most commentators in the debate, the most serious objection against Russell's version of structural realism has been that of M.H.A. Newman in a critical review of *The Analysis of Matter*. Newman argued against Russell's claim that we can know only the (abstract) structure of the external world, alleging that this makes scientific knowledge *trivial*. Indeed, the only way to avoid the triviality accusation, according to him, is to abandon ESR. In the ensuing years, Newman's review received little attention until Demopoulos and Friedman (1985) unearthed it.[52] Let us now take a brief look at Newman's formulation of the objection and consider its two main claims.

First, Newman takes ESR knowledge claims to be trivial, for he takes Russell's structuralism to amount to assertions of the following type: "[*t*]*here* is a relation *R* such that the structure of the external world with reference to *R* is *W*" (1928: 144). He argues that, aside from indicating the required cardinality, these assertions are not saying anything of importance since we can derive the same assertions for any given class by appeal to the following theorem: "For given any aggregate A, a system of relations between its members can be found having any assigned structure compatible with the cardinal number of A" (140). In other words, given the right number of objects we can set up any structure we like. Yet, we expect knowledge of the external world to be the outcome of empirical investigation not of *a priori* reasoning. Indeed, the only information that requires empirical investigation under Russell's view, according to Newman's argument, is information about the size of a given class.

Second, Newman argues that the only way to avoid the triviality accusation is to give up ESR. This much, according to him, is evident in the idea "that it is meaningless to speak of the structure of a mere collection of things, not provided with a set of relations", and "[t]hus the only important statements about structure are those concerned with the structure set up … by a given, definite, relation" (140). The sole way to avoid trivialization, Newman holds, is to specify the particular relation(s) that generate(s) a given structure. That is, if we uniquely specify R, instead of just saying 'There is a relation R that has a certain structure W', the fact that R has structure W

[52] Solomon (1989) points out that Newman's objection had been unsuccessfully employed by R. B. Braithwaite (1940) in a review of Arthur Eddington's *The Philosophy of Physical Science*. Solomon argues that Braithwaite did not correctly understand Newman's objection. Moreover, he argues that Eddington (1941), in his reply to Braithwaite, despite being confused about the notion of structure, should have realised that the objection was inapplicable in his case. See below for some brief remarks on Eddington's account. Finally, I have discovered that McLendon (1955), as it seems independently of Newman, also raises the triviality accusation against Russell's position.

is no longer trivial. The problem is that to specify R, one inevitably goes beyond the epistemic commitments of the structural realist, thereby abandoning ESR.

In their article, Demopoulos and Friedman take Newman's objection as the definitive refutation of structural realism. They parade Russell's concession of the point in a letter to Newman (see Russell (1968: 176)) and his subsequent abandonment of the idea that our knowledge of the physical world is *purely* structural.[53] Interestingly, their presentation of Newman's objection is mainly a reconstruction that focuses on the Ramsey-sentence approach to theories. Following Grover Maxwell's suggestion, they argue that "it is possible to extract from the book [i.e. Russell's (1927)] a theory of theories that anticipates in several respects the Ramsey-sentence reconstruction of physical theories articulated by Carnap and others many decades later" (1985: 622).[54] After all, if all we can know about the external world is that there are relations that have certain properties, then the Ramsey-sentence seems like a good candidate to express such statements, because it existentially quantifies over all theoretical predicates – remember that relations are merely 2+*n*-place predicates – thereby allowing only assertions about properties of such properties or properties of such relations. Demopoulos and Friedman argue that if a theory is consistent and all its observational consequences true, then the truth of its Ramsey-sentence follows as a theorem of set theory or second-order logic. On the basis of the above association between the Ramsey-sentence and structural realism, they claim that Russell's position collapses into phenomenalism.[55] Given the gravity of Newman's objection and associated results, I will devote chapter four to a thorough analysis of these issues.

4. The Years in Between

After Russell, the next systematic epistemic structural realist was Grover Maxwell. In between the two, a number of eminent philosophers espoused different forms of structuralism, but these were not systematically developed and have not contributed much to the current debate. In this section I briefly note their views.

[53] For Russell's post-1928 work on structuralism, see his *An Inquiry into Meaning and Truth* and *Human Knowledge*.
[54] Ladyman (1998) has argued that the Newman objection is identical to an argument put forward by Jane English (1973).
[55] The only thing that distinguishes phenomenalism and Russell's structural realism, according to them, is that the latter makes a cardinality assumption with regard to the external world.

It was Demopoulos and Friedman who first pointed out that Moritz Schlick's position in *General Theory of Knowledge* is quite similar to Russell's structural realism. Like Russell, Schlick distinguishes between structure and quality/content and holds that our knowledge of the world is restricted to its structure. Unlike Russell, Schlick rejects the idea that we know the structure of our experience. For him, the term 'knowledge by acquaintance' is an oxymoron. We can know the structure of the world but we are only acquainted with the content or quality of our experience. Schlick thus draws a line between knowledge and acquaintance that perfectly coincides with his distinction between structure and content/quality.

One of the oddest types of structuralism ever proposed is that of Arthur Eddington (see his (1939)). In Eddington's mind, our knowledge of the world is structural. Thus far his epistemological stance is in agreement with Russell's and not at all unreasonable.[56] The oddity can be found in his rejection of an idea common to most scientists and philosophers of science, i.e. that our knowledge of the physical world is at least justified *a posteriori*. According to Eddington, our knowledge of the physical world is purely *a priori*![57] Needless to say, it is hard to get used to the idea that a statement as implausible as this one comes from a physicist of such grand stature.

The implausibility of his position notwithstanding, it is worth bringing up one of the main motivations for Eddington's structuralism, namely group theory. The spread of group theory in the twentieth century, from geometry to quantum mechanics, seems to have made a lasting impact on his philosophy.[58] As Eddington acknowledges, his understanding of the notion of structure is group-theoretical. He thus says: "What sort of thing is it that I know? The answer is structure. To be quite precise it is

---

[56] Steven French (2003) offers a more detailed analysis of Eddington's structuralism. Among other things, he argues that Eddington's structuralism has both epistemological and ontological implications, the latter leading to a position similar to the one advocated by French himself, viz. ontic structural realism. This form of structural realism is discussed in a section below.

[57] As Solomon (1989) has pointed out, this rejection makes Eddington's structuralism immune to Newman's objection because the latter is directed at claims that our knowledge of the external world is purely structural and a-posteriori.

[58] For more on this, see Steven French and James Ladyman (2003a: 50-51) but also French (2003). French and Ladyman argue that one other major motivation for Eddington's structuralism was the implications quantum physics had for the issue of the individuality of particles.

structure of the kind defined and investigated in the mathematical theory of groups" (147).

Another structuralist from the same period as Schlick and Eddington is Ernst Cassirer. French and Ladyman (2003a: 38-41) recently resuscitated Cassirer's views. More precisely, they make a convincing case that Cassirer advocates an ontological version of structuralism, according to which relations, and hence structures, are the primitive ontological components of the world. Cassirer certainly drew ontological lessons from the developments of the quantum and relativistic revolutions. He thus asked questions like "Is there any sense in ascribing to them [i.e. electrons] a definite, strictly determined existence, which, however, is only incompletely accessible to us?" (1936: 178). His answer to this question and others like it is a resounding 'no', since he conceives of electrons not as individuals but simply "describable as 'points of intersection' of certain relations" (180). He thus seems to reject the traditional object-based ontology for a relation-based ontology that reconceptualises an object in terms of relations.[59]

At around the same time as these authors, Carnap made several decisive steps towards structuralism. That Carnap had structuralist inclinations was first suggested by Demopoulos and Friedman (1985). In the *Aufbau*, as is well known, Carnap advocates the reconstruction of all scientific concepts on the basis of private experience. Yet, it is unclear what precisely Carnap wants to achieve (see Creath (1998)). Some, for example, suggest that Carnap simply tried to reduce physical objects to observable phenomena, implying a phenomenalist project. Against this interpretation, Demopoulos and Friedman suggest that there is an undeniable structuralist streak in the *Aufbau*. More specifically, they claim that for Carnap, only those statements that express the structure of experience reveal the objectivity of science. Here's a telling passage quoted by Demopoulos and Friedman:

> Science wants to speak about what is objective, and whatever does not belong to the structure but to the material (i.e., anything that can be pointed out in a concrete ostensive definition) is, in the final analysis, subjective. One can easily see that physics is almost altogether desubjectivized, since almost all physical concepts have been transformed into purely structural concepts (1928, §16).

[59] See also Cassirer (1944).

Carnap, they point out, sets up a program of defining scientific concepts as 'purely structural definite descriptions'. The important point to note is that these definite descriptions contain only logical vocabulary. This is a move similar to the Ramsey sentence, the only difference being that Carnap turns all the terms, i.e. not just the theoretical ones, into variables.

Recent work by Psillos (1999; 2000) has uncovered that Carnap defended a more robust form of structuralism in the fifties and sixties. For example, in 'The Methodological Character of Theoretical Concepts' Carnap holds that theoretical variables range over natural numbers but only because the domain of the naturals has a kind of structure that is isomorphic to the structure of the domain of the theory. Carnap signifies the importance of structure over its elements, saying that "the structure [of the domain of the theory] can be uniquely specified but the elements of the structure cannot" (1956: 46). In the years that follow, his structuralism becomes even more pronounced. The most important development is his reinvention of the Ramsey-sentence approach, under the name of 'the existentialised form of theories'.[60] He avoids a realist interpretation by holding that in Ramseyfication the theoretical terms are to be replaced by variables that range over mathematical entities.

Carnap's agenda, throughout this period, seems to have been to uphold a neutral stance towards the realism-instrumentalism debate. As made obvious above, however, his insistence on the interpretation of theoretical variables ranging over mathematical entities, as opposed to physical entities, tips the balance in favour of the instrumentalist side. In a move to avoid instrumentalism, Carnap explains that the variables in his system have two interpretations, one extensional and one intensional.[61] From an extensional point of view, the theoretical variables of the Ramsey-sentence range over mathematical entities. From an intensional point of view, the theoretical variables of the Ramsey-sentence can be seen as ranging over physical entities in that the intensions of theoretical terms are physical concepts not

[60] Psillos cites a letter from Carnap to Hempel (dated February 12, 1958), where Carnap reveals that he had read Ramsey many years before he developed his own existentialized form of a theory but had completely forgotten about it.
[61] See Psillos (1999: 54) where he cites a letter from Feigl to Carnap (dated July 21, 1958).

mathematical ones. This tips the balance in favour of the realist side since he allows the Ramsey-sentence to make existential statements about unobservable entities.

Carnap struggles with these issues through various manuscripts, letters, and articles.[62] As Salmon (1994) indicates, it is not until Grover Maxwell's intervention that Carnap's attitude towards the Ramsey-sentence settles. Through Maxwell's influence, Carnap comes to see the Ramsey-sentence as incompatible with instrumentalism, since it can both attain a truth-value and make existential statements about physical entities. Even though Carnap adopts this view by 1974, he, unlike Maxwell, neither associates the Ramsey-sentence with structural realism nor embraces the latter. With these brief remarks on Carnap's structuralism completed, it is time to turn to the Ramsey-inspired structural realists, starting with Maxwell.

5. Epistemic Structural Realism, Ramsey-Style

Maxwell

In the late sixties, Grover Maxwell published a number of articles, defending an epistemic version of structural realism that owes much to Russell. Maxwell traces the position to Poincaré, Schlick, Wittgenstein and, naturally, Russell himself.[63] Echoing his predecessors, he speaks of the inability to have direct knowledge of the external world in distinctly Kantian terms:

> On the one hand there is the realm of phenomena. These are wholly *in the mind* (in our sense). Of the phenomena and only of the phenomena do we have *direct knowledge*. On the other hand, there are the things in themselves, and here our divergence from the views of Kant is great; although we have no *direct* knowledge of the latter, the bulk of our common sense knowledge and our scientific knowledge *is* of them… all of this knowledge is purely structural (1968: 155).

Closely adhering to Russell's version of structural realism, Maxwell urges commitment to the view that "all of the external world including even our own bodies is unobserved and unobservable" (152). He is thus using the term

[62] In the end, Carnap manages a type of neutrality, but one that is between realism and the Ramsey-sentence, not realism and instrumentalism (see Psillos (1999: 58-61)).
[63] He also mentions Beloff (1962), Mandelbaum (1964), Aune (1967), Pepper (1967). Maxwell (1970b: 24) also claims that it is similar to the 'critical realism' of, among others, Roy Wood Sellars as well as to the representative realism of Locke provided certain modifications are made.

‘unobservable’ in a way that is different from its use today. Like Russell, he does not discriminate between macro and micro-physical objects. For them, the term ‘unobservable’ denotes the set of all things inhabiting the external world, i.e. the set of all non-mental entities. Their claim, of course, is not that our observations have no causal origins in the external world, but rather that what we directly observe is ‘wholly in our mind’. Unless otherwise noted, I will henceforth employ Russell and Maxwell’s meaning for the term ‘unobservable’.

Despite their agreement on what ‘unobservable’ denotes, there are certain differences between Russell and Maxwell that are worth pursuing. One difference is that Maxwell dissociates himself from reifying observable units, avoiding reference to things like sense-data, sensibilia, percepts, etc. (151). Instead, he places the spotlight on the linguistic level, with observation sentences and predicates as primitives. Somewhat paradoxically, Maxwell is best known for his critique of the distinction between observational and theoretical terms (see his (1962)). Yet, he here seems to advocate a strong distinction between observable and unobservable that is essential for his version of structural realism. The apparent tension is dissolved if one takes into account that, for him, the entire external world is unobservable. That is, this way of delineating the observable from the unobservable avoids the kinds of objections Maxwell raised in his earlier work. For example, seeing through instruments is no longer a threat to the observable-unobservable distinction since *all* seeing is restricted to the perceptual world.

Given the sense of ‘unobservable’ just sketched, how can knowledge of a wholly unobservable external world be had? The answer, says Maxwell, lies in the causal theory of perception. An important feature of this theory is that “it is *not* essential to the position [i.e. structural realism] that the sense impressions or perceptual experiences, or whatever we decide to call them ‘resemble’ the physical objects which may be among their causal antecedents” (1968: 155). What is necessary is that “at least a certain subset of the features of the [sense] impression are isomorphic with a subset of the features of the physical object” (156). Without this type of ‘resemblance’, Maxwell insists, there can be no knowledge of the external world. His justification for this requirement proceeds via familiar Russellian techniques such as the claim that causal chains leading up to our perceptions are structure-preserving (1970b: 25) and the H-W principle (1968: 156).

Maxwell, like Russell, argues that the motivation for the causal theory “is virtually forced upon us by common sense as well as by science” (1970b: 23). In some limited sense, this is right. Most of us, after all, would agree that the causes of our perceptions originate in the external world. However, there is no widespread agreement on how the ‘information’ coded in our perceptions represents the external world if it does so at all. In other words, the claim that perceptions preserve the structure of their causes is more difficult to swallow. Maxwell admits that “there are no purely logical or purely conceptual reasons that there be structural similarities between objects in the external world and items in our experience” (25). Nevertheless, he claims that well-confirmed theories support this assumption, arguing that “if such [structural] similarities were fewer or, even, virtually nonexistent, knowledge of the physical realm would be more difficult to come by but not necessarily impossible” (25).

Maxwell claims that we cannot know the first-order properties of physical objects; we can only know their second or higher order properties, what he calls ‘structural properties’ (18). This is supposed to follow from the idea that first-order properties of phenomena, like colours, need not resemble the first-order properties of their causes. Maxwell’s conclusion is that “[w]*hat holds of colors must also be true for all of the first order properties that we perceive directly*” (19) [original emphasis].

Maxwell praises Russell, among other things, for the reconciliation of realism with the logical positivist verificationist principle. This is achieved, Maxwell claims, through Russell’s principle of acquaintance and his distinction between knowledge by acquaintance and knowledge by description. The principle of acquaintance is a close relative of the verificationist principle, for it states that to understand a proposition we must be acquainted with all of its constituents. With some perhaps not so trivial adjustments to the terminology, Maxwell transposes this idea to the current context, claiming that all descriptive terms in a meaningful sentence must refer to ‘items’ of our acquaintance, i.e. all descriptive terms *must* be observation terms (as opposed to theoretical terms).[64] Yet realism requires that we have knowledge of items with which we are not acquainted. This is where Russell's knowledge by description comes in, for it allows an object to be known by a list of descriptions – i.e. without our first being acquainted with it. Needless to say Maxwell takes knowledge by description to be the same as knowledge via theory.

As I mentioned earlier, one of Maxwell's contributions to the debate is the bridge he forges between the Ramsey-sentence approach and structural realism. It is at this point that the utility of the principle of acquaintance and the acquaintance vs. description distinction becomes evident. According to Maxwell, knowledge representation via the Ramsey-sentence approach validates both the principle and the distinction. This is so, because the Ramsey-sentence approach existentially quantifies over all theoretical terms but leaves all observation terms intact. Maxwell explains:

> [We] can formulate propositions that refer to unobservable properties or to classes of unobservable things by means of existentially quantified predicate variables and other purely logical terms plus terms whose direct referents are observables. Fortunately any theory whatever can be transformed without loss of significant content into such a proposition. It is only necessary to replace the conjunction of the assertions of the theory by its Ramsey sentence (16).

In accordance with Russell's principle of acquaintance, the 'items' that theoretical terms supposedly refer to, unlike the items of observation terms, are not 'ingredients' of a proposition. For Russell, this means that sentences expressing such a proposition will not contain a name or descriptive constant that refers directly to the alleged items. Diverging from Russell's viewpoint, Maxwell argues that there is a sense in which a proposition *refers* to the items that its theoretical terms prescribe. It refers to them indirectly, through "(1) terms whose *direct* referents are items of acquaintance and (2) items of a purely logical nature such as variables, quantifiers and connectives" (1970a: 182-3).

The advantage of employing the Ramsey-sentence approach is that its assertions are restricted to properties of properties of unobservables, i.e. it does not uniquely

[64] He thus assumes that the terms 'observation' and 'acquaintance' are co-extensive (1970a: 182). Notice that his deliberate choice of the term 'item' reflects his avoidance of what exactly the nature of the objects of our acquaintance is.

identify the properties of unobservables. This *seems* in accord with Maxwell's view that we do not have epistemic access to the first-order properties of unobservables.[65] Nonetheless, "our (Ramseyfied) theories tell us that they exist and what some of *their* (second and higher order) properties are" (1970b: 19) [original emphasis].

To appreciate the marriage between structural realism and the Ramsey-sentence approach, it is worth considering one of Maxwell's examples. Suppose that given numerous observations we pronounce the truth of the following sentence: $(\forall x)(\forall y)[(Ax \,\&\, Dx) \supset (\exists y)\, Cy]$ where *A* and *D* are theoretical predicates which stand for 'is a radium atom' and 'radioactively decays' respectively, and *C* is an observation predicate which stands for 'is an audible click in a Geiger counter'. If this sentence is true then its Ramsey-sentence, namely $(\exists\psi)\,(\exists\varphi)\,(\forall x)\,(\forall y)\,[(\psi x \,\&\, \varphi x) \supset (\exists y)\, Cy]$ where 'ψ' and 'φ' are predicate variables, will also be true. The principle of acquaintance holds that we cannot know sentences like the first one, because they mistakenly include fully interpreted theoretical predicates, i.e. *A* and *D*. The Ramsey sentence version circumvents this problem by merely asserting that such properties exist. Maxwell explains that our knowledge of these properties "is by description and, as in all such cases, we refer to them not by predicate constants, but indirectly by means of purely logical terms plus an observation term, in this case, 'C' " (1970a: 186-7).

Despite the strong case that Maxwell makes, Russell's version of structural realism and the Ramsey-sentence approach are inconsistent. True, both Russell and Maxwell advocate a notion of structure that identifies properties preserved by isomorphic mappings.[66] It is also true that the notion of abstract structure I presented earlier seems ideal for the purposes of both. In spite of this agreement, the Ramsey-sentence of a theory preserves the *logical structure* of the whole theory, something directly at

[65] As I argue in the next chapter (see section three), the Ramsey-sentence approach is not in accord with Maxwell's idea about first-order properties, for it quantifies over any theoretical properties, i.e. not just those that are first-order.
[66] In a puzzling footnote, Maxwell notes that the account he offers in his (1968) is incomplete and incorrect in that "structure should not be identified with form; rather it is form plus causal connections with experience" (154). I do not know what to make of this, though my suspicion is that he might be attempting to fend off objections on how much the notion of abstract structure can tell us about the world.

odds with Russell's insistence that we infer the structure of the world from the structure of our perceptions.

To elucidate the point, consider the following example. Suppose that we have in our hands a theory, call it 'K', and that all it says about the world is captured by the claim: $(\forall x)\ [(T_1x \supset T_2x)\ \&\ (O_1x \supset \neg O_2x)]$.[67] Now, according to Russell, we find out about the structure of the physical world through the structure of observations. First of all, we take the concrete observational structure of K, i.e. $(\forall y)\ (O_1y \supset \neg O_2y)$, call it '$O_K$'. We then deduce the abstract structure of O_K, i.e. $(\exists\Phi)(\exists\Psi)(\forall y)\ (\Phi y \supset \neg\Psi y)$, call it '$A_K$'. Finally, via principles H-W and MR we postulate that there is a unique concrete physical structure, call it 'P_K', which instantiates A_K and whose domain members are the causes of the domain members of the concrete observational structure. We can express P_K as $(\forall y)\ (Fy \supset \neg Gy)$, where F and G are predicates referring to physical properties. Qua structural realists, we do not have epistemic access to the properties F and G are referring to, so we cannot say that we know P_K. All that we can say is that we know that there exist two predicates that: 1) refer to the physical properties that cause observable O_1 and O_2 and 2) that these predicates instantiate the predicate variables in A_K. We can call this last claim 'K_P'. The point that I am making here is that K_P is obviously different to the Ramsey-sentence of K, R(K): $(\exists\Theta)(\exists\Sigma)\ (\forall x)\ [(\Theta x \supset \Sigma x)\ \&\ (O_1x \supset \neg O_2x)]$. One major difference is that the Ramsey sentence of K asserts the existence of *at least* two physical properties, whereas K_P asserts the existence of *just two* physical properties. Moreover, the latter states that the two properties are the causal antecedents of O_1 and O_2, something R(K) does not do. Another major difference is that the logical properties of R(K) and K_P, at least in this example, are different. That suffices to establish that the two methods, i.e. Ramseyfication and Russell's method, are not equivalent. No wonder then that even Maxwell remarks in passing that "the Ramsey sentence is *approximately* equivalent to Russell's contention that we do have knowledge of the structural properties of the unobservable" (1970b: 17) [my emphasis].

[67] Worrall objects that this example is artificial, for K does not involve any intricate logical relations between the observational and theoretical terms. This, according to him, makes the theoretical part of the sentence content-free. I will address this issue in section three of chapter four.

Worrall and Zahar

Worrall's and Zahar's variety of epistemic structural realism, initially also branded 'syntactic realism', is inspired by Poincaré's historical arguments, and in this respect differs from both Russell's version and Maxwell's version. The recent interest in structural realism was instigated by the publication of Worrall's 'Structural Realism: The Best of Both Worlds?' a decade and half ago. Worrall there argued that a sensible position in the scientific realism debate needs to take into consideration two warring arguments: the no-miracle argument and the pessimistic induction argument.[68] In short and as already sketched in chapter one, PI holds that since predictively successful scientific theories have eventually been discarded, we have inductive evidence that even our current theories, despite their great successes, will also be discarded one day. NMA holds that realism is the *only* view that does not make the predictive success of science a miracle. Worrall offers ESR as a position that underwrites both of these arguments and situates itself midway between constructive empiricism and traditional scientific realism. It underwrites the NMA because it argues that the success of science reflects the fact that we have got the structure of the world right. It underwrites PI because it concedes that non-structure gets abandoned.

Following Poincaré, Worrall takes the Fresnel-Maxwell case as historical evidence for ESR. He indicates that the structure of Fresnel's theory, as it is for example expressed through his equations for the relative intensities of reflected and refracted light at the boundary between two transparent media of differing optical densities, was carried over to Maxwell's theory unscathed. Thus, Worrall argues, if we look at theory change solely from the perspective of mathematical equations, the Fresnel-Maxwell case counts as evidence for the essentially cumulative development of science.[69] The underlying assumption is that it is reasonable to hold that what survives theory change is what has really latched on to the world. According to Worrall, Fresnel was completely wrong about the *nature* of light, viz. that light consists of vibrations that are transmitted through an all-pervading medium, the ether. Fresnel was probably right, however, about its *structure*, i.e. that optical

[68] Worrall traces PI and NMA to both Poincaré and Duhem (see (1989: 140-2)).
[69] Heinrich Hertz's often quoted comment that 'Maxwell's theory is the system of Maxwell's equations' is congenial with Worrall and Poincaré's claim that the essence of the theory is the relations it postulates.

effects depend on something or other that vibrates at right angles to the direction of propagation of light, just as required by the equations.

A question that naturally arises from the above exposition of Worrall's views is whether the mathematical continuity found in the above case is a widespread phenomenon within the history of science. Worrall grants that the Fresnel-Maxwell case is 'unrepresentative' in that Fresnel's equations are entailed by Maxwell's theory without any modifications.[70] It is more often the case that equations of an older theory reappear only as limiting cases of equations in a newer theory. Indeed, the two great theories of the twentieth century, viz. the theory of relativity and quantum mechanics, depart from classical physics in ways that prima facie seem difficult – some people have argued impossible – to reconcile.

Redhead (2001a), himself an ESR sympathiser, cites two cases where the structural continuity between old and new is difficult to maintain. The first case involves the relationship between Minkowskian and Galilean space-times. Unlike Galilean space-time, Minkowskian space-time admits a non-singular metric. If, however, we let the speed of light tend to infinity, the metric becomes singular. This leads to the disappearance of the relativity of simultaneity, allowing for the recovery of Galilean space-time. The second case involves the relation between the Poisson and Moyal bracket formulations of classical and quantum mechanics respectively. The latter formulation generalises the former by introducing non-commutative multiplication for phase space functions. If we set Planck's constant to zero, commutativity is recovered and so is the Poisson formulation.

Redhead's two cases are meant to illustrate an abrupt qualitative discontinuity between the new and the old. Regardless of this discontinuity, Redhead notes an apparent affinity between old and new structures:

> Qualitatively new structures emerge, but there is a definite sense in which the new structures grow naturally, although discontinuously, out of the old structures. To the mathematician introducing a metric in geometry, or non-commutativity in algebra are very natural moves. So looked at from the right perspective, the new

[70] That the Fresnel-Maxwell case is atypical has also been pointed out by, among others, Howson (2001), Redhead (2001a), and Kitcher (2001).

> structures do seem to arise in a natural, if not inescapable way out of the old structures (19).[71]

In other words, if, like the mathematician, we see how natural the leap is from old to new structure, then we realise that the discontinuity is not debilitating. Seeing as this argument rests on a metaphor, it is no wonder that Redhead is reticent regarding its force. A major task for the structural realist then is to find a way to make concrete the correspondence relation between old and new structures.

Zahar has recently claimed that a proper defence of ESR requires a departure from standard semantics. By interpreting relations only through their relata, he maintains, standard semantics fails to give due priority to the relations, which are, after all, the focus of structural realism. Here is what Zahar says:

> …according to structural realism, we often have a good reason for supposing that 'R' [i.e. a specific relation] reflects a real connection between elements about whose intrinsic nature we know next to nothing. The conditions under which we are entitled to make such a realist claim obtain whenever we have a highly unified hypothesis H which both involves R and explains a whole host of seemingly disparate facts in a non-adhoc way (2001:38).

Implicit in this passage is an association between knowledge of the intrinsic nature of objects and classical semantics. In rejecting the former, Zahar believes that we must also reject the latter. The presumed association, however, is highly dubious since not knowing the intrinsic nature of objects does not force us to abandon the characterisation of relations in terms of individuals. We can simply stick with the less radical view that the individuals are known only up to isomorphism, expressing our knowledge of relations as higher order claims about sets of individuals. What is more, Zahar's and Worrall's support for the Ramsey-sentence approach does not seem to square with his call for a new semantics. Either, the Ramsey-sentence approach plus the associated classical semantics works, in which case there is no need for a new semantics, or it does not work and that is (potentially) why we must look for a new semantics.

---

[71] See also his (2001b: 346-347).

Another interesting development has seen the reconciliation of structural realism with a position proclaimed by many (see, for example, Niiniluoto (1999)) as its main competitor in the realist camp, namely entity realism. In a noteworthy article, Chakravartty (1998) has sought to bring the two together under the banner of his own position, 'Semi-realism'. He argues that the properties we detect in experiments should be central to both accounts. Commitment to the existential claims of entity realism, says Chakravartty, can be achieved only through relying on relations between detectable properties. Conversely, these relations, which are the focus of structural realists, contain substantive information about entities. Thus, he concludes, properly understood entity and structural realism "entail one another; they are, in fact, one and the same position: semirealism" (407).

6. Psillos' Objections

In a recent succession of articles ((1995), (2000a), (2001a), 2001b)) and a book (1999, ch. 7), Psillos has attacked various versions of structural realism, especially those of Russell, Maxwell and Worrall. Since a proper exposition of these criticisms would take rather long, I merely list them here and ask the reader to wait for a detailed treatment in the ensuing chapter. The seven most important objections that emerge from Psillos' attack, all of them directly challenging ESR, are:

(PS1) ESR commits us only to uninterpreted equations but these are not by themselves enough to produce predictions (1999: 153-4).

(PS2) Structural continuity through theory change can be explained better by traditional scientific realism than by ESR (1999: 147-8).

(PS3) Some non-structural theoretical content is retained in theory change, and this is better supported by current evidence and more likely to be true than non-structural theoretical content in the past (1999: 147-8).

(PS4) The structure vs. nature distinction that ESR appeals to cannot be sustained (1999: 157).

(PS5) ESR faces a dilemma: On the one hand, the H-W principle by itself can only establish a relation of embeddability between the external world and the 'world' of percepts, not a relation of isomorphism as required by ESR. Without a relation of isomorphism, the structural realists cannot establish inferential knowledge about the structure of the external world. On the other hand, H-W together with its converse, viz. different stimuli/physical objects imply different percepts, allow for the establishment of isomorphic relations but, in doing so, concede too much to idealism (2001a: S13-S16).

(PS6) The claim that the first-order properties and relations of unobservables are unknowable in principle cannot be justified (1999: 156; 2001a: S20-21).

(PS7) Knowing the abstract structure of the external world is not enough since it merely amounts to knowing formal properties such as transitivity, symmetry and reflexivity (2001a: S16-S17).

7. Ontic Structural Realism

An altogether different species of structuralism has been proposed by James Ladyman (1998). Ladyman argues that structural realism should be understood not just as an epistemological, but also as a metaphysical position. He claims that this much is suggested by Worrall's version of structural realism, which, according to Ladyman, is ambiguous between the two manifestations. Yet, neither Worrall nor any other ESR-ist adopts any substantive metaphysical positions but rather asserts the epistemic inaccessibility to physical objects beyond the level of isomorphism. Steven French (1998; 1999; 2003) has joined forces with Ladyman (see their (2003a) and (2003b)) in advocating what they call 'Ontic Structural Realism' or OSR for short. As I have already indicated, they claim to have traced the roots of OSR to Ernst Cassirer. But let us take things from the beginning.

Appealing to some of the aforementioned objections to structural realism, particularly Newman's objection and Psillos' objection that the distinction between structure and non-structure cannot be drawn, Ladyman hastily concludes that ESR is incapable of solving the problem of ontological discontinuity through theory change.[72] He ties this problem to a type of underdetermination that originates in the philosophy of physics, namely whether elementary particles are individuals. A solution to this latter problem requires drastic measures, according to Ladyman:

> What is required is a shift to a different ontological basis altogether, one for which questions of individuality simply do not arise… So we should seek to elaborate structural realism in such a way that it can defuse the problems of traditional realism, with respect to both theory change and underdetermination. This means taking the structure as primitive and ontologically subsistent (1998a: 420).

This is the crux of OSR. Crudely put, OSR prescribes that all that exists in the world is structure. Consequently, all that can ever be known about the world is structure.

The motivation for OSR comes from considerations about modern and, in particular quantum, physics.[73] In classical physics, elementary particles are taken to be indistinguishable individuals. More precisely, they are only distinguishable with respect to their spatio-temporal coordinates but not with respect to any other properties they posses. Their individuality is thought of as something over and above these latter properties. The quantum view of elementary particles, say French and Ladyman, underdetermines the metaphysics of elementary particles. That is, they can be viewed as either individuals or non-individuals.

To illuminate this point let us take French's example of two indistinguishable particles that are distributed over two states (see his (1998)). The scenario obviously offers four possibilities: (1) particles *a* and *b* in state *A*, (2) particles *a* and *b* in state *B*, (3) particle *a* in state *A* and particle *b* in state *B*, and (4) particle *a* in state *B* and particle *b* in state *A*. Under the orthodox view of quantum statistics, (3) and (4) are thought of as one and the same possibility with nothing distinguishing between them.[74] That is, according to the Bose-Einstein statistics implicit in the orthodox interpretation of quantum theory, these two possibilities are considered to be the very same thing. French takes this to mean that the particles must be thought of as non-individuals. He concedes, however, that there is another view within quantum statistics that, at least in principle, takes (3) and (4) as distinct. From this, he

---

[72] It is puzzling how Ladyman comes to think that these problems have anything to do with the problem of theory change.
[73] Paradoxically, French (1998) argues that we cannot read metaphysics off current physics.
[74] In classical statistical mechanics, (3) and (4) are thought of as distinct.

concludes that quantum physics underdetermines, i.e. is neutral, between the view of particles as individuals and that of particles as non-individuals.

Astonishingly, French and Ladyman claim that this underdetermination supports OSR. Yet, at most, the underdetermination seems to raise doubts about the individuality of particles.[75] A defence of the view that we should throw away an individual-based ontology and reconceptualise the role of individuals in terms of structures, as French and Ladyman suggest, would at least require demonstrating that elementary particles are non-individuals. Their own insistence on underdetermination between the two possibilities defeats any such approach. If anything, the underdetermination counsels agnosticism between the two views of particles and, by extension, agnosticism about OSR.[76]

It is worth noting that in his original article, Ladyman offers the model-theoretic (a.k.a. semantic) approach to theories, according to which theories are conceived of as sets of models, as a general framework for the treatment of theories. Together with French, they have since then extended this framework with the so-called 'partial structures approach', developed first by Newton Da Costa (see, for example, Da Costa and French (1990)). Among other benefits, this approach allegedly provides a better representation of continuity through theory change, especially those cases where continuity is only approximate. As French and Ladyman have admitted, however, the radical shift in ontology requires a new semantics to go with it. What remains wanting, is the fleshing out of this much-advertised new semantics. Many have questioned the very possibility of conceiving of objects as structures. Moreover, even if the vaunted reconceptualisation were possible it is doubtful that this would be a good reason to abandon an object-based ontology.

In view of the fact that my dissertation deals solely with the epistemological dimension of the scientific realist debate, OSR will not be investigated further. It is mentioned here only in order to cover all major developments in structuralism of the

[75] This objection, as well as many others, has also been raised by Tian Yu Cao (2003a; 2003b), Anjan Chakravarrty (2003) and Matteo Morganti (forthcoming).
[76] Ladyman seems to have had a change of heart. When I raised this point at a BSPS lecture given by French and entitled "*From Poincaré's Crutch to Melia's Weasel: Having One's Ontological Cake and Eating it too*" Ladyman agreed that what the underdetermination argument warrants is agnosticism.

natural sciences. Notwithstanding the perhaps insurmountable difficulties it faces, OSR is at the cutting edge of metaphysics, its proposal as radical as they get.

8. Empiricist Structuralism

Van Fraassen (1997; 1999) has vehemently attacked structural realism, both its epistemic and ontic forms, arguing instead for an empiricist version of structuralism, which he aptly calls 'empiricist structuralism'. He agrees with Worrall that there is a preservation of structure through theory change, but argues that the type of structure involved is the structure of the phenomena, not the structure of the unobservables (1999: 30-1).[77] In van Fraassen's eyes there are two realms of scientific investigation: 1) the phenomena and 2) the mathematical structures. We represent the structure of the phenomena with the help of mathematical structures.

Van Fraassen claims that the empiricist can explain how and why earlier theories were successful. Instead of the realist explanation that requires old theories to have latched on to the structure of the unobservables, his alleged explanation requires that the new theories imply "approximately the same predictions for the circumstances in which the older theories were confirmed and found adequately applicable" (25). This, according to van Fraassen, doubles up as a criterion for theory acceptance. That is, a new theory must at least be able to make approximately the same confirmed predictions as the old one. It also satisfies the no miracles intuition, continues van Fraassen, without making the success of science a miracle, "because in any theoretical change both the past empirical success retained and new empirical successes *were needed as credentials* for acceptance" (25) [original emphasis].

The motivation for van Fraassen's structuralism is different from any of the ones we have seen so far. It is worth quoting him in full:

> According to the semantic approach, to present a scientific theory is, in the first instance, to present a family of models - that is, mathematical structures offered for the representation of the theory's subject matter. Within mathematics, isomorphic objects are not relevantly different; so it is especially appropriate to refer to mathematical objects as "structures". Given that the models used in science are mathematical objects, therefore, scientific theoretical descriptions are structural;

[77] As van Fraassen notes: "There was something they [i.e. the theories] got right: the structure, at some level of approximation, of those phenomena" (31).

> they do not "cut through" isomorphism. So the semantic approach implies a structuralist position: science's description of its subject matter is solely of structure (1997: 522).[78]

Given that mathematical objects can only be described up to isomorphism, van Fraassen says, our use of mathematical structures to describe the phenomenal world makes us structuralists. The motivation is thus primarily linguistic, in that he is arguing that language, in particular mathematics, gives rise to, and perhaps even necessitates, structuralism.[79] This linguistic motivation is reminiscent of the view, held by Poincaré and Russell, that nothing other than structure is transmissible. It is worth noting that, despite the jargon used by semantic theorists, the semantic approach is superfluous in the above argument since one need not be a semantic theorist to hold the two premises, i.e. that mathematical objects are describable up to isomorphism and that we use mathematical objects to represent the world.

Following in van Fraassen's footsteps, Otávio Bueno (1997; 1999; 2000) argues for a position that he calls 'structural empiricism'. His position inherits some of the main features of constructive empiricism, such as the notion of empirical adequacy, but also van Fraassen's recent emphasis on structures. His notion of structure, however, is a bit more idiosyncratic. Like French and Ladyman, Bueno relies on a partial structures approach to scientific theories. Within this framework, he introduces variant notions of empirical adequacy, such as the notion of degrees of empirical adequacy, characterised in terms of the notion of partial isomorphism (see his (1999: section 3)). Indeed, Bueno takes himself as extending van Fraassen's account by fleshing out a more flexible relation between structures, provided by the partial structures approach. This has been a move already suggested by van Fraassen (see his (1997: 524)), and Bueno acknowledges as much.

9. The Main Structural Realist Obstacles

Given the above elaboration of the historical development of structuralism and in particular structural realism, we can identify four main obstacles that structural realists need to somehow account for:

[78] See also his (1999: 31-2).
[79] Though van Fraassen does not elaborate on this point, I presume that his claim is not restricted to the language of mathematics but any language whose statements can be translated into the language of mathematics.

(SRP1) The Newman objection: The ESR claim that all we can know about the world is that it has a certain abstract structure makes scientific knowledge trivial. It presumably makes it trivial because saying, as structural realists do, that for a given class there exists a system of relations that specifies an abstract structure is not saying anything of empirical importance, since this claim follows a-priori by appeal to theorems of set theory and logic plus knowledge of the cardinality of the given class. In order to avoid the triviality accusation, appeal must be made to non-structural considerations. This amounts to abandoning pure ESR.

(SRP2) The structural discontinuity objection: There is insufficient historical evidence for structural continuity through theory change. The Fresnel-Maxwell case is atypical. Most current theories' immediate predecessors are, even at the level of structure, discontinuous with their successors.

(SRP3) Psillos' medley of objections: PS1-PS7.

(SRP4) The Empiricist Structuralist Challenge: There is continuity of structure through theory change, but it is continuity of the structure of phenomena not of the structure of unobservables.

These are added to the obstacles faced by scientific realism outlined in chapter one. The only exception is SRP2, which, naturally, overlaps with RP2 to a certain extent.

10. Conclusion

The history of structuralism in the natural sciences is rich and varied. Among the many structuralist positions, ESR, especially Worrall's version, has been hailed by many as a refreshing new hope for realism. As we have seen, it has also been heavily criticised. The rest of this dissertation will be an evaluation of ESR in light of the objections raised against it and, more broadly, the objections raised against traditional versions of realism. First in line is chapter three where I address Psillos' objections PS1- PS7, who spearheads the critique of structural realism. In addressing these objections I will try to clarify and make precise the notions and principles on which structural realism depends. Chapter four is devoted to the Newman objection, SRP1, hailed by many as *the* fatal blow to structural realism. In chapter five I pursue a historical case study in an effort to address the historical objections SRP2, and RP2. Following that is a chapter on underdetermination where RP1 and, for reasons that will become clear later, SRP4 are tackled. Finally, the seventh chapter offers a glimpse into the future directions the research of this dissertation can be taken.

3

Recent Objections

Among ESR's critics, Psillos seems the most committed to the position's overthrow. In the last few years he has launched a barrage of objections against ESR (see his (1995), (1999), (2000a), (2001a), and (2001b)). The aim of this chapter is threefold: 1) to evaluate Psillos' offensive against both the Poincaréan/Worrallian and the Russellian versions of the position, 2) to elaborate more fully what ESR involves, and 3) to suggest improvements where ESR is indeed failing.

1. Introduction

Psillos has praised Worrall's revival of the Poincaréan version of ESR, saying that it gives us an important insight into the scientific realism debate. The insight it gives is that we need not believe to an equal degree all that a scientific theory ascribes to the world. In spite of this acknowledgement, Psillos thinks that ESR faces a number of insurmountable objections. I have already listed these in the previous chapter but I reproduce them here for the reader's convenience. The first three objections have been raised primarily against the Poincaréan/Worrallian version of ESR:

(PS1) ESR commits us only to uninterpreted equations, but these are not by themselves enough to produce predictions, and, therefore, do not deserve all the epistemic credit (1999: 153-4; 2001a: S21).

(PS2) Structural continuity through theory change can be explained better by traditional scientific realism than by ESR (1999: 147-8).

(PS3) Some non-structural theoretical content is retained in theory change, and this is better supported by current evidence and more likely to be true than non-structural theoretical content in the past (1999: 147-8).

One of Psillos' objections seems to apply to all ESR-ists, namely:

(PS4) The structure vs. nature distinction that ESR appeals to cannot be sustained (1999: 157).[80]

Psillos has also put forward three objections that are directed at the Russellian version of ESR. These are:

(PS5) ESR faces a dilemma: On the one hand, the H-W principle by itself can only establish a relation of embeddability between the external world and the ‘world’ of percepts, not a relation of isomorphism as required by ESR. Without a relation of isomorphism, the structural realists cannot establish inferential knowledge about the structure of the external world. On the other hand, the converse of H-W, viz. different stimuli/physical objects imply different percepts, together with H-W allow for the establishment of isomorphic relations but, in doing so, concede too much to idealism (2001a: S13-S16).

(PS6) ESR cannot justify the claim that the first-order properties and relations of unobservables are unknowable in principle (1999: 156); (2001a: S20-21).

(PS7) Knowing the abstract structure of the external world is not enough since it merely amounts to knowing formal properties such as transitivity, symmetry, and reflexivity (2001a: S16-S17).

In what follows, I evaluate these objections by separating them into two groups, one comprising the Worrallian/Poincaréan objections, the other comprising the Russelian objections. Albeit having an equal right to be in both groups, I place PS4 in the first group because Psillos raises this objection in the context of criticising Worrall’s ESR. Finally, since the second group of objections is thematically identical to objections made by Michael Bradie (1977) and Bas van Fraassen (1997), I include these there.

[80] See also Ladyman (1998) and van Fraassen (1999).

2. Terminological Issues

I must first clarify some terminological issues, which, as we shall shortly see, are the sources of some of the above objections. Psillos does not sufficiently explicate the notion of structure, and, as a consequence, draws some mistaken conclusions about the commitments made by ESR.[81] He employs a number of terms, some of which introduced by the structural realists themselves, that loosely refer to what the structural realists have in mind but that are also misleading in their own peculiar way. These are: 'mathematical structure of theories', 'the logico-mathematical structure of theories', 'mathematical content of theories', 'the mathematical form of laws', 'mathematical equations' and 'uninterpreted mathematical equations'.

The first one, viz. 'mathematical structure of theories', may be too narrow. If we take logic as not subsumed under mathematics, then we are leaving out structures specifiable by logic but not by mathematics. For obvious reasons, this problem is remedied by the term 'the logico-mathematical structure of theories'. Both terms, however, as well as the term 'mathematical content of theories' may be too broad in that there is plenty of mathematical machinery which does not play any representative role.[82] Typically, structures taken to represent the physical world are embedded in broader mathematical structures. The excess mathematical structure is obviously not the target of the structural realist's commitments. Hence, to say that a structural realist is interested in the whole mathematical content of theories is misleading at best.

The next term, 'the mathematical form of laws', is also misleading for at least two reasons. One important reason is that the notion of structure should not be restricted to laws. Laws typically express relations between physical entities, properties and relations, but they are not the only theoretical statements that do so. Functions, equations, symmetries, principles, covariance statements, etc., postulate relations between terms that can usually be expressed set-theoretically in the above-mentioned way.[83] Take, for example, the inequality relations of momentum-position, $\Delta p_x \Delta x \geq$

---

[81] See also Redhead's criticism of Psillos (2001b: 345).

[82] See Redhead's (2001a) for an interesting discussion of so-called 'surplus structure', i.e. mathematical structure that has no representative role.

[83] It is worth noting that whether we get to call something 'law', 'principle', or 'equation' is often a historical accident.

h/2, and time-energy, $\Delta E \Delta t \geq h/2$, where $\Delta(z)$ denotes 'spreads' of the value of a measurable quantity *z*, *x* a position co-ordinate, p_x the momentum at *x*, *E* energy, and *t* time. These are relations, and hence specify structures, just as much as Newton's inverse-square law and Boyle's law for gases. For the same reason the last two terms on the above list, 'mathematical equations' and 'uninterpreted mathematical equations', are problematic since they restrict the applicability of the notion of structure to equations. As a matter of fact, since the structural realist's epistemic concerns are with relations, theoretical statements expressed in natural language can also qualify as specifying structures so long as they are expressing relations that have logico-mathematical properties. For example, the statement 'Diamonds are harder than topaz gemstones and topaz gemstones are harder than apatite minerals' is entailed by Moh's scale of hardness, and reflects an ordering of minerals that, among other logico-mathematical properties, exhibits the property of transitivity.

The other reason why the term 'the mathematical form of laws' is misleading is that it is not entirely clear what 'mathematical form' means. Perhaps Psillos is alluding to the idea expressed by the last term on our list, viz. 'uninterpreted mathematical equations', with particular emphasis on 'uninterpreted'. Yet ESR does not subscribe to uninterpreted equations as Psillos suggests. It is to precisely this issue that I now turn to in order to tackle Psillos' first objection.

3. The Objections Against the Poincaréan/Worrallian ESR

Objection PS1

Scientific realists, argues Psillos, deny the ESR claim that "*all* of what is retained is empirical content and (uninterpreted) mathematical equations" (1999: 147) [original emphasis]. The reasoning is that "mathematical equations alone – devoid of their theoretical content – [cannot] give rise to any predictions… [p]redictions require theoretical hypotheses and auxiliary assumptions" (153). Hence, Psillos concludes, uninterpreted mathematical equations cannot be entirely responsible for the success of the scientific theories in which they appear. This claim reflects an objection that Psillos echoes throughout his work.[84]

Is the structural realist arguing that uninterpreted equations are entirely responsible for the success of scientific theories? More specifically, is the structural realist arguing that we should believe only in uninterpreted equations? A careful review of the literature reveals that no structural realist ever supported such a view. Even Worrall, the subject of Psillos' objection, comes close to holding such a view but does not take the plunge. He comes dangerously close, for example, when in arguing that only structure gets preserved through theory change, he asserts that "Fresnel's equations are taken over completely intact into the superseding theory – reappearing there *newly interpreted* but, as mathematical equations, entirely unchanged" (1996: 160) [my emphasis]. If one looks at the context in which this sentence was uttered, as I will soon be doing, one can ascertain that by 'newly interpreted' Worrall is referring to the reinterpretation of these equations under new ontological assumptions. He does not require any other type of reinterpretation of the equations, or that the equations be entirely uninterpreted.

If Psillos is referring to the interpretation that assigns values to the terms of the equation, then he has completely misread the structural realist project. This latter type of interpretation links the terms of an equation – or any other relation for that matter – to our observations, thereby allowing for the production of predictions. This in turn makes verification of the equations possible. Take Worrall's example of Fresnel's equations: (1) $R/I = \tan(\theta_1-\theta_2) / \tan(\theta_1+\theta_2)$, (2) $R'/I' = \sin(\theta_1-\theta_2) / \sin(\theta_1+\theta_2)$, (3) $X/I = (2\sin\theta_2\cdot\cos\theta_1) / (\sin(\theta_1+\theta_2)\cdot\cos(\theta_1-\theta_2))$ and (4) $X'/I' = 2\sin\theta_2\cdot\cos\theta_1 / (\sin(\theta_1+\theta_2)$, where θ_1 and θ_2 are the angles made by the incident and refracted beams with the normal to a plane reflecting surface, and I, R, and X represent the amplitudes of vibration of the incident, reflected, and refracted beams respectively; these are the square roots of the intensities of the components polarised (1) in the plane of incidence I^2, R^2, and X^2, and (2) at right angles to the plane of

[84] For example, earlier in the book he says "it is best not to treat theories as abstract structures, but instead to appeal to the success of *interpreted* scientific theories in order to argue that the kinds posited by them populate the world" (1999: 69). In his (2001a), he repeats: "…in empirical science we should at least seek more than formal structure. Knowing that the world has a certain formal structure… allows no explanation and no prediction of the phenomena" (S21).

incidence I'^2, R'^2, and X'^2. The interpretations of the *angles* and the *intensities* are indispensable to the successful application of the equations. Each of these interpretations assigns a measurable and hence broadly construed *observable*, as opposed to theoretical, property to a term.

Worrall does not question the interpretation of terms θ_1 and θ_2 as angles made by the incident and refracted beams, or of terms I^2, R^2, and X^2 as the intensities of the components polarised. Questioning these would be tantamount to renouncing one of the most spectacularly successful set of equations proposed in the nineteenth century; equations that produced such amazing and unexpected predictions as the bright spot at the centre of the shadow of an opaque disc when placed in the path of light coming from a single slit.

Nowhere in the above discussion of the interpretations to the terms of Fresnel's equations have I mentioned the 'ontological' interpretation that Fresnel attached to them, i.e. that light consists of vibrations transmitted through an all-pervading medium, the ether. But it is only this *ontological* interpretation that Worrall specified as being re-interpretable, not anything else. The ontological interpretation affects only the amplitudes I, R, and X which in Fresnel's framework are understood as vibrations of the ether. One can simply reinterpret these as amplitudes of some sort of vibration/oscillation without any loss of predictive power. That is, Worrall questions what kind of thing is vibrating or oscillating: Is it the ether, the electric and magnetic field strengths, or something else? According to him, we should remain agnostic only with regard to what is doing the vibrating, i.e. only with regard to the 'ontological' interpretation of Fresnel's equations. In other words, we hang on to the idea that *something* is doing the vibrating without being able to specify what that something is beyond the level of isomorphism.

The above example illustrates that the structural realist is very much in need of interpreted equations. Indeed, Psillos' accusation that the structural realist subscribes only to uninterpreted equations rests on a serious misreading of the ESR position. The structural realist does subscribe to interpreted equations, but attempts to distinguish between interpretations that link the terms to observations from those that do not. The hoped-for outcome is interpreted equations that represent relations between measurable, in a broad sense observable, things. In Worrall's version of ESR this information is represented via the Ramsey-sentence, which preserves the interpretations of observables.

Objections PS2 & PS3

Psillos acknowledges the historical claim that structure, in one form or another, often survives scientific revolutions. But he points out that this phenomenon requires an explanation. One explanation, according to him, is the following:

> One might argue simply that retention at the level of equations is merely a pragmatic feature of scientific practice: the scientific community finds it just convenient and labour-saving to build upon the mathematical work of their predecessors. This predilection for mathematical equations, the argument would go on, signifies just the conservativeness of the scientific community rather than anything about the real relations in the world (1999: 152).

Psillos' point is that Worrall needs to show why structural continuity between successive theories is not merely the result of convenient scientific practice, but rather the result of mathematical structures accurately representing the structure of the world.

It is, of course, quite obvious that the same objection can be raised about any type of continuity through theory change. Is it just an accidental, convenient or conservative feature of history or does it reflect a latching onto the world? Thus, it is not just the structural realist that needs to provide a bridge between historical continuity and accurate representation. Psillos, like many others, acknowledges that the defence of realism in any of its forms needs an argument that provides such a bridge. For most, including Worrall and Psillos, the NMA fits the bill. Thus, the bridge can, presumably, be constructed through an appeal to the predictive success of a theory. In other words, the fact that the surviving elements of scientific revolutions are employed in the production of successful predictions makes it unlikely that these elements are not accurately representing the world.

According to Psillos, the structural realist cannot appeal to traditional versions of the NMA, since they attempt to correlate predictive success with more than just the structural features of a theory. Thus, Psillos argues, if NMA is to support ESR, it must be formulated in a way that structures, and *only* structures, get credit for the predictive success of a theory. He offers two reasons why he thinks that this cannot be achieved: 1) uninterpreted equations cannot produce predictions or explain anything so should not take all the credit, and 2) non-structural components are preserved across scientific revolutions and these are better supported by current evidence and more likely to be true than the structural elements.

As the reader might have noticed, 1 and 2 correspond to PS1 and PS3 respectively. Psillos uses them to buttress PS2, i.e. the objection that traditional scientific realism can explain the preservation of structure through theory change better than ESR. Thus, Psillos' case for PS2 hinges on the potency of PS1 and PS3. I have already shown that PS1 lacks potency. To reiterate last section's resolution, the structural realist never championed uninterpreted equations. What about PS3?

This objection is more serious and therefore demands closer scrutiny. Psillos claims that the historical record exhibits continuity beyond the structural level, thereby lending credence to the view that 'non-structural' components of a theory deserve at least some credit for its predictive success. It is only the combination of mathematical equations with theoretical hypotheses and auxiliary assumptions, he holds, that suffices to produce a theory's predictions. According to him, "scientists now have good reason to believe that the [non-structural] content of current theories… is better supported by the evidence, and, hence, more likely to be true" (147).[85] In other words, not only is non-structural theoretical content retained, but it is also better supported, and, therefore, more likely to be true than non-structural theoretical content in the past.[86] Indeed, Psillos maintains that scientific realists "can *explain* the fact that mathematical equations have been retained through theory

[85] Psillos' phrasing equivocates the contrast class. Does he mean that current theoretical content is better supported and more likely to be true than old theoretical content *or* that theoretical content is better supported and more likely to be true than structure? I take him to mean the former since the latter conflicts with his belief that the preservation and success of structure is an important aspect of the realist picture.

[86] Compare this to what he says a page later, viz. that it is "more likely to be true than false" (148). It should be obvious that *A* being more likely to be true than *B* is not the same thing as *A* being more likely to be true than false.

change by saying that they form an integral part of the well-supported and (approximately) true theoretical content of theories" (147) [my emphasis].

The sole, and somewhat detailed, illustration of such alleged non-structural components that Psillos offers, appears in his examination of Fresnel's theory of light, where he lists three assumptions that Fresnel used in the derivation of his laws:[87]

> (a) A *minimal mechanical assumption* that the velocity of the displacement of the molecules of ether is proportional to the amplitude of the light-wave.
> (b) The *principle of conservation of energy* [vis viva] during the propagation of light in the two media. Applying the principle of the conservation of energy to the effective components of light in the interface of the two media, he arrived at a general relation of the form $\sin\theta_2\cdot\cos\theta_1\,(1-R^2) = \sin\theta_1\cdot\cos\theta_2\cdot X^2$.
> (c) A *geometrical analysis* of the configuration of the light-rays in the interface of two media. (158) [abbreviated].

According to Psillos, these are theoretical assumptions that are 'fundamentally correct' but cannot be completely accounted for in structural terms. In any case, they were purportedly carried over from Fresnel's to Maxwell's theory, and therefore deserve at least some of the credit for the predictive success of those theories.

The first thing to note here is that Psillos seems to endorse the view that the success of Fresnel's laws must rub off onto the premises actually employed to derive these laws. Yet, this type of reasoning is fallacious. It is common knowledge that if evidence *e* confirms a hypothesis *H*, it does not follow that it confirms any theory that entails *H*. The point can also be made in the context of deductive logic. A valid argument with a true conclusion need not have a single true premise. The premises may very well be true or at least approximately true, but that is an issue that needs to be confirmed independently. In the current context, (a) is thought to be false because of the reference it makes to the ether.

Setting aside the above fallacy, I find no good reason to view (a), (b), and (c) as 'non-structural'. The first of these, the minimal mechanical assumption, bears all the hallmarks of a structural component since it states a mathematical relation between

[87] What follows is a shortened excerpt that contains all the important points.

two things, viz. the amplitude of the wave and the velocity of the displacement of the molecules of the ether. The fact that the latter of the two is now considered a fictional quantity just means that the relation stated by (a) is not verifiable and can thus be replaced by an electromagnetic facsimile.[88] It is not strange then, to find in Psillos an eventual acknowledgement that the minimal mechanical assumption is not really performing a substantive role in the derivation of Fresnel's laws. He thus says that the only assumption required is to "take energy as a function of the square of the amplitude of the light waves" (159). Indeed, Psillos notes that Fresnel himself had recognised that "no specific assumptions about the trajectories of the ethereal molecules were necessary" (159).[89] The resulting assumption is devoid of reference to the ether, expresses a measurable relation, and, thus, is completely sanctioned by ESR.

It is even more puzzling to see the second assertion, viz. the principle of conservation of vis viva, on Psillos' list of non-structural elements, since this principle expresses a mathematical relation. The principle had been stated in mathematical form as early as the seventeenth century and, despite the appearance of competing accounts, it continued to be treated mathematically until it was replaced by the modern principle of the conservation of energy. One form this principle has taken is the following: $m_1v_1^2 + m_2v_2^2 = m_1(v_1')^2 + m_2(v_2')^2$ where *m* is mass, *v* is velocity, different subscripts indicate different bodies, and the prime indicates the velocity of those bodies after the collision. Notice that all the terms in the equation are measurable and hence broadly construed observable.

Finally, no realist would support the view that geometrical analysis, (c) on Psillos' list, represents any aspect of the world. Geometrical analysis is simply a tool available to the theorist to facilitate modelling and calculation. Its preservation through scientific revolutions, therefore, has no epistemic significance for the realist, structural or other. Even if it had epistemic significance, I do not see how this would

[88] That not all structure gets preserved is a point addressed in chapter five of this dissertation.
[89] Jonathan Bain also makes this point when he says that, what Psillos calls the 'minimal mechanical assumption', "was used solely to express the energy associated with a light-wave as the square of its amplitude with no essential reference to the medium of oscillation. Hence, again, one can argue that the aether was not used in the derivation" (1998: 163).

help Psillos' case since geometrical analysis involves nothing but mathematical structures and, as such, would support ESR, not traditional scientific realism.

All in all, it seems unclear how Psillos' example stands as evidence for continuity at the non-structural level.[90] If anything, his three assumptions appear to be thoroughly structural, and, hence, encourage the correlation between preservation of structure and predictive success. Given these results and the fact that Psillos fails to provide any evidence or arguments to support the claim that the content of current theories is better supported by the evidence and thus more likely to be true than that of older theories, PS3 seems unwarranted. Even if we find clear cases of non-structural component preservation, we must still ask whether such components are essential in the prediction-making and explanatory aspects of theories.[91] As I indicated a few paragraphs ago, Psillos' case for PS2 depends on the potency of PS1 and PS3. Since both PS1 and PS3 seem impotent, I conclude that, pending further evidence and arguments, PS2 remains unsubstantiated.

A more powerful objection to ESR that does not appear on Psillos' list is SRP2. Many authors, including Worrall in his original treatment, have rightly pointed out that the neat preservation of equations found in the Fresnel-Maxwell case is atypical in the history of science (see, for example, Howson (2001), Kitcher (2001), and Redhead (2001a; 2001b)). This is an important objection whose discussion I must postpone until chapter five. There, I will consider, among other things, whether the correspondence principle can assist the structural realist to find more evidence of structural preservation.

Objection PS4

One of the weaker features of Worrall's work on ESR concerns the way in which he contrasts structure to other things. Psillos rightly criticises Worrall for not being clear on "what exactly the distinction he wants to draw is" (1999: 155). While Worrall sometimes talks about the structure of a theory versus its theoretical

[90] Redhead makes a similar observation when he says: "Psillos presents detailed case studies for the examples of caloric and ether but what the discussion boils down to seems to be that structural aspects of the old theory are preserved in the new theory" (2001b: 344).

[91] In chapter five, I argue that there are cases of non-structural preservation, but point out that these do not seem to be essential in the prediction-making aspects of theories, i.e. they do not have any independent confirmation.

interpretation, this being sanctioned by his advocacy of the Ramsey-sentence approach, at other times he talks about the structure of an entity or process versus its nature. Regrettably, he does not explain exactly what he means by nature.

Psillos' critique begins by complaining that the use of the term 'nature' is anachronistic. To talk of 'nature', Psillos says, "over and above [the] structural description (physical and mathematical) of a causal agent is to hark back to medieval discourse of 'forms' and 'substances'… [but such] talk has been overthrown by the scientific revolution of the seventeenth century" (155-6). Without a doubt, the term 'nature' carries too much unwanted baggage with it, having been used in numerous philosophical debates for a variety of reasons. What exactly is meant by it in the present context?

Russell, Poincaré, Maxwell, and Worrall all appeal to the term because of the Kantian undertones of their epistemology. The idea is that we do not have direct access to things-in-themselves, or to 'the nature of things', since direct access is limited to perceptions or phenomena. Unlike Kantian epistemology, knowledge of things-in-themselves can be had under ESR, yet it is indirect, i.e. mediated through perception, and only of a structural kind.

How can we best express this idea of natures? One approach, *implicit* in Worrall's work, is to reduce talk about natures to talk about theoretical interpretations. The aim here is to turn the structure vs. nature distinction into the more familiar structure vs. theoretical interpretation distinction. The latter, as I have already mentioned, is sanctioned by the Ramsey-sentence approach, which strips a theory's theoretical terms of their interpretation and leaves the logical structure and observational interpretation intact. Since Ramsey-sentences make assertions about the properties of theoretical properties, the theoretical properties themselves are presumably the unknowable natures.

This understanding of the notion is similar to Russell's understanding.[92] Russell thinks that we can only have knowledge of the logico-mathematical properties of the properties that things-in-the-world possess, i.e. we can only have knowledge of the abstract structure. Demarcated thus, the nature of things-in-the-world is restricted to that part of physical properties whose description goes beyond isomorphism.[93] In other words, we can know *all* physical properties (of any order) up to isomorphism. That this knowledge does not uniquely specify the physical properties is a trivial point, and will be made clear with examples in the last subsection of section four of this chapter.[94] Nature thus refers to any non-isomorphically specifiable part of physical properties.

What I have just said suggests a widening of the rift between Russell and Grover Maxwell. According to Maxwell, the nature of things-in-themselves is restricted to just their first-order properties. Russell's view, by contrast, takes the nature of things-in-the-world to be restricted to that part of physical properties whose description goes beyond isomorphism. To give an example of their disagreement, take a second order property of a physical object. Whereas for Maxwell this can be knowable, for Russell that part of the second-order property that is not captured by an isomorphic description will be unknowable.

As I pointed out in chapter two, Maxwell is influenced by Russell's idea that the properties of phenomena need not resemble the properties of their external world causes. However, he mistakenly restricts these properties to first-order properties.[95] But, why, we may ask, should second (or higher)-order properties of phenomena necessarily resemble the second (or higher)-order properties of their causes? It is not clear where Maxwell acquired this idea, but it is certainly not a consequence of his accepting the Ramsey-sentence approach. The Ramsey-sentence quantifies over *any* theoretical properties. It thus does not force its advocates to espouse an epistemic distinction between first-order and higher-order theoretical properties. Owing to Maxwell's interpretation of Russell, Psillos takes the distinction to be the central

---

[92] One must not forget that there are also important differences between Russell's ESR and Ramsey-style ESR, notably those discussed in the previous chapter.
[93] Thus 'nature' in this context is not restricted to the essential properties of physical objects, but covers accidental ones too.
[94] In the same section, I will make clear how properties can be known up to isomorphism.
[95] He may have thought that Russell held this view.

tenet of epistemic structural realism, and, as a consequence, needlessly raises objection PS6 to counter it.[96]

Having looked at the principal way in which structural realists understand the structure vs. nature distinction, let us return to Psillos' critique. The main objection that Psillos raises is that "it is doubtful that [the distinction] is well-motivated" because: (P1) "the nature and structure of an entity form a continuum" and (P2) "the nature of an entity, process, or physical mechanism is no less knowable than its structure." (155).[97] Take P1 first. According to Psillos, the nature of a theoretical entity is not distinct from its structure. When scientists talk about the nature of an entity they "talk about the way in which this entity is structured" (155). Indeed, Psillos offers as an example the concept of 'mass', saying that "by discovering more about the properties of mass [including its structural properties] we discover more about its nature" (156). This is just P2, according to which, knowing the structure of an entity means knowing its nature, and so, presumably, the structure of an entity cannot really be clearly distinguished from its nature.

There are various problems with both P1 and P2. Let us consider problems with P2 first. Despite having criticised the term 'nature' as anachronistic, Psillos in the above quotation takes it as signifying all the properties that entities possess. Defined in this way, it is obvious that knowing the (abstract) structure, i.e. the logico-mathematical properties, of an entity just means knowing some properties of that entity, and, hence, something about its nature. Should we decide to understand 'nature' as Psillos does, knowing the (abstract) structure of an entity is knowing something about its nature. Even so, the advocates of ESR can still maintain that the nature of an entity cannot be *completely* known since properties of external world entities, according to them, can only be known up to isomorphism. Psillos does not provide any specific arguments to counter this last claim.

---

[96] That is not to say, of course, that Psillos' objection is not still worth mounting against Maxwell's particular brand of ESR.

[97] Psillos borders on contradiction when, on the one hand, he claims that there is something beyond structure that gets carried over through theory change and, on the other, he argues that the distinction between structure and non-structure cannot be drawn clearly.

Alternatively, we can adopt the Russellian understanding of the term, according to which 'nature' simply refers to that part of properties of external world objects which is left out of an isomorphic description. Better yet, we can baptise some new unloaded term and infuse it with Russell's idea. After all, what is important is what the term denotes. The distinction between structure and non-structure would then express the distinction between the logico-mathematical properties of external world objects on the one hand, and that part of those properties going beyond the logico-mathematical description on the other. In sum, P2 seems groundless and reduces to no more than terminological quibbling.[98]

What about P1, i.e. the claim that structure and nature form a continuum? Consider what Psillos has to say:

> An exhaustive specification of this set of properties and relations leaves nothing left out. Any talk of something else remaining uncaptured when this specification is made is, I think, obscure. I *conclude*, then, that the 'nature' of an entity forms a continuum with its 'structure' (156-157) [my emphasis].

Suppose for this discussion, that by 'nature' we mean what Psillos means, i.e. presumably all the properties possessed by a given entity. First of all, let me reiterate that by 'structure' the structural realist means the logico-mathematical properties of physical objects. This means that there could not be a complete overlap between a set so specified and a set that contains *all* properties concerning an entity. That is, the set of logico-mathematical properties of an entity is a proper subset of the set of all its properties. More to the point, from the view that the properties specified by structure and those specified by nature coincide, it does not follow that they form a continuum. A continuum presupposes two distinct and opposite ends that define an interval between which lies a set that can be brought into a one-one correspondence with the reals. One would assume that what Psillos means by a continuum here is that on one end we find structure and on the other end we find nature. The continuum analogy can be used to express the idea of no privileged dividing line, but it is inconsistent with the idea that the extension of the predicate 'structure of an

[98] Redhead raises a similar point: "Surely part of what we mean by the nature of an entity is the structural property of the relations into which it enters. I don't at all disagree with this point. But this is really a *semantic red herring*. All that the structural realist needs to claim, on my account, is that part, i.e. the structural part, of the nature of the posited physical entities is all that we can claim to know" (2001b: 346) [my emphasis].

entity' is a proper subset of the extension of the predicate 'nature of an entity' or even with the idea that the two predicates have the same extension.

In sum, Psillos is right to criticise Worrall for not being clear on what the structure vs. nature distinction represents. As we have seen, however, the distinction can be drawn quite clearly so long as we define 'nature' as designating the non-isomorphically specifiable part of the external world.

Before we carry on to the next section, I want to consider Worrall's recent suggestion that there is no such distinction between structure and non-structure. More accurately, he suggests that it is not meaningful to speak of non-structural theoretical content since all theoretical assertions are structural.[99] When Poincaré says that 'the ruins of science may be still good for something', Worrall argues, he means that in reality there are no ruins. Here's Poincaré's passage in full:

> The ephemeral nature of scientific theories takes by surprise the man of the world. Their brief period of prosperity ended, he sees them abandoned one after another; he sees ruins piled upon ruins; he predicts that the theories in fashion to-day will in a short time succumb in their turn, and he concludes that they are absolutely in vain. This is what he calls the *bankruptcy of science*.
>
> His scepticism is superficial; he does not take into account the object of scientific theories and the part they play, or he would understand that the ruins may be still good for something ([1905]1952: 160).

Worrall's claim is that, instead of understanding the ruins of science as a metaphor for the non-structural stuff that gets left behind in the wake of a scientific revolution, we should recognize that there are no ruins. How could there be, he asks, if a scientific theory properly construed is just the Ramsey-sentence?

That Poincaré can be read in this way is arguable. I will not dwell on this however. What is even more arguable is the idea that the structure vs. non-structure distinction becomes obsolete if one construes scientific theories as nothing other than their Ramsey-sentence counterparts. It seems to me that so long as scientists posit metaphysical theses about the properties of physical objects, there will be ruins in the aftermath of scientific revolution. Even if scientists are wrong in doing so, it happens

[99] He expressed this opinion in various talks and also in private communication with me.

and that means ruins are generated. Worrall suggests that the ruins are only apparent. Even so, I don't see why this interpretation forces us to abandon the structure vs. non-structure distinction, and, more importantly, how structural realism can survive without the distinction. How else can structural realism be understood as underwriting the pessimistic induction argument, if not for the ruins? Indeed, the very fact that Ramseyfication is a process by which we strip away theoretical interpretations should be sufficient to dissuade Worrall from giving up the structure vs non-structure distinction. After all, the stripping away must be directed at something other than structure.

4. The Objections Against the Russellian ESR

Russell's Principles Revisited

Michael Bradie's critique of Russell's principles serves as a good launch pad into a discussion that will throw more light on the Helmholtz-Weyl principle (H-W) and the Mirroring Relations principle (MR). Let us take H-W first. Bradie protests that the principle is strictly speaking false. This, according to him, is made "clear from the case of colour blind persons or others whose receiving mechanisms (brains) are defective in some way" (1977: 444).

This objection is clearly misdirected at the H-W principle. Colour-blindness gives rise to perceptions of the same colour, where people without the condition perceive different colours. It is, thus, a case of same percepts given different stimulus/physical object. This counts against the converse of the H-W principle, namely that different stimuli/physical objects imply different percepts. Let us call this converse the 'W-H principle' (just for the sake of convenience not because either Helmholtz or Weyl advocated it). As I shall indicate soon, Russell shrugs off the W-H principle because he recognises that we often have different stimuli/physical objects that lead to the same percepts.

Bradie's example fails to counter the H-W principle. But other examples, some of which involve defective perceptual mechanisms, are more compelling. It is, for instance, possible that the same stimulus can give rise to different perceptions in different subjects.[100] Indeed, one would expect this to be the case when the subjects concerned have very different neurophysiologies. A stimulus that produces the sensation of redness in me could produce a completely different sensation in, let's say, a Martian.

Though indeed correct, I do not think this last one is an effective counterexample to H-W. Russell insists that since our access to other minds is indirect, for all we know the same stimulus gives rise to different perceptions in different people. This is vindicated by neuroscience, which holds that brain activity in experiments where the same stimulus is given to different subjects is not identical; at least not in the token-token sense. This result is compatible with H-W, since it states that different percepts *in any given subject* arise because of *that* subject's exposure to different stimuli. The emphasis here is on individuals. H-W does not state that two or more individuals each of whom experiences a different percept will have been exposed to different stimuli, but rather that the same individual who experiences different percepts will have been exposed to different stimuli.

H-W, properly understood, is not easily violated. It is hard to imagine that the same person would form a different perception if exactly the same stimulus was present at two different times. Of course, we can imagine that changes in a person's neurophysiology over time could bring about such a result. This, however, does not seem to be the norm but rather the exception. Most individuals would, when given the same stimulus, identify the same percept most of the time. Were this not to hold then neuroscience in particular, and science as a whole, would be difficult, if not impossible, to pursue. Quine, indicating that we should not even require sameness of perceptions but rather similarity of perceptions, reinforces the argument for the indispensability of H-W, saying that "[p]erceptual similarity is the basis of all expectation, all learning, all habit formation" ([1995]1998: 19). One need only suppose the contrary to see how everyday reliable inferences become unavailable.[101]

---

[100] Remember that H-W can be stated in its contrapositive form, i.e. 'If same stimuli, then same percepts'. When stated like this, it is easier to see that the principle is violated by cases where the same stimuli give rise to different percepts.

[101] If the same stimuli gave rise to different percepts more often than not, then we would have to accept that whenever members of a team synchronize their watches a miracle takes place.

The fact that H-W refers to individuals, instead of groups, does not mean that it has no intersubjective or objective implications. After all, the inference from different percepts is supposed to tell us all something objective about the external world, namely the presence of different stimuli. Perhaps more importantly, H-W's reference to *any given individual* amounts to a universal generalisation, i.e. it holds for all individuals. For those who would like a formulation that is a little less centred on the individual, we could offer the following reformulation: 'Same objective state of affairs, same perception', where by 'objective state of affairs' we include facts about sense organs. This last formulation brings out more clearly that which objectively matters, so that any individual in the same objective state of affairs would have the same sensation. This could be further modified in accordance to Quine's qualification that similarity, not just strict sameness, is in order.

Let us see how H-W works in practice. We already said that even though the same stimulus may induce different percepts in different individuals, the very same stimulus will almost always induce the same percept in the same individual. This spectacular fact allows for agreement between individuals. Even if two or more individuals have different perceptual experiences in the presence of the same stimulus, language allows them to express their agreement that each is experiencing the same thing. For example, if, whenever stimulus *A* is present, individual *X* experiences *B* and individual *Y* experiences *C*, where B ≠ C, then even if *X* and *Y* call their experiences by different names, their consistent use of those names in the presence of *A* allows for the discovery that they are talking about the same thing. Figure two below illustrates this point.

Although individuals *X* and *Y* do not experience the same percept, but *B* and *C* respectively, and they give their own name, *b* and *c* respectively, to stimulus *A*, each experiences the same percept at various times, t_1, …, t_n, i.e. always percept *B* for *X* and percept *C* for *Y*. Despite the fact that percepts, like *B* and *C*, are private, each individual's consistent identification of stimulus *A* as the cause of their percept is sufficient ground for agreement at time t_n that *b* and *c* name the same stimulus, which they decide to call 'd'. Quine makes an almost identical point when he says: "…if two individuals jointly witness one scene, and subsequently jointly witness another scene, and the one witness's global stimulations on the two occasions qualify for him as perceptually similar, usually the other witness's stimulations will so qualify for the other witness " (20).

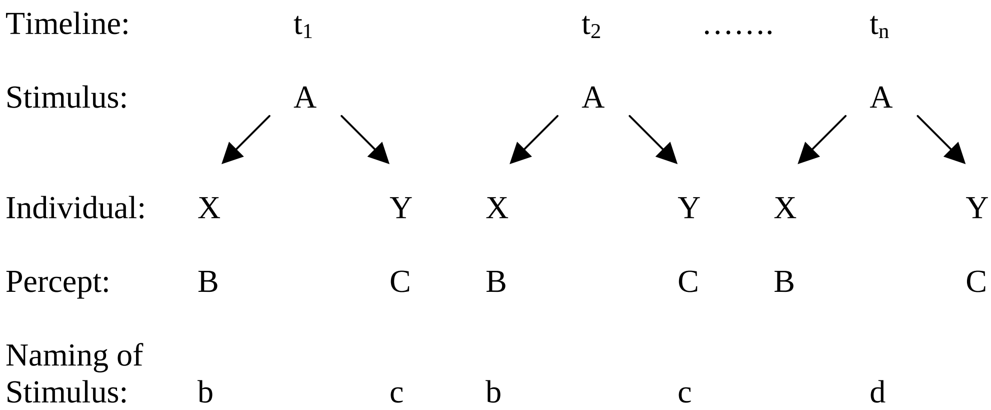

Figure 2: Agreement between individuals in light of different percepts.

A slight modification of the above conditions, administering two different stimuli that lead to two different perceptions for each individual, allows us to extend this point. In accordance with H-W, each individual would reason that given the two different percepts there are two different stimuli. Each individual's consistent identification of each stimulus with a particular percept would eventually lead to the discovery that both individuals agree that they are talking about the same two different things.

Let us now turn to MR – the principle that relations between percepts are isomorphic to relations between their non-perceptual causes. A sensible question to ask is whether Russell argues for it. Here's one argument I have uncovered:

> Let us confine ourselves to the angular co-ordinates [as an example]. My point is that the relations which physics assumes in assigning angular co-ordinates are not identical with those which we perceive in the visual field, but merely correspond with them in a manner that preserves their logical (mathematical properties). This *follows* from the assumption that any difference between two simultaneous percepts implies a *correlative* difference in their stimuli (1927:252) [my emphasis].[102]

[102] Russell's focus on the correspondence between angular co-ordinates and the visual field simply reflects the fact that he is giving a concrete example of what he thinks holds in general. Also, the reference to simultaneity does not appear to be a necessary ingredient of H-W.

Thus construed the argument is valid but only because Russell sneaks the term ‘correlative’ into the formulation of H-W found in this passage and thereby guarantees the preservation of the logico-mathematical properties of relations. But this is obviously question begging since he assumes what he wants to derive. The argument properly construed is invalid. H-W does not by itself entail that relations between percepts have the same logico-mathematical properties as (i.e. are isomorphic to) relations between stimuli. In short, it is possible that H-W is true and MR false.

Perhaps the argument is an enthymeme. Earlier in the book, Russell made the remark that “This principle [i.e. H-W] together with spatio-temporal continuity, suffices to give a great deal of knowledge as to the *structure* of stimuli” (227). Russell’s principle of spatiotemporal continuity asserts, roughly speaking, that non-perceptual causes and their perceptual effects are spatiotemporally continuous, i.e. there is a continuous series of events that links the cause to the effect. Setting aside any difficulties facing the principle itself, it is still not all that clear why spatiotemporal continuity can help with the task at hand.[103] Maybe Russell’s idea is that any spatiotemporal relations between percepts mirror spatiotemporal relations between their causes in the physical world. Even if true, this postulate would not license the inference from non-spatiotemporal relations between percepts to non-spatiotemporal relations between their non-perceptual causes. One could, of course, reduce all physical relations to spatiotemporal ones, in which case the only relations between percepts that matter for our purposes are spatiotemporal. Without speculating further, pending a clear account of how the principle of spatiotemporal continuity is meant to support MR we must remain sceptical.

Perhaps the implicit premise we were looking for in Russell’s argument was not the right one. Another candidate that might turn the enthymeme into an explicit argument can be found in this passage:

> …perception as a source of knowledge concerning physical objects would be impossible if there were not, in the physical world, semi-independent causal chains,

[103] Among the difficulties facing the principle is one pointed out by Bradie, namely that Russell never explains how spatio-temporal continuity works between the space of percepts and the space of physics (1977: 444-5).

> or causal lines as we may call them. The light which comes to us from a printed page contains retains the structure of the page; if it did not, reading would be impossible (314).

The idea here is that structure is preserved via the assumption that the causal chains leading from a physical object to the formation of our perception of it are sufficiently independent of causal interference. This allows us to get a clear picture of the cause of our perception. Russell emphasises the overall reliability of visual perception, saying that "light-waves travel with extraordinarily little modification through empty space, and without very great modification through a clear atmosphere" (164).[104]

This is certainly an important principle, without which the preservation of any information about the external world would be impossible. But it does not really establish that the relations between percepts are isomorphic to the relations between their external world causes. Otherwise put, the principle of semi-independent causal chains is a necessary condition for the preservation of structure or relations, but certainly not a sufficient one.

Bradie sums up the nature of the difficulty well:

> But what gives us the right to infer structure? We get the right by assuming certain postulates to be true. But, then why should we not assume certain other postulates to be true… This is an extremely important question which needs to be thoroughly examined. Without a clear reason for preferring structural properties to qualitative properties, Russell's epistemological position seems highly arbitrary (1977: 450).

Why, as Bradie indicates, should we accept MR and not some other principle(s) that preserve(s) 'qualitative' properties over structural ones?

But surely, it is hard to imagine how we can ever have knowledge of the external world without MR. Epistemological realism requires belief in the correspondence between language and reality, i.e. belief in semantic realism. As many have argued, the only type of correspondence that says anything coherent about the world is one

---

[104] Russell goes on to speak of the 'causal purity' of light-waves and concludes that "[t]his is the physical merit of sight as a source of knowledge concerning the external world" (164).

that says something about the relations objects in the world stand in.[105] That is, it is assumed that the correspondence between our theories/sentences/mental entities/percepts and the external world reveals the latter's relations. To be sure, traditional scientific realists like Psillos criticise ESR for offering too little, claiming that we do not only have knowledge of structure but also of something more.[106] A more pressing question to the above then is whether there is licence to infer anything more than structural properties? Since the scope of this question coincides with that of objection PS6, I defer its discussion for later.

Percepts, Phenomena and Observation Sentences

Before resuming my evaluation of Psillos' objections against Russellian ESR, there is one more issue that needs brief discussion. There is a deep worry that commitment to percepts, and other such items of direct acquaintance, entails commitment to 'the given'. Among others, Wilfrid Sellars famously attacked, and to the eyes of many defeated, the myth of the given, i.e. the idea that experience is given without prior conceptualisation. Is there reason for concern that percepts reincarnate the given?

No such reason exists. As is well known, Russell abandoned the idea of sense-data, the allegedly pure objects of our perception, by the time he wrote the *Analysis of Mind*, replacing them with percepts.[107] Unlike sense-data, percepts are conceived of as impure, yet, nonetheless, "the only part of [the world] that we can know without the help of rather elaborate and difficult inferences" (1927: 264). Russell's percepts can, thus, be thought of, not as entirely un-conceptualised, but as un-conceptualised a mental entity as we can possibly get.

We need not even appeal to entities that sound as elusive as percepts. In my discussion of Grover Maxwell's ESR in chapter two, I pointed out that he avoids the reification and mystification of perceptual units, trading talk of percepts for talk of

---

[105] In the context of theories of truth, Marian David says: "A correspondence theory is usually expected to… tell us about the workings of the correspondence relation, about the nature of facts, and about the conditions that determine which truthbearers correspond to which facts. It is natural to tackle this by construing correspondence as an *isomorphism* between truthbearers and facts" (2002) [original emphasis].

[106] In so far as traditional realists accept that we have knowledge of the world's structure via a correspondence principle they should not find MR unreasonable.

[107] See ch.2, §3 for more on this transition.

observation sentences and predicates.[108] This shift is meant to focus on less objectionable units, like observation sentences and predicates, yet leave Russellian ESR essentially unaffected. All that is required is some sort of correlation between the epistemically relevant aspects of perception and observation statements/predicates. It seems clear to me that this correlation comes naturally. Consider my discussion of the intersubjective/objective implications of H-W once more. The idea was that individuals employing H-W could reach the same conclusions in the presence of the same stimulus despite experiencing different percepts. How would they come to such an agreement? As I argued earlier, *language* allows distinct individuals to express their agreement that each is experiencing the same thing. The particular language involved is observational. The epistemically relevant aspect of perceptions is thus transmissible through observation sentences and predicates.[109] In sum, even though percepts are 'mental' entities and observation sentences are linguistic ones, the shift of focus from one set of entities to the other does not compromise the main tenets of Russellian ESR, since nothing epistemically relevant is lost.

One further issue concerns the relationship that phenomena have to percepts. It is not exactly clear what the term 'phenomenon' denotes. Sometimes phenomena are thought of as entities existing solely within our mind, much like percepts – if not identical to them. This reading is suggested by the etymology of the term, for in its original form it literally means 'appearance'. A different reading takes phenomena to be intermediaries between the physical objects and our mind, existing independently of the latter. In spite of the disagreement between these two readings, it is generally agreed that phenomena have an immediate effect on the mind.[110] It is this *immediateness*, and the epistemic privilege that it purportedly confers, that is so crucial for the empirically motivated Russellian brand of ESR. Thus, whether we talk about phenomena or percepts does not matter much in the given context.[111]

---

[108] Of course, there is no need to deny the existence of perceptual entities; there is only a need to avoid making any substantive metaphysical assumptions about their nature.
[109] Van Fraassen believes that observation sentences can capture the epistemically relevant content of phenomena.
[110] For a notable exception see Bogen and Woodward (1988). According to them, *data* can be observed, whereas *phenomena* "in most cases are not observable in any interesting sense of that term" (306).
[111] Should one insist on distinguishing between percepts and phenomena, it would still be plausible to maintain that the relationship between physical objects, phenomena, and percepts takes the following

Henceforth, and unless otherwise noted, I will employ the terms ‘percept’, ‘observable’, ‘phenomenon’ and any of their derivatives more or less interchangeably, not because I assume that they necessarily amount to the same thing, but because, as I have just argued, such usage does not compromise the spirit of Russellian ESR.

Objection PS5

The First Horn of the Dilemma

According to Psillos, the H-W principle can only establish a relation of embeddability between the external world and the world of percepts, falling short of a relation of isomorphism required by ESR. Without a relation of isomorphism, Psillos argues, structural realists cannot establish inferential knowledge about the external world. But in what way exactly is the H-W principle able to establish a relation of embeddability but not one of isomorphism?

Let us identify any set of percepts by the letter P and any set of external world causes, i.e. stimuli/physical objects, by the letter C. Psillos argues that the H-W principle cannot give us isomorphic mappings between P and C. To remind the reader, the H-W principle expresses the following conditional: If different percepts, then different stimuli/physical objects. This principle guarantees that a given C will have at least as many members as the corresponding P. More importantly, the principle is equivalent to saying that there is an injective mapping f: $P \rightarrow C$, such that if x and y are distinct members of P, then their images in C, fx and fy, are also distinct. Now, for a mapping to be isomorphic it is not enough to be injective, it must also be surjective, i.e. for every $c \in C$, there is at least one $p \in P$, such that $f(p)=c$. To establish this, we need the converse of H-W, viz. W-H: if different stimuli/physical objects, then different percepts.[112] H-W and W-H together guarantee that there is a bijective mapping between the set of percepts and the set of

form: At one end lie the physical objects and at the other the observation sentences. In between lie the percepts and the phenomena, the former causally originating from the latter which, in turn, causally originate from physical objects.

[112] Notice that a *purely* surjective mapping leaves open the possibility that two or more percepts may correspond to one and the same cause.

stimuli/physical objects. A relation of isomorphism can be established when this bijective mapping also preserves all the relations found in the structured domains.[113]

Psillos correctly points out that the H-W principle cannot by itself give us the much-desired relation of isomorphism. He is wrong, however, in arguing that the H-W principle allows us to establish embeddability relations, unless he is using the term ‘embedding’ to mean injective mappings. The term ‘embedding’ is more often reserved for injections that also map relations, hence appeal to this notion implies one is already dealing with a structured domain. As we have just seen, H-W only takes us as far as plain injective mappings, i.e. it does not map any relations – something required by embeddings. More formally, an embedding of structure S_1 (A, R) into structure S_2 (B, R´) is a one-one mapping *f* of A into B such that: (1) $f(a_j)=b_j$ for all $a_j \in A$ and (2) $<a_1, \ldots, a_n> \in R_i$ iff $<f(a_1), \ldots, f(a_n)> \in R'_i$ for all $i \in I$ and all $a_1, \ldots, a_n \in A$. It is worth noting that embeddings are isomorphic mappings of a particular kind. In general, we can say that a structure S_1 is embedded in a structure S_2 *if and only if* S_1 is isomorphic to a substructure of S_2.

More crucially, Psillos is wrong in assuming that ESR requires a commitment to isomorphic relations only. Russell acknowledges that the W-H principle is unreliable because we often have different stimuli that lead to the same percepts. This can be easily illustrated in cases involving distance, as Russell’s example shows: “If we are observing a man half a mile away, his appearance is not changed if he frowns, whereas it is changed for a man observing him from a distance of three feet” (1927: 255). Because of such examples, Russell suggests that “differences in the percept imply differences in the object, but not vice-versa” (339-340). Moreover, he recognises the limitations of the inferential powers of the H-W principle, when it is not accompanied by W-H. Paradoxically, Psillos takes note of this when he says that “[p]recisely because Russell doesn’t have the converse principle, he talks of ‘roughly one-one relation’ ” (2001a: S15). In other words, Psillos acknowledges that Russell was never committed to isomorphic relations only.[114]

[113] This is an important detail that Psillos fails to mention in his discussion.
[114] The above illustrates why Russell refrains from saying that we can know the structure of the physical world and instead holds that we can ‘infer a great deal’ about it.

The more serious objection Psillos puts forward is that it is meaningless to speak of *roughly* one-one relations. Yet, even without the help of W-H, MR is a strong enough principle to guarantee inference at the isomorphic level. This can be seen in the following way: Injective mappings can easily be given inverse mappings, i.e. for *any* injective mapping f: $D \rightarrow E$ we can give an inverse mapping f^{-1}: $E' \rightarrow D$ where E' = ran f. That is, E' contains as its members all and only those objects that are contained in the range of $f =_{\mathrm{df}} \{fx: x \in \mathrm{dom}\ f\}$. Notice that by doing so we immediately satisfy the requirement of a surjective mapping, since for every object in E' there is at least one – in this case only one – corresponding object in D. In short, we get a bijective mapping between D and E' – where E' may or may not equal E.[115] The MR principle, i.e. that relations between percepts have the same logico-mathematical properties as relations between their external world causes, allows us to turn a bijective mapping into an isomorphic one, for it allows us to preserve any relations the set of external world causes may have.

Psillos complains that "[f]rom a realist viewpoint, it should at least in principle be possible that the (unobservable) world has 'extra structure', i.e., structure not necessarily manifested in the structure of the phenomena" (S15).[116] If there is such extra structure, he continues, the required relation between the world of percepts and the external world should be that of embeddability not isomorphism. Yet, Psillos argues, ESR cannot be upheld by appeal to embeddability since under this relation "the structure of the percepts doesn't determine the *domain* of the stimuli" (S16) [original emphasis].[117]

Let us first reflect on the idea that the unobservable world could have some extra structure that is not manifested in the structure of the phenomena. There seems to be no reason why ESR should be inconsistent with this idea. ESR simply says that structures of phenomena mirror the structures of the unobservable world. It requires that every phenomenal structure has a corresponding unobservable structure. It does

[115] E′ will be different from E only if the cardinality of E is greater than the cardinality of D.
[116] Notice that Psillos uses the terms 'percept' and 'phenomenon' interchangeably. Though it is good practice to keep the two apart, I follow Psillos in using them interchangeably since they do not compromise the spirit of Russellian ESR.
[117] This quote appears in the midst of Psillos' discussion of the second horn of the dilemma but can be mustered here since it is an objection to the view that ESR sanctions embeddings.

not require the converse, i.e. that every unobservable structure has a corresponding phenomenal structure. In other words, ESR is compatible with the idea that not every unobservable structure will have a corresponding phenomenal structure, i.e. that the unobservable world may have extra structure. The real objection to ESR in this context would be to show that at least some structures of phenomena do not mirror the structures of the unobservable world.

What exactly does Psillos mean when he says that embeddings do not support ESR, since under embeddings the structure of the percepts does not determine the domain of the stimuli/physical objects? One way of understanding this claim is to take the absolute determination of the domain of the stimuli/physical objects as the complete description of the domain's objects. If this is the case, his argument clearly falters on account of the fact that the relation of isomorphism does not require such a determination either. Indeed, one of the central points of ESR is that the stimuli/physical objects along with their properties and relations cannot be fixed absolutely, but only *up to* isomorphism.[118] In short, this sort of underdetermination is not only compatible with ESR but constitutive of it. The only other plausible reading of Psillos' claim is that a relation of isomorphism requires that the sets mapped have the same cardinality, whereas embeddability allows one to infer the minimum size of the set from which the range of the mapping is drawn. This difference does not amount to anything significant because there is no clause in ESR that requires the exact determination of the cardinality of a given set.

As Psillos admits, Russell's epistemic commitments are restricted to embeddings. These, as I have indicated, still offer isomorphic mappings, albeit of a special kind, namely that the structure of perceptions is isomorphic to a substructure of the external world. This still allows inferential knowledge from the structure of perceptions to the structure of the external world. Indeed, that the whole of the unobservable world might have extra structure, i.e. structure not reflected in the structure of the phenomena, is a possibility that is consistent with ESR. Thus Psillos' first horn of the dilemma crumbles.

---

[118] This type of underdetermination is independently backed up by a host of different arguments. One such argument is the ontological relativity argument propounded by Quine (1969). Demopoulos and Friedman (1985: 628), in particular, make a compelling case for this.

The Second Horn of the Dilemma

It is not entirely clear what Psillos means when he says that H-W and W-H allow inferences at the level of isomorphism but concede too much to idealism. In support of this claim he quotes certain passages from Hermann Weyl, where it seems that Weyl takes W-H to be "the central thought of idealism" and asserts that "science concedes to idealism that its objective reality is not given but to be constructed" (1963: 117). On the basis of this quote, Psillos complains that it should not be a priori false for a realist that there is a divergence between the structure of the world and the structure of the world of percepts. According to Psillos, "[f]or all we know, the unobservable world may differ from the world of phenomena not just in its 'intrinsic nature', but in its structure too." (2001a: S16).

At least two things lend themselves to scrutiny here. First, Weyl seems to be too quick to judge W-H the central thought of idealism. Whether one is an idealist depends on what one takes the stimuli (referred to in the W-H principle) to be. Nothing prevents one from taking stimuli to be objects in the mind-independent physical world, or, as Russell does, as physical events. Moreover, there is no inconsistency in holding the belief that there is a mind-independent physical world and the belief that this world can only be known up to isomorphism. The point needs no belabouring. W-H does not imply belief in idealism. Even if it did, ESR does not employ W-H.

Second, and more importantly, when Psillos argues that it should not be a priori false that there is a variance between the structure of the external world and the structure of perceptions, he mistakenly implies that this is the ESR-ist view. I do not see any good reasons why this should be the case. Nobody would deny that perceptual apparatus can *sometimes* malfunction. This means that the structure of the external world need not always be correctly reflected in the structure of our perceptions. A similar qualification should be made with regard to H-W. As we saw earlier, *most* individuals would, when given the same stimulus, identify the same percept *most* of the time. Hence, ESR-ists can, and should, accept the view that some variance between the structure of the external world and the structure of perceptions exists. This qualification does not fundamentally undermine their programme, for the overall reliability of inferential knowledge about the structure of the external world is safeguarded.

Objection PS6

The claim that we can know only the structure of the world, charges Psillos, is ambiguous. It may mean one of three things: (a) that everything is knowable apart from the individual objects, or (b) that everything is knowable apart from the individual objects and their first-order properties, or (c) that everything is knowable apart from the individual objects, their first-order properties and their relations. Each of these, Psillos claims, creates a different version of epistemic structural realism. But which one do we choose? In other words, where exactly do we draw the line between what is knowable and what is not? Psillos thinks that option (c) "is the only characterisation of ESR which can impose a principled limitation on what is knowable" (S21). But (c), according to him, is questionable since it commits us to the idea that some properties are unknowable *in principle*. He says:

> …it isn't clear why the first-order properties of unobservable entities are unknowable. They are, after all, *part and parcel of their causal role*. So, if all these entities are individuated and become known via their causal role, *there is no reason to think that their first-order properties, though contributing to causal role, are unknowable* (S17) [my emphasis].[119]

It is thus implied that traditional varieties of scientific realism, of which Psillos is an advocate, are more reasonable than ESR because they do not preclude first-order properties from being knowable in principle.

Let us, first of all, make a clarification. Although (c) comes close to a faithful characterisation of ESR, it misrepresents the position in one important respect. ESR does not hold that we have absolutely no knowledge of the first-order properties of external world objects. This is a mistake that, as we saw earlier, originates in Grover Maxwell's misconstruction of Russell's account. ESR holds that *all* properties of external world objects are knowable up to isomorphism. More precisely, ESR really means (c′): Everything in the external world, i.e. objects, properties, and relations, is knowable up to isomorphism.

[119] See also (1999: 156).

Since I presented isomorphism as a relation that holds between relations or structures, I must explain what is meant by the claim that *objects* and *properties* can be known isomorphically. Structures specify objects, relations, and, potentially, one-place properties.[120] We take abstract structures to represent a certain isomorphism class of concrete structures, i.e. to represent the concrete structures isomorphically. Given this character of abstract structures, the contents of their domains of objects and any one-place properties (understood as sets) cannot be uniquely specified. Only their cardinalities and the (logico-mathematical properties of the) relations they stand in can be specified. For example, we can say that property P has three objects and property Q has two objects and that a certain relation R with formal properties X, Y, and Z holds between objects in P and Q.[121] Thus, to say that we know objects or properties isomorphically just means that we know them to the extent that they are specified by abstract structures.

Despite Psillos' misconception of Russellian ESR, his objection can be reformulated thus: Why should properties of the external world be epistemically inaccessible beyond the level of isomorphism? This question is equivalent to the question I promised to return to in the section on Russell's principles, namely: Is there licence to infer anything more than structural properties?

A satisfactory answer to this question can be given and finds some support in science. Optical science, for instance, tells me that when I see a coloured object it is the result of incident light waves of a given wavelength hitting my retina and producing nerve impulses that travel all the way to my brain where the relevant perception is formed. It thus tells me that colour gives us some information about the external world, namely that incident light waves of a given wavelength are hitting my retina.[122] If I see two otherwise *perceptually identical* objects, one of which is red and the other green, then, *ceteris paribus*, I postulate that there must be some

[120] Recall that in the beginning of chapter two it was indicated that a structure may also specify one-place properties and not just relations, i.e. not just 1+*n*-place properties, where *n* is a positive integer.
[121] An additional, crucial, component of ESR knowledge claims that is not contained in the isomorphism claim and is therefore sidelined here is the idea that the physical system exemplifying an abstract structure S* can be indirectly identified as that system which is causally responsible for the concrete observational structure that led us to infer S*.
[122] I am aware of the voluminous philosophical literature on colours. What I say here bodes well with eliminativist theories of colour, according to which physical objects have no colour. For more on this and other theories see David Hilbert (1998).

difference in the two objects responsible for the difference I identify in perception.[123] In optics, this difference arises from the different properties of the surface of the two objects, which determine the wavelength composition of the light reflected from them. This is just the H-W principle in use. I infer that there must be a non-empty set of properties that one object has while the other does not. (NB: If I am colour blind I may not be able to tell the difference but that just means that I will not pick up on this relation. The H-W principle holds that provided we identify a difference in perception, we should postulate that this corresponds to a difference in the world. It *does not* guarantee that we will identify a difference. It is the W-H principle that requires that there be a corresponding difference in perception provided there is a difference in the world).

On the basis of Russell's programme of epistemological reconstruction, although I can infer that there is a non-empty set of properties that one object has while the other does not, I cannot infer exactly what these properties are. To gain more information about these properties and their objects I must make more observations. In particular, I must find out whether any relations hold at the perceptual level. Placing the two objects under the microscope, for example, would presumably reveal such relations. Supposing MR to hold, I infer that the perceptual relations revealed under the microscope reflect relations between the constituent parts of the objects. So, at best, I know certain relations between these constituent parts of the objects, but I do not know the constituent parts themselves. But knowing a relation without knowing the relata just amounts to knowing the logico-mathematical properties of the relation. This is equivalent to saying that we know these relations, and the structure they specify, up to isomorphism. More pertinently, it means that the properties of the relata can only be specified up to isomorphism.

Psillos insists that first-order properties are 'part and parcel of their causal role' and thus must be knowable. ESR does not deny that first-order properties are an essential or integral component of the causal chains that lead up to our perceptions. But it is

[123] One potential worry here might be that colour, as well as other such properties, do not correspond to things or the structure of things in the world, but rather are products of our neurological apparatus and the external world. Such an objection would miss the point however since, as I have detailed in chapter two, the external world, according to Russell and Maxwell, encompasses all that is non-perceptual which includes the neurological. After all, is not our nervous system composed of physical entities?

one thing to argue for this, and quite another to claim that it shows we have epistemic access to the first-order properties, or indeed higher-order properties, of physical objects beyond the isomorphic level. Pending a more detailed argument explaining why this is the case, I do not see any force to this dimension of Psillos' argument.

Objection PS7

Van Fraassen has recently complained that ESR is too weak a position to be realist, for it leads to the view that "[s]cience is now interpreted as saying that the entities stand in relations which are transitive, reflexive, etc. but as giving no further clue as to what those relations are." (1997: 516). Similarly, Psillos thinks that the position's commitment to abstract structure makes it too weak to be a genuine realist position. In this vein, Psillos echoes Van Fraassen's concern:

> [Grover Maxwell] thought that formal properties, such as transitivity, are purely structural, but added that 'not all structural properties are purely formal'… Yet, he leaves us in the dark as to what these non-formal structural properties which are referred to by theories are (2001a: S16-17).

The implication here is that Maxwell's recognition of the insufficiency of formal properties to capture all that is of epistemic worth prompts him to view them as a proper subset of structural properties. Hence, both Psillos and van Fraassen object that knowing the abstract structure of the external world is not enough since it merely amounts to knowing its formal properties. In what follows, I reply to this objection by arguing that the characterisations of structure that structural realists adhere to are broader than some of their opponents say or imply.

Hereafter and until otherwise noted, let us call 'formal properties' of relations *only* the properties of reflexivity, irreflexivity, transitivity, symmetry, and antisymmetry. Neither Van Fraassen nor Psillos specify what exactly belongs to this collection of properties but give as examples those properties just mentioned. This vagueness is part of the problem, as we shall shortly see.

First of all, it should be acknowledged that formal properties are unable to uniquely determine specific relations. An example will help illustrate the point. Take the following two relations $R_1 = \{<1,2>, <2,1>, <1,1>\}$ and $R_2 = \{<1,2>, <2,1>, <2,2>\}$ defined on set $U = \{1,2,3\}$. Both relations share the same formal properties, in the sense specified above, i.e. they are not reflexive, not irreflexive, not transitive, symmetric, and not antisymmetric. Suppose, for the moment, that R_1 is the target relation. Knowing these formal properties plus the content of U does not allow one to infer which of the two relations is the target. In fact, the underdetermination is even deeper since there are a number of other relations defined over the same set that have the same formal properties. More specifically, there are seven other relations, viz. $R_3 = \{<1,2>, <2,1>, <3,3>\}$, $R_4 = \{<1,3>, <3,1>, <1,1>\}$, $R_5 = \{<1,3>, <3,1>, <2,2>\}$, $R_6 = \{<1,3>, <3,1>, <3,3>\}$, $R_7 = \{<2,3>, <3,2>, <1,1>\}$, $R_8 = \{<2,3>, <3,2>, <2,2>\}$, and $R_9 = \{<2,3>, <3,2>, <3,3>\}$.

This underdetermination resurfaces at a more abstract level, i.e. the class of isomorphic relations. This is the level at which the Russellian structural realist claims that our knowledge becomes delimited. First of all, we abstract[124] from our domain U to domain U' where the only information we end up having about U' is its cardinality, namely that it has three objects. We can give arbitrary labels to these objects. Let us call them '*a*', '*b*', and '*c*'. We can then proceed to abstract from each concrete relation: i.e. from R_1 to $R_1' = \{<a,b>, <b,a>,<a,a>\}$, from R_2 to $R_2' = \{<a,b>, <b,a>,<b,b>\}$, from R_3 to $R_3' = \{<a,b>, <b,a>, <c,c>\}$, etc.[125] Notice that the formal properties are of course preserved despite the process of abstraction. Hence, the same underdetermination also holds here for the simple reason that the new relations share the same formal properties.

Having established that knowing the formal properties is not sufficient to determine the target relation, at the concrete or abstract level, we can now proceed to show that this underdetermination is more severe than the one faced by ESR. According to ESR, relations can be known up to isomorphism. Suppose, as before, that the target relation is R_1. Knowing R_1 then means not being able to describe it beyond isomorphism. Five other relations are isomorphic to R_1, i.e. R_2, R_4, R_6, R_8, and R_9. This eliminates three out of the nine relations, i.e. R_3, R_5 and R_7, when compared to

[124] The process of abstraction referred to here is the same as that explained in ch. 2, §3.
[125] It is important to note that nothing dictates between $R_1' = \{<a,b>, <b,a>,<a,a>\}$ and $R_1' = \{<a,b>, <b,a>,<b,b>\}$. In this context, the two are equivalent. My choice is thus inconsequential.

the underdetermination that ensues from knowing just the formal properties.[126] Hence, the underdetermination has been considerably reduced. The result is even better if the target relation is R_3. In that case, only two other relations are isomorphic to it, i.e. R_5 and R_7. This eliminates the other six relations, reducing the underdetermination even more dramatically. The example clearly illustrates that knowing the formal properties of a relation, in the sense we have unpacked the notion of 'formal properties' above, is often times not the same thing as knowing the abstract structure/isomorphism class to which this relation belongs.[127] The latter can be much more informative than the former.

At first glance then, an isomorphism seems to be able to preserve more than just formal properties. But there is a trick here. We have defined the term 'formal properties' in a narrow way. The reason for doing so lies in the fact that Van Fraassen's and Psillos' objection draws precisely on such a narrow meaning, or at least on the ambiguity that looms. In reality, we call 'formal properties' any properties that can be described in the language of mathematics. But whatever can be described in the language of mathematics will be described only up to isomorphism. As Michael Resnik aptly remarks "no mathematical theory can do more than determine its objects up to isomorphism" (1981: 529). This is a view not just typical of structuralists in the philosophy of mathematics but also of the wider mathematical community. An isomorphic mapping preserves any formal, i.e. logico-mathematical, properties we can think of. Hence, all that can be captured by an isomorphism is a list of formal properties, where the latter concept is now properly understood in a broad way. This is probably what Russell had in mind when he made such comments as "[l]ogical properties include all those which can be expressed in mathematical terms" and "structure is what can be expressed by mathematical logic, which includes mathematics" (1927: 251, 254).

Since science employs mathematical objects as surrogates for its own objects, it is reasonable to argue that scientific objects can themselves only be described up to isomorphism. I indicated in the previous chapter that van Fraassen pushes this

[126] R_1', for example, cannot be instantiated by R_3, since the same constant cannot name two different things. That is, for R_3 to instantiate R_1', one of *a* or *b* would have to name either 1 and 3 or 2 and 3.
[127] Sometimes the severity of the underdetermination is the same in both cases. Consider, for example, intended relation $R_I = \{<1,1>\}$ defined on set $U_I = \{1, 2\}$.

argument with regard to phenomenal objects, which, in his opinion, are the only objects science should study. Considering that van Fraaseen adopts this view, it is ironic that he then turns around to criticise the ESR-ists for not being able to describe relations beyond isomorphism. If anything, on this issue, the two positions seem to stand or fall together.

One may complain that we still end up with underdetermination on our hands. The appropriate response to this complaint is that ESR delimits underdetermination by setting constraints on the type of objects, properties, and relations that are suitable candidates for the role of correctly representing the ontology of the world, namely those that belong to the same isomorphism class. This limit must be appreciated for it leaves realism with some breathing space. It allows us to get some, admittedly loose, grip on reality.

5. Conclusion

In reply to Psillos' objections we can now give the following answers:

Reply to PS1: ESR does not involve a commitment to uninterpreted equations. Worrall's version, in particular, involves a commitment to structures (including equations), whose observation terms are fully interpreted and whose theoretical terms are presumably implicitly defined through their logical relations with one another and with the observation terms. This just amounts to the Ramsey-sentence approach to theories. Contrary to Psillos' objection, such structures have the capacity to produce observable predictions. On the basis of this objection, the claim cannot be made that structures do not deserve all the epistemic credit.

Reply to PS2 and PS3: Psillos employs PS1 and PS3 to support PS2. PS1, as I have just reiterated, is unwarranted. The examples of preserved non-structural components that Psillos gives in support of PS3 are either devoid of epistemic significance or unmistakably structural. The further fact that Psillos does not back up his claim that the content of current theories is better supported by the evidence and thus more likely to be true than that of older theories, leaves PS3 unwarranted. Thus, PS2 remains unsubstantiated.

Reply to PS4: Though the distinction between structure and nature is unclear in Worrall's work, the structural realist can appeal to Russell's distinction which is both precise and wards off Psillos' objections. In particular, Psillos' claim that the nature of an entity is no less knowable than its structure cannot be upheld if we adopt the Russellian view that 'nature' just means the non-isomorphically specifiable part of entities. Moreover, Psillos' idea that the nature and structure of an entity form a continuum is a badly chosen and ineffective metaphor since: a) Russell's definition allows for a crisp distinction between nature and structure, and b) it is inconsistent with the idea that the extension of the predicate 'structure of an entity' is a proper subset of the extension of the predicate 'nature of an entity', or even with the idea that the two predicates have the same extension.

Reply to PS5: The so-called 'dilemma' for the upward path to ESR is ill conceived. The first horn of the dilemma rests on a misrepresentation of ESR as a position that relies strictly on isomorphic mappings. In any case, given MR, it was shown how one can arrive at isomorphic mappings from weaker claims about injective mappings. Psillos' claim, that the structural realist cannot account for the possibility that the unobservable world may have extra structure not manifested in the perceptual world, was shown to be false. ESR just requires that every perceptual structure has a corresponding unobservable structure, not vice-versa. The bottom line is that some inferential knowledge about the structure of the world can be safeguarded, and that is all that is needed. The second horn of the dilemma rests on the mistaken idea that H-W and W-H concede too much to idealism. Contra Weyl, whom Psillos supports, it was indicated that W-H is not the central tenet of idealism for it is compatible with epistemic realism about the mind-independent physical world. At any rate, ESR does not require commitment to the W-H principle. Nor does ESR stipulate that there is absolutely no variance between the structure of the external world and the structure of our perceptions. Rather, there is evidence to believe that such variance is limited enough to be incapable of significantly undermining the reliability of inferential knowledge about the structure of the external world that ESR demands.

Reply to PS6: ESR does not support the view that first-order physical properties are completely unknowable. Rather, it supports the view that *all* physical properties can be known up to isomorphism. This mistake notwithstanding, the question can still be asked as to why knowledge should be restricted thus. The answer to this question takes the following form: We can know the physical relations without knowing the relata, on the presupposition that these relations are reflected in relations between observables. Knowing a relation without knowing the relata just amounts to knowing the logico-mathematical properties of the relation. This, in turn, amounts to knowing the isomorphism class to which these relations belong. Indeed, to know just the isomorphism class entails that the properties of the relata cannot be uniquely identified, but only up to isomorphism. Psillos' further claim that first-order properties must be knowable in their entirety because they are part and parcel of their causal role does not, as it stands, carry any weight since no argument is provided to support it.

Reply to PS7: Knowing the abstract structure of the world does not merely amount to knowing narrowly construed formal properties such as symmetry, reflexivity, etc. The narrow construal is unsatisfactory since we take formal properties to be all the logico-mathematical properties. In fact, this is the way Russell understood the concept. Under such a broad construal, the notion of abstract structure coincides fully with that of formal properties. This is so if one takes the view, which is correct to my opinion, that mathematical objects can only be described up to isomorphism. Since mature science is mathematical, it is not unreasonable to assume that scientific objects can only be described up to isomorphism. Finally, although underdetermination is unavoidable, ESR imposes significant curbs on its impact, curbs that realists should rest content with.

4

THE NEWMAN OBJECTION

1. Introduction

The aim of this chapter is to evaluate whether ESR can withstand a cluster of objections that fall under the umbrella of the 'Newman objection'. I will first present a detailed account of the objection and some associated results. This will be followed by an assessment of various replies to the objection that have been proposed in recent years. Finally, I will offer my own solution to the objection.[128]

2. Newman's Bifurcated Challenge

As we have seen, it was not until *The Analysis of Matter* (1927) that Russell wholeheartedly embraced and developed the structuralist viewpoint. In 1928, one year after its publication, M.H.A. Newman, a famous mathematician and subsequently Bletchley Park code-breaker, published a critical notice of Russell's book. There he argued that the ESR claim that we can know only the (abstract) structure of the external world trivialises scientific knowledge. Moreover, he argued that the only way to get around this problem involves giving up ESR.

The First Fork: *ESR Knowledge Claims are Trivial*

Newman begins by noting that Russell's view, that we have no knowledge of the *physical* relations over and above their formal (i.e. structural) features, amounts to the assertion that "[*t*]*here* is a relation *R* such that the structure of the external world with reference to *R* is *W*" (1928: 144). By this, I presume, he means that the structural realist cannot state the particular relations that the external world instantiates. Rather, the structural realist can only state that there exist such relations of which the only thing we can know is their abstract structure *W*. Newman then urges us to consider the logical theorem that for "any aggregate A, a system of relations between its members can be found having any assigned structure compatible with the cardinal number of A" (140). According to this theorem, the mere number of members in an aggregate entails that there are systems of relations

[128] A shorter version of this chapter has been published as Votsis (2003b)

definable over those members having any specified structure. Thus saying, as structural realists do, that for a given class there exists a system of relations that specifies a structure, is not saying much since this claim follows as a matter of logic by employing the above noted theorem plus the cardinality of the given class. But surely anything that is known about the external world must be discoverable empirically, not a-priori. Yet the only thing open for empirical determination under Russell's view, according to Newman's argument, is the cardinality of the given class.

Newman's argument can be understood as a *modus tollens*. If epistemic structural realism is true, then scientific knowledge imparts information only about the cardinality of the external world. But surely, science gives us more knowledge than this. Therefore, epistemic structural realism is false.

We can now state the theorem upon which Newman's result is based:

(NT) Newman's theorem: Let $S = (U, R_1, \ldots, R_k)$ be a structure and V be a set. Suppose that there is an injection $\rho: U \rightarrow V$. Then, there exists a structure S' whose domain is V, and which has a substructure isomorphic to S.

The proof for this theorem can be given as follows:[129] We begin by defining the image of mapping ρ as $\rho(U) := \{x \in V: \exists\alpha \in U, \rho(\alpha) = x\}$. From this we know that $\rho(U) \subset V$ and since ρ is injective we know that $\rho: U \rightarrow \rho(U)$ is a bijection. Its inverse is thus, $\rho^{-1}: \rho(U) \rightarrow U$. We can now define a relation R_i', for each n-place relation R_i on U, on the set $\rho(U)$ as follows: $R_i' := \{(x_1, \ldots, x_n) \in V^n : x_1, \ldots, x_n \in \rho(U) \wedge (\rho^{-1}(x_1), \ldots, \rho^{-1}(x_n)) \in R_i\}$. In other words, R_i' is an n-place relation on V. Note that it follows from the definition of each R_i' that $\forall\alpha_1, \ldots, \alpha_n \in U, (\rho(\alpha_1), \ldots, \rho(\alpha_n)) \in R_i'$ iff $(\alpha_1, \ldots, \alpha_n) \in R_i$. This is the condition for an isomorphism. By repeating this for every relation R_i on U we define relations R_i' on V and hence have a structure $S' = (V, R_1', \ldots, R_k')$. If we now take the *restriction* of S' to the subdomain $\rho(U) \subset V$ we observe that it is just the substructure $(\rho(U), R_1', \ldots, R_k')$ which is isomorphic to S, i.e. $(\rho(U), R_1', \ldots, R_k') \approx S$. This just means that save for cardinality constraints

[129] Many thanks to Jeff Ketland for providing this proof.

we can impose any structure on a set; that structure being of course set up by *some* relation(s). Thus, saying that 'There exists some relation R which has a specified structure S' is not saying much, since that follows trivially modulo cardinality constraints.

The Second Fork: *ESR cannot be Salvaged*

Newman correctly points out "that it is meaningless to speak of the structure of a mere collection of things, not provided with a set of relations" and "[t]hus the only important statements about structure are those concerned with the structure set up … by a given, definite, [set of] relation[s]" (140). The only way to avoid trivialization, according to him, is by specifying the particular relation(s) that generate(s) a given structure. That is, if we uniquely specify R, instead of just saying 'There exists some relation R which has a specified structure W', the fact that R has structure W is no longer trivial. The problem is that to uniquely specify R, one inevitably goes beyond the epistemic commitments of the structural realist, i.e. ESR in its pure form must be abandoned.

Let us remind ourselves of the structural realist's epistemic commitments. Russell claims that we can at most know the abstract structure of physical relations but not the relations themselves, for we have no epistemic access to them.[130] A consequence of this view is that there is an underdetermination of the physical relations by the abstract structure (URS for short), since infinitely many such relations correspond to any such structure. For some, this is the essence of Newman's problem. Ladyman certainly thinks so when he says: "[t]here are serious difficulties… which were raised by Newman (1928)… the basic problem [being] that structure is not sufficient to uniquely pick out any relations in the world" (1998: 412). To explain the objection better, consider first the following example (taken out of context from Newman's paper) that illustrates two different relations that share the same abstract structure:

> Let a set, *A*, of objects be given, and a relation *R* which holds between certain subsets of *A*. Let *B* be a second set of objects, also provided with a relation *S* which holds between certain subsets of its members… For example *A* might be a random collection of people, and *R* the two-termed relation of being acquainted. A *map* of *A*

[130] The use of the term 'physical relation' carries no intention, either by Russell or by me, to reify the relations found between physical objects.

> can be made by making a dot on a piece of paper to represent each person, and joining with a line those pairs of dots which represent acquainted persons. Such a map is itself a system, *B*, having the same structure as *A*, the generating relation, *S*, in this case being “joined by a line”. (1928: 139)

Newman, like most mathematicians, tends to avoid using different symbols for a *set* and for some *structure* with that set as domain, since it is usually clear from the context what he is referring to. For the sake of clarity, we shall use asterisks to indicate a structure as opposed to the set that constitutes the structure’s domain. Thus, structure *A** and structure *B** have set *A* and set *B* respectively as their domain. The two relations, ‘being acquainted’ and ‘joined by a line’, are undoubtedly distinct from one another both intensionally (i.e. what they mean differs) and extensionally (i.e. what they denote differs). In this context, however, they are employed in such a way that the structures they give rise to are isomorphic to one another. Hence, they share the same abstract structure. Now suppose that we are interested in only one of these relations but we have epistemic access to neither. If all we have knowledge of is (abstract) structure, as the structural realist suggests, we cannot distinguish between the two relations. And, of course, we do not just have two relations to choose from but infinitely many, since there can be infinitely many bijective mappings that preserve the same structural properties.

Newman explores several ways in which the structural realist might try to distinguish between intended and unintended systems of relations, i.e. structures. Two of these ways stand out. The first one is an attempt to dress the distinction as one between real and fictitious relations. Newman defines a relation as fictitious when “the relation is one whose only property is that it holds between the objects that it does hold between” (145). Real relations can then be implicitly defined as those relations that have more than just this property. This, according to Newman, is obviously not going to be of help, since the *only knowledge* a structural realist would have of the real relations is exactly the same knowledge he would have of the fictitious ones, viz. that they hold between some objects. But what if we know something about these objects apart from their having a given structure, could we not then claim to have a way to distinguish the real relations from the fictitious ones? For instance, if we fix the domain in the above example to the set *A*, i.e. the set whose members are people, then, at least prima facie, there is no longer a question of being unable to distinguish between the two relations 'being acquainted' and 'being joined by a line'.

Anticipating this reply, Newman argues that even if the domain of the objects has been specified, we are still left with the problem that we must "distinguish between systems of relations that hold among members of a *given* aggregate" (147) [my emphasis]. Demopoulos and Friedman elaborate that "[t]his is a difficulty because there is *always* a relation with the [given] structure" (1985: 628-629). This point can be extended by showing that isomorphic relations, defined over the same domain, can yet be different in some important respects. Example: Suppose we make another mapping of *A* by painting a line between all, and only those, people who are acquainted. Let us call the resulting structure 'C*' and its generating relation 'T'. Notice that *C** is isomorphic to *A**, which means that they share the same abstract structure, let us call it 'S*'. Notice also that *C** and *A** have the same set of objects as their domain, viz. set *A*. However, *A** and *C** are generated by different relations – at least if these relations are considered as relations-in-intension. *A** is generated by *R* while *C** is generated by *T*. Thus, knowing the abstract structure S* and fixing the domain to set *A* does not allow us to uniquely pick out the so-called intended relation, whatever that relation may be. But surely, the argument goes, being able to point to the intended relation must be an essential part of scientific enterprise and knowledge.

Newman's other attempt to distinguish between intended and unintended relations takes the form of a distinction between important and unimportant (or trivial) relations. But how is this distinction to be made, Newman asks, if we are to "compare the importance of relations of which nothing is known save their incidence [i.e. occurrence] (the same for all of them) in a certain aggregate" (147). The only way to do that without giving up ESR, Newman reasons, would be to take the term 'importance' as one of "the prime unanalyzable qualities of the constituents of the world", something he considers completely absurd (147).

Newman concludes that if we are to avoid trivialization, we must surrender "the 'structure vs. quality' division of knowledge in its strict form" (147). But to surrender this distinction, he claims, is to render ESR false. As Demopoulos and

Friedman explain "since it is indisputably true that our knowledge of structure is non-trivial – we clearly do not stipulate the holding of the structural properties our theories have – it cannot be the case that our knowledge of the unperceived parts of the world is *purely* structural" (1985: 630).[131]

In a letter sent to Newman shortly after the publication of his critical notice, Russell acknowledged that he was wrong in saying that only the structure of the physical world can be known (see his (1968: 176)). Thereafter, Russell abandoned pure ESR (see for example Russell (1948)), and never returned to address Newman's problem.

A Note on the Ramsey-Sentence

As indicated in chapter two, Demopoulos and Friedman recast Newman's objection against the Ramsey-sentence approach. Their main point is that if a theory Θ is consistent and all its observational consequences true, then the truth of Θ's Ramsey sentence is guaranteed and hence structural realism (nearly) collapses into phenomenalism. Indeed, the only thing separating ESR from phenomenalism, say Demopoulos and Friedman, is the cardinality constraint. This is so because it is taken to say something, i.e. how many types of objects and properties exist, about the unobservable world.[132]

Building on Demopoulos' and Friedman's work, Jeff Ketland (forthcoming) argues that Ramseyfication "yields a claim which is, roughly speaking, *equivalent* to the claim that the theory is empirically adequate" (2).[133] More precisely, he argues that saying that the Ramsey sentence of a given theory Θ, R(Θ), is true is equivalent to saying that Θ has an empirically correct full model whose theoretical domain possesses the right cardinality. By 'empirically correct' he means that the *observable* domain and relations designated by the model is isomorphic to the appearances.[134] From this, he concludes that "the 'structural content' of a theory Θ, at least *if* it is

[131] I presume that by 'unperceived parts of the world', they simply mean what Russell and Maxwell mean by the term 'unobservable' (see ch.2, §5).
[132] Russell raises this point in the aforementioned letter to Newman, saying, however, that it "is not a point upon which I wish to lay stress". He also notes that the cardinal number under consideration must be finite if the claim is to have any empirical significance.
[133] Page numbers are from the preprint.
[134] Ketland assumes a two-sorted interpreted language where first-order variables range over two domains, one for observable and another for unobservable/theoretical objects.

identified with what R (Θ) 'adds' to the claim that Θ is empirically adequate, is just this Newman-esque cardinality constraint" (11). Thus, Ketland's claim is that structural realism à la Ramsey is not very different from anti-realist positions such as constructive empiricism. The difference, again, lies in the cardinality constraint, which presumably says something about the unobservable world.

3. Various Replies to the Objection

As we have seen, Newman argues that if we want to represent empirically determinable scientific knowledge, structure is not enough. Epistemic structural realists such as Grover Maxwell, John Worrall, Michael Redhead, and Elie Zahar share so much with Russell's structural realism that it is no surprise that Newman's objection is dusted off and pitted against them. Ever since the Demopoulos and Friedman paper, many authors (see, for example, van Fraassen (1997), Ladyman (1998), Psillos (1999) and Ketland (forthcoming)) have appealed to the objection to attack some version of ESR. In this section, I will be appraising replies to the objection offered by Psillos, Redhead, French, Ladyman, Worrall and Zahar.

Psillos

Although not a structural realist, Psillos (1999; 2000a; 2001a; 2001b) has ventured to show that the only way to overcome Newman's objection is to espouse a traditional scientific realist position, forsaking the purely structural claims of ESR. In particular, Psillos argues that if the world is structured in a unique way, then we can simply pick out the right relations, and, hence, the right structure. He thus says, "[i]f the domain is already 'carved up' in natural kinds, then it is a far from trivial exercise to find a network of relations which generate a certain formal structure… instead of being a matter of logic, this issue becomes an open empirical problem" (1999: 68). How does he propose to pick out the right relations? Psillos only hints at how this may be done when he says that what is "required is getting the extensions right, i.e. identifying those and *only those* extensions which mark off the boundaries of – and relations among – natural kinds" (68).

The claim that the world is carved up in natural kinds does not, by itself, solve the Newman problem. After all, if the claim is true, it merely guarantees the existence of physical properties and relations, but says nothing about our epistemic access to them. Psillos, in fact, acknowledges that the natural kinds claim is compatible with ESR. He says "[t]he thesis that the world possesses a unique natural-kind structure is surely compatible with the claim that, without also specifying *what exactly* these natural kinds are, of this structure the only thing that *can* be asserted is that it exists" (68) [original emphasis]. What Psillos could say in addition, and perhaps he implies it, is that we have direct access to the external world. Under these conditions, we can empirically discover its physical relations and, hence, its unique natural-kind structure. The problem with this suggestion is that it simply assumes what ESR denies, i.e. that we have direct access to the world. Psillos does not offer any justification for this assumption and for this reason we must leave this discussion aside.

Strangely enough, though Psillos admits that the natural kinds claim is compatible with ESR, he claims that acceptance of the claim leads to two unwanted consequences for ESR: (1) ESR is rendered false, and (2) the Ramsey-sentence approach, that some structural realists advocate, is rendered unsatisfactory. Does Psillos' natural-kinds answer to Newman's objection, if it indeed is an answer, exclude ESR from partaking in it?

Let us consider Psillos' allegedly unwanted consequences for ESR one at a time. It is not entirely clear what he means by the first one. Perhaps what he wants to say is that the natural-kinds claim, i.e. the claim that the world is carved-up in natural kinds, is itself 'non-structural'. That is, it is an assumption over and above those allowed by ESR-ists, so if it is adopted one cannot maintain a purely structural knowledge of the world. Yet the natural-kinds claim is an ontological assumption about what the world is like, and, as such, it adds nothing to the epistemological commitments of the structural realist. Moreover, even if these commitments were to be burdened with a non-structural claim, it is far from clear that the resulting 'impure' ESR must be abandoned for the sake of full-blown realism, as Psillos ultimately wants. Unless Psillos provides some justification as to why we should accept the first consequence, or, indeed, any of the other claims I just mentioned, ESR emerges unscathed.

Disappointingly, Psillos' justification for his second consequence is also mere hand waving. He claims that the Ramsey-sentence approach proves unsatisfactory,

> …for when it comes to the defence of the thesis that science captures the structure of the world already carved up in natural kinds, it is best not to treat theories as abstract structures, but instead to appeal to the success of *interpreted* scientific theories in order to argue that the kinds posited by them populate the world (69).

In other words, the success of interpreted scientific theories presumably provides a better defence of the natural kinds claim. The implication here is that the structural realist commits himself to uninterpreted theories, an accusation that we have witnessed on Psillos' list of objections under PS1. I argued in the previous chapter (see section three) that this objection fails since the structural realist does not subscribe to uninterpreted, i.e. purely syntactic theories. Rather, the structural realist requires that the observational terms of the theory be interpreted. In particular, the Ramsey-sentence of a theory contains interpreted observational terms, so it cannot be the case that the structural realist is committed to entirely uninterpreted theories. Similarly, the Russellian account commits one to, among other things, concrete observational structures.

Redhead

Redhead's (2001a; 2001b) stab at solving Newman's problem involves an attempt to dissociate ESR from the Ramsey-sentence approach. He dismisses the Ramsey-sentence formulation, viz. $\exists\Phi$ (W(Φ, O)), and opts for a logically stronger formulation, viz. W(R, O), where Φ is an existentially bound theoretical variable, *O* an observational term, *W* is an abstract structure, and *R* a specific *physical* relation.[135] Obviously, the truth of W(R, O) cannot be a trivially established. But how do we justify our specification of *R* in the first place? Redhead's solution is that we take *R* as being "hypothesised in some explanatory theoretical context so it exists as an ontological posit" claiming that "all that we have epistemic warrant for is the second-order [i.e. abstract] structure of [W]" (2001b: 346). Thus, we may remain epistemologically agnostic with regard to the specific relation *R*, while still asserting knowledge of its abstract structure *W*.

---

[135] The formulations are abbreviated for the sake of expediency.

The perhaps not so obvious difficulty with Redhead's solution is that our epistemic warrant for (abstract) structure *W* must somehow be derived from the nonexistent epistemic warrant of the hypothesised physical relation *R*. Psillos makes precisely this point in his reply to Redhead's review of his book, when he says that he "cannot see how we can have epistemic warrant to believe that the structure is [W] which is not parasitic on having epistemic warrant to believe that R is the required (definite) relation" (2001b: 369). He goes on to say that "[i]t is the knowledge of (or commitment to) the definite relation R… that issues the warrant to believe that its structure is [W] and hence issues the warrant that the relevant domain has structure [W]" (369). Even though the (abstract) structure *W* of a relation *R*, i.e. the broadly construed formal properties of *R*, is logically weaker than *R*, and thus presumably more epistemologically warranted, one needs first to establish that *R* has some epistemic warrant in order to be able to infer *W*. The problem is that admitting that *R* has epistemic warrant amounts to abandoning pure ESR, for we can no longer claim that only abstract structure can be known.

French and Ladyman

Ontic structural realists French and Ladyman have yet another solution to Newman's objection. They argue as follows:

> Worrall's approach is thoroughly embedded in the so-called syntactic view of theories that adopts first-order quantificational logic as the appropriate form for the representation of physical theories. We will not rehearse our reasons here, but we consider this approach to be deeply flawed, not only because of its inadequacy in reflecting scientific practice, but also because of the pseudoproblems that arise once one has adopted it. So for example, the Newman problem is obviated if one does not think of structures and relations in first-order extensional terms (2003a: 33).

If, in their view, we forgo the syntactic view of theories of which the Ramsey sentence approach is a cornerstone, then the Newman problem will no longer be an obstacle. Their alternative is the model-theoretic approach to theories. At no point, however, do they really elaborate how the Newman problem can be obviated with the adoption of this approach.[136]

[136] For an argument against the idea that by abandoning extensionalism we can avoid Newman's objection see Demopoulos and Friedman (1985: 629-30).

The problem seems equally threatening to ontic SR, regardless of which approach one takes to theories, for taken together (ontic SR and the Newman problem) lead to the view that we can have no non-trivial knowledge of the world. That is, assuming the following two premises are true, (a) The Newman problem, i.e. if all we can know about the world is structure we cannot know anything of importance, and (b) all we can know of the world is structure, for all there exists is structure, we derive (c) we cannot know anything of importance about the world. There is nothing particular in ontic SR, or any of the accompanying claims that French and Ladyman make, that saves it from Newman's problem.

Worrall and Zahar

Another reply to Newman's objection originates in Worrall (2000) who, as we have seen, advocates a version of ESR augmented by the Ramsey-sentence approach. He argues that provided a distinction between observational and theoretical terms is made, "the fact that, for every [theory] S, S and its RS [Ramsey sentence] S* are empirically equivalent, [entails that] S* is co-falsifiable with S and hence for any falsifiable S, its RS S* cannot tell us *only* about the size of the universe" (8). In other words, if the theory itself is empirically falsifiable then its Ramsey-sentence must also be empirically falsifiable for the two are empirically equivalent. Hence, the Ramsey-sentence of such a theory cannot be trivial.

An elaboration of this argument can be found in an appendix to Zahar (2001), co-authored by Worrall and Zahar. There it is claimed that Russell's mistake was to opt for a purely structural representation that contains no observational terms:

> It was admittedly unfortunate that in his 1927 Russell spoke of a *purely* structural description of reality being inferred from perceptual results. The fault lies… with the implicit assumption that once the inference is carried out, an exclusively structural account is obtained in which no observational terms occur (238) [original emphasis].[137]

[137] Worrall, in private communication, claims that the point does not come out correctly in the text just quoted. According to him, it is Newman who mistakenly viewed Russell as giving a purely structural description that affords no observational terms. Russell's only fault, claims Worrall, was to forget that his whole view was based on the acquaintance/description distinction, a view that necessitates observational terms.

A sentence expressing this pure structure would, of course, be completely uninterpreted. If that is then Ramseyfied, the only information it would yield would be that the unramseyfied sentence has at least one model whose domain can be taken to be countable.[138] Worrall and Zahar instead opt for sentences with interpreted observation terms. Ramseyfying these sentences keeps the observation terms intact, i.e. interpreted, which means that the resulting Ramsey-sentences are empirically falsifiable and, therefore, not trivially satisfiable.

Worrall and Zahar claim that Newman's objection applies only if a distinction between observational and theoretical terms is *not* made. Indeed, according to them, the distinction must be made. Once made, they say, Ramsey-style ESR only faces Demopoulos' and Friedman's extended objection, namely that what the Ramsey-sentence asserts 'over and above' its observational content is reducible to logic or mathematics. In more detail, Demopoulos, Friedman and recently Ketland argue that Ramsey-style ESR collapses to phenomenalism/constructive empiricism, since the only epistemically-relevant thing left (after the reduction of what is 'over and above' to logic/mathematics) in a Ramsey-sentence is the observational content. To be more precise, the accusation is one of *near collapse*, since, as Demopoulos, Friedman, and Ketland admit, there is always the something 'over and above' to worry about.

Against these claims, Worrall and Zahar argue that "[t]his 'over and above'… seems to us to be not only badly defined but also indefinable in any non-trivial way" (240). They cite two reasons for this. First, the Ramsey-sentence does not logically follow from its observational content. The Ramsey-sentence contains empirical generalisations and hence cannot be deduced from the observational content of the theory since, according to them, the latter contains no such generalisations. According to them, to include empirical generalisations in the observational content would be to go "against the canons of even the most liberal version of empiricism" (241). In this sense, they argue, the Ramsey sentence makes non-trivial claims, i.e. claims over and above the observational content that are not reducible to logic or mathematics.

[138] Strictly speaking, Worrall and Zahar should not call this 'Ramseyfication' since no observation terms are present in the original sentence.

Second, Worrall and Zahar argue that if we were to count empirical generalisations as part of the observational content of a theory, "then the Ramsey-sentence *might* well be one of them; in which case Demopoulos' and Friedman's thesis would collapse into the trivial claim that the Ramsey-sentence follows from itself" (240) [my emphasis]. What Worrall and Zahar seem to say in this blurry passage is that under a broad construal of the observational content, i.e. a construal that includes empirical generalisations, the Ramsey-sentence itself can be regarded as purely empirical. In that case, there is absolutely nothing 'over and above' to worry about, hence Demopoulos' and Friedman's claim loses its potency. Presumably, this empirical understanding of the Ramsey-sentence is not threatening to Worrall and Zahar, because they assume that the observational content broadly construed goes beyond what an anti-realist is willing to accept.

There are several inadequacies with Worrall's and Zahar's defence of the Ramsey approach to epistemic structural realism. First off, contra Worrall's and Zahar's remarks, there is a clear sense in which at least part of the meaning of 'over and above' is fixed. To wit, a Ramsey-sentence says something about the cardinality of the theoretical domain and how it is structured. Moreover, Demopoulos, Friedman, and Ketland complain that there is nothing in the Ramsey-sentence that restricts the theoretical domain to *those and only those objects* that are unobservable.[139] If no such restriction can justifiably be made, Ramsey-sentence assertions are to be understood as indiscriminately applying to abstract objects, including mathematical entities. The claim that a given number of mathematical objects can be structured in a particular way then becomes a matter of logic/mathematics; that is, it follows from Newman's theorem. The challenge for the Ramsey-style ESR-ist is to show whether or not a restriction can be justifiably imposed on the 'theoretical' domain, so that theoretical variables range over all and only those objects that are unobservable, i.e. physical objects.[140]

[139] Ketland correctly recognises the need to have an interpreted language where first-order variables range over *two* domains, one for theoretical entities (i.e. unobservables, mathematical entities, etc.) and the other for observables.

[140] My inclination is that a restriction to the theoretical domain can be justified as an ontological assumption and, hence, without any compromise on the epistemic commitments of structural realism.

Another problem with Worrall and Zahar's defence is their narrow construal of the concept of observational content. The construal is highly questionable, for empiricists, like van Fraassen, have traditionally classified empirical generalisations as part of the observational content of a theory. Perhaps there is no good reason for empiricists to do so, but Worrall and Zahar need to explain why we should suddenly alter our understanding of empiricism. Likewise, I am unaware of any realist positions in the literature that take commitment to empirical generalisations as sufficient support for realism.[141] Again, a lot more needs to be said on this issue before Worrall's and Zahar's argument can gain credibility.

The problems do not stop there. If empirical generalisations provide sufficient support for structural realism, why do we need Ramsey-sentences at all? The answer would most probably be that a Ramsey-sentence does not merely provide empirical generalisations, but also makes existential and generalised claims about theoretical predicates. Worrall and Zahar fail to explain why we need these existential claims, if, as shown above, what delineates structural realism from empiricist antirealism is empirical generalisations. Moreover, their assertion that the theoretical terms found in existential claims are 'indissolubly entangled' with the observational terms remains suggestive, pending further elaboration on their behalf. How, one may well ask, could theoretical terms be indissolubly entangled with observational ones if indefinitely many Ramsey-sentences, with syntactically different existential claims, are compatible with the observational content broadly construed? To create such 'rival' Ramsey-sentences, all one needs to do is take the observational content of a given Ramsey sentence and simply construct other Ramsey sentences at will by appending syntactically different theoretical statements to the original observational content and then Ramseyfying.

The doubtfulness of the claim that theoretical terms are indissolubly entangled with observational ones is reinforced by the following considerations. Suppose Φ is a theory that entails a set of sentences S, and that S can be subdivided into two disjoint sets: O which contains true observation sentences and T which contains theoretical

[141] Even Russell, who comes close to supporting such a view, demands that empirical generalisations mirror the structure of the physical world. In doing so, he commits himself to something more than the empirical generalisations, namely to the (at best) isomorphic correspondence between the physical world and the phenomenal world.

sentences. Suppose further that *O* contains *all* the observation sentences true of the world, and nothing else. We know that *Φ* and its Ramsey-sentence *R*(*Φ*) are empirically equivalent. So if *Φ* entails *O*, then *R*(*Φ*) entails *O*. Since *O* contains all true observation sentences (and no false ones), no observation sentence can undermine the theoretical part of the Ramsey-sentence. But then, how can one decide between various different manifestations of the theoretical part of the Ramsey-sentence? For example, how could we decide between $(\exists\psi)(\exists\phi)(\forall x)[(\psi x \text{ v } \phi x) \supset Ox]$ and $(\exists\psi)(\exists\phi)(\forall x)[(\neg\psi x \,\&\, \phi x) \supset Ox]$?

Jane English (1973) has proved that two or more Ramsey-sentences with the same observational consequences cannot be inconsistent. The proof is a reductio. Here's a reconstruction: We start by assuming that there are two inconsistent Ramsey sentences, $R(T_1) = (\exists\phi_1)\ldots(\exists\phi_n)T_1\,(O_1, \ldots, O_k; \phi_1, \ldots, \phi_n)$, and $R(T_2) = (\exists\psi_1)\ldots(\exists\psi_m)T_2\,(O_1, \ldots, O_k; \psi_1, \ldots, \psi_m)$ that are compatible with all possible observations. Reversing the process of Ramseyfication, we substitute distinct predicate letters for the second-order variables and get $T_1(O_1, \ldots, O_k; F_1, \ldots, F_n)$ and $T_2(O_1, \ldots, O_k; G_1, \ldots, G_m)$. Trivially, T_1 & T_2 implies $R(T_1)$ & $R(T_2)$. If the pair $R(T_1)$ and $R(T_2)$ is inconsistent, so is T_1 and T_2. The next step requires appeal to Robinson's consistency theorem, which holds that $T_1 \cup T_2$ is satisfiable if and only if there is no sentence in the common vocabulary of their languages such that one theory entails this sentence and the other entails its negation.[142] By the completeness theorem we infer that T_1 and T_2 are inconsistent if and only if there is at least one sentence in the common vocabulary of their languages such that one theory entails this sentence and the other entails its negation. The common vocabulary in this case is the observational vocabulary, for Ramseyfication turns the theoretical vocabulary into variables. Hence, it follows that there is a sentence O in the observational vocabulary of both $T_1(F_i)$ and $T_2(G_i)$, such that $T_1(F_i) \supset O$, and $T_2(G_i) \supset \neg O$. Suppose O is true. By modus tollens, we infer that T_2 is false in virtue of its being not compatible with all possible observations. Since T_2 and $R(T_2)$ are empirically equivalent, this means that $R(T_2)$ is also not compatible with all possible observations. Contradiction! Suppose O is false. By modus tollens, we infer that T_1 is false in virtue of its being not

[142] The theorem is stated thus: Let L_1 and L_2 be two languages where $L = L_1 \cap L_2$. Suppose that T_1 and T_2 are theories satisfiable in L_1 and L_2 respectively. Then $T_1 \cup T_2$ is satisfiable if and only if there is no sentence σ in L such that $T_1 \vDash \sigma$ and $T_2 \vDash \neg\sigma$.

compatible with all possible observations. Since T_1 and $R(T_1)$ are empirically equivalent, this means that $R(T_1)$ is also not compatible with all possible observations. Contradiction! Therefore, any Ramsey sentences that are compatible with all possible observations cannot be inconsistent with one another, i.e. they will be consistent.

Although the starting assumption in the above proof is that the Ramsey-sentences are compatible with *all possible observations*, we can replace this with the assumption that the Ramsey-sentences have the *same set of consequences* expressed in the observational vocabulary. Notice that it still follows that there is a sentence O in the observational vocabulary of $T_1(F_i)$ and $T_2(G_i)$, such that $T_1(F_i) \supset O$, and $T_2(G_i) \supset \neg O$. This, in effect, means that the two theories do not share the same set of consequences expressed in the observational vocabulary, something that contradicts our starting assumption. Thus we can conclude that any two Ramsey sentences that have the same observational consequences cannot be inconsistent, i.e. they will be consistent.

The severity of the problem becomes apparent when one considers that there are indefinitely many inconsistent alternatives to a theory that have the same observational consequences. The Ramsey-sentences of all of these theories are logically consistent with each other. Yet, the original (non-Ramseyfied) theories were inconsistent! The reason for this mismatch stems from the substitution of theoretical predicates for existentially quantified variables. Take, for example, two contradictory statements such as $(\forall x)Px$ and $(\forall x)\neg Px$, where *P* is a theoretical predicate. When we Ramseyfy these statements we get $(\exists\phi)(\forall x)\phi x$ and $(\exists\psi)(\forall x) \neg\psi x$ respectively. Notice that nothing prevents $\neg\psi \equiv \phi$.

Worrall and Zahar suggest that we can choose the right Ramsey-sentence by recourse to such considerations as unity and simplicity. Here’s what they say:

> The more demanding structural realist can go much further than this, but only by appealing to a metaphysical principle similar to the one invoked by both Poincaré and Einstein. The latter put it as follows: the world *is* the realisation of what is mathematically simplest, while Poincaré held the degree of unity-cum-simplicity of a hypothesis to be an index of its truth-likeness (2001: 249).

Presumably then, the right Ramsey-sentence is the one that offers the most unified and simple account of the observables and unobservables. Needless to say, this claim requires fleshing out if it is to be taken seriously. Why should we take the notions of unity and simplicity as epistemically significant? It is well known that such notions are notoriously difficult to pin down, and that anti-realists like van Fraassen dismiss them as merely pragmatic considerations. Suppose for argument's sake that they are epistemically significant. Why should there be a *unique* Ramsey-sentence that captures all the true observational consequences and offers the most unified and simple account of the observables and unobservables? I will return to this issue in chapter six, where I will be considering arguments from underdetermination.

In sum, the Ramsey-style structural realist has a lot of explaining to do before she can gain a foothold in the scientific realism debate.

4. Overcoming Newman's Objection

I have already indicated in my discussion of Grover Maxwell's ESR (see ch.2, §5) that Russell's version of structural realism is incompatible with the Ramsey-sentence approach. Attacks on the Ramsey-sentence rendition of ESR, led by Demopoulos and Friedman and more recently by Ketland, therefore miss their target when it comes to Russell. I will argue in this section that upon closer scrutiny the accusation of triviality mounted against Russell's version of ESR fails, because the position's knowledge claims are informative. Moreover, I will expose as a myth the idea that we can uniquely pick out physical relations or indeed their relata.

First of all, it should be made clear that if *all* the structural realist is arguing for is the claim that there exist relations with particular structures, then this is obviously trivial for the reasons Newman mentions. But no structural realist makes such a claim! Russell, in particular, claims that our epistemic commitments extend to cover REC1, REC2, and REC3. That is, concrete observational structures, abstract structures (i.e. isomorphism classes) corresponding to the concrete observational structures, and the existence of concrete physical structures that are causally responsible for the concrete observational structures and belong to the same isomorphism classes as them. According to this view, we *empirically* identify logico-mathematical properties of the unobservable world via the assumption that causally linked concrete observational structures and concrete physical structures share the same abstract structures, i.e. belong to the same isomorphism classes.

Is Russell's claim trivial? Before we can answer this question, we first need to understand what exactly is meant by the characterisation that ESR knowledge claims are trivial? Customarily, the term 'trivial' means that the information on offer is of little or no importance. So, in what way are the knowledge claims of ESR of little or no importance? The well-rehearsed answer is that the information these claims offer can also be derived *a-priori* from set theory together with a cardinality constraint. Hence, the answer goes, the only important information contained in the structural claims concerns the cardinality of the domain. This seems to imply that any information contained in a statement that is also derivable a-priori lacks importance.

There is a very simple and straightforward way to show that this view is mistaken. Consider the following example: Take the numbers 133 and 123. I can, restricting myself solely to arithmetic, perform various operations on these numbers. One such operation is addition. Similarly, if I had two collections of 133 and 123 physical objects respectively, I could count them one by one and would reach the same result.[143] Despite the similarities, there is an important difference between the two cases. In the latter case, the result is a property that is then ascribed to the physical world, in particular to the physical objects under consideration, and not merely an exercise of arithmetic. This claim is warranted by the employment of an *empirical method* to arrive at the given number. The fact that arithmetic allows me to perform a calculation between the numbers 133 and 123 a-priori – numbers that happen to be the same as those involved in the two collections of physical objects – does not mean that the information I reached counting the physical objects in the two collections is devoid of importance. One need only consider the consequences had I made an error in counting.

The same type of argument can be applied to the issue at hand. Provided that we have the right cardinality, we can set up any structure – and by extension any

[143] Assuming that the objects retain their individuality when I add them.

abstract structure – we want a-priori just by appeal to set theory and in particular the theorem employed by Newman. But we can also reach the same abstract structure a-posteriori. Empirical investigation leads us to the discovery/postulation of relations between observables. By deduction, we can then get the abstract structures of these relations. Appeal to H-W and MR, principles that are questionable but not the target of Newman's objection, allows us to infer that relations between observables and the corresponding relations between unobservables share the same abstract structures. The *method* of arriving at the abstract structures is at least partly empirical in that the discovery of relations between observables is an empirical matter. Thus, the fact that set theory allows me to arrive at the same abstract structures a-priori does not mean that the information I have reached via observation is devoid of importance.

One further consideration should make the general point sharper. It is a feature of the expressive power of mathematics that it can give us all the structures that satisfy any given cardinality constraint. No structure is privileged in this sense. Russell's a-posteriori method guarantees that some structures are privileged over others. According to this method, we choose those abstract structures that are instantiated by concrete observational structures. We choose them because of the assumption that concrete observational structures are isomorphic to their causal antecedents, namely concrete physical structures. In the above example, this would be analogous to the fact that although arithmetic allows me to sum any two numbers, there is only one number that can be correctly ascribed to the aggregate of the two collections of physical objects under consideration.

The critic may object that the weight of Newman's objection is that knowing only the abstract structures is not enough. But why is it not enough? Demopoulos and Friedman suggest, borrowing a concept from Quine, that without appeal to a background theory the abstract structure cannot single out the intended from the unintended interpretations:

> From a contemporary, model-theoretic standpoint, this is just the problem of intended versus unintended interpretations: Newman shows that there is always some relation, R, (on the intended domain) with [abstract] structure W. But if the only constraints on something's being the intended referent of 'R' are observational and structural constraints, no such criterion for distinguishing the intended referent of 'R' can be given, so that the notion of an intended interpretation is, in Quine's

> phrase, provided by our background theory, and hence, cannot be a formal or structural notion in Russell's sense (1985: 633).

Demopoulos and Friedman correctly point out that observational and structural constraints are not sufficient to determine the referent of ‘R’. Indeed, Quine argues that a background theory is required to fix the interpretation. Yet, he also argues that this fixing is by no means absolute. For Quine, a background theory provides an interpretation to the logical form of a theory by “picking a new universe for its variables of quantification to range over, and assigning objects from this universe to the names, and choosing subsets of this universe as extensions of the one-place predicates, and so on.” (1969: 53-54). Even with a background theory at hand, however, “the intended references of the names and predicates have to be learned rather by ostension, or else by paraphrase in some antecedently familiar vocabulary” (54). But, Quine goes on, “the first of these two ways has proved inconclusive”, since it faces the usual problems of the indeterminacy of reference. That is, learning by ostension cannot uniquely identify the underlying physical objects, as the familiar example of the term ‘Gavagai’ illustrates. The second way then “is our only recourse; and such is ontological relativity” (54). In other words, paraphrasing in some antecedently familiar vocabulary just brings us back to where we started, for it is an appeal to another background theory. As Quine notes, “[since] questions of reference of the sort we are considering make sense only relative to a background language [or theory], then evidently questions of reference for the background language make sense in turn only relative to a further background language [or theory]” (54). That, of course, leads to a regress. The moral of the story is that the very choice of ontology, i.e. the background theory, is a relative matter.

The above illustrates that Demopoulos and Friedman’s appeal to Quine, in order to support their claim that we can avoid the problem of unintended interpretations by employing a (non-structural) background theory, rests on a serious misrepresentation of his work. Ontological relativity is the idea that we cannot eliminate unintended interpretations altogether, i.e. that we cannot uniquely pick out physical relations or indeed their relata. It is in this spirit that Quine says “we cannot require theories to be fully interpreted, except in a relative sense” (51). What we can do, according to him, is impose observational, behavioural, and structural constraints to narrow down the number of unintended interpretations.

Some peculiarities of Quine's approach should be noted. Quine promotes the idea that one can reinterpret a theory, i.e. give it a new background theory, without any essential loss. He favours reinterpretation because he thinks that "what makes sense is to say not what the objects of the theory are, absolutely speaking, but how one theory of objects is interpretable or reinterpretable in another" (51). In more detail:

> Suppose… that in the statements which comprise the theory, that is, are true according to the theory, we abstract from the meanings of the nonlogical vocabulary and from the range of the variables. We are left with the logical form of the theory, or, as I shall say, the *theory form*. Now we may interpret this theory form anew… (53).

Though this technique is similar to the Ramsey-sentence approach, it must not be confused with it. The main difference is that, unlike the Ramsey-sentence approach, Quine's technique affects (by re-interpretation) both theoretical and observational terms. In Quine's eyes, the difference between the two is that "Ramsey's treatment brings out indeterminacy of reference not by *reinterpretation*, but by *waiving the choice of interpretation*" ([1995]1998: 74) [my emphasis].[144] Whether this marks a real difference is an open question that will not be pursued here.

Importantly, Quine draws an epistemological lesson that is similar to that of epistemic structural realism. This comes out clearly in many of his writings. For example in the discussion section of Grover Maxwell's original article on structural realism, he makes the following comment: "One central plank in Professor Maxwell's platform is that our knowledge of the external world consists in a sharing of structure. This is to my mind an important truth, or points towards one" (1968: 161). A decade and a half later he repeats, "structure is what matters to a theory, and not the choice of its objects" (1981: 20). Also, in a more recent article he says: "The conclusion is that there can be no evidence for one ontology over against another, so long anyway as we can express a one-one correlation between them. Save the

[144] Note also that Quine's position is called 'global structuralism', because he advocates structuralism about concrete and abstract objects alike. F.P. Ramsey advocated structuralism only with regard to theoretical objects.

structure and you save all" (1992: 8). Hence, far from being critical of ESR, as Demopoulos and Friedman have suggested, Quine's views and especially his idea of a relativised background theory lend more credence to it.

Returning to URS, i.e. the underdetermination of the physical relations by the abstract structure, we can say that this type of underdetermination is not a real threat for ESR but rather another way of endorsing it. According to the epistemic structural realist, the limit of knowledge is description up to isomorphism. That may seem too little for some but this knowledge is already a considerable improvement over complete ignorance and, more importantly, over the knowledge claims of the empiricist. To understand the extent to which this is an improvement, let us contrast URS with all-out underdetermination. Suppose we have absolutely no information about the world, i.e. we know nothing of the objects that inhabit it, nothing of their number, none of their properties and relations, etc. This state of knowledge allows for a complete underdetermination of what the world is like if it were to be accurately described. Thus any logically possible object, property, and relation is a candidate for the role of correctly representing the furniture of the universe. Consider what happens if we adopt ESR. Certain relations between observables, and hence their corresponding abstract structures, are singled out for their ability to produce highly accurate predictions. These abstract structures can be thought of as setting constraints on the type of objects, properties and relations that are suitable candidates for the role of correctly representing the ontology of the world. In terms of our discussion of intended vs unintended interpretations, this just means that the abstract structures will set constraints as to which background theories/ontologies/interpretations are allowed i.e. are consistent with the abstract structures.

The overall claim is not that the problem of unintended interpretations does not pose an epistemic obstacle. Rather, the claim is that it is a kind of obstacle that realists can live with and, if the structural realist or Quinean is right, one they must live with, for there is no *empirically justifiable* way in which we can uniquely pick out the ontology of the world. A 'non-structural' realist may of course object that there are ways in which we can justifiably eliminate underdetermination altogether or at least restrict it even further by appealing to non-structural considerations. This is a legitimate reply but one which needs to be backed up by evidence. Until that happens the ‘non-structural’ realist cannot substantiate his claim that ESR cannot deliver as much knowledge of the world as can be had.

5. Conclusion

Newman’s objection and the associated results are not the knock-down punch to ESR as critics maintain. Yet, the perception that it is a knock-down punch persists in the wider philosophical community. In a way, this is not surprising, seeing as the attempt to rebut Newman’s objection is a recent phenomenon. Indeed, if my assessment of the various replies offered to the objection is correct, most attempts to rebut the objection have been largely unsuccessful.

Worrall and Zahar’s attempt represents the only exception. They convincingly argue that Ramsey-style ESR’s knowledge claims cannot be trivial, thereby disarming Newman’s objection. Things do not look as rosy for Ramsey-style ESR, however, when it comes to associated objections. As I have argued, whether Ramsey-style ESR avoids collapsing into phenomenalism/constructive empiricism depends on whether its advocates can justify a restriction of the domain over which the theoretical variables (of the Ramsey-sentence) range to unobservables. More damagingly, Worrall and Zahar’s narrow, and unargued for, construal of the concept of observational content goes against the concept’s conventional meaning. Indeed, their suggestion that commitment to empirical generalisations is enough to distinguish ESR from non-realist empiricism flies in the face of well-established ideas. At any rate, if support for structural realism comes through empirical generalisations, it is not at all clear why we need Ramsey-sentences. The only defence to this claim that Worrall and Zahar could muster is their assertion that theoretical variables in Ramsey-sentences are ‘indissolubly entangled’ with the observational terms. Yet, as I have argued, this assertion remains at best highly suggestive. The same holds for their appeal to the notions of unity and simplicity, by which they hope to show how we can choose between competing Ramsey-sentences.

Being endemic to the Ramsey-sentence formulation of ESR, these objections do not affect Russellian ESR. That leaves open the question whether this version of ESR can overcome the Newman objection. I hope to have shown that Russellian ESR is impervious to this objection. Newman unfairly attributed the simplistic claim ‘There exists a relation with a particular abstract structure’ to ESR. No epistemic structural realist accepts such a claim, which is obviously trivial. The Russellian structural realist, in particular, holds that we empirically identify logico-mathematical properties of the external world via the assumption that causally linked concrete observational structures and concrete physical structures share the same abstract structures, i.e. belong to the same isomorphism classes. To be precise, the Russellian ESR claim is this: 'There is a unique physical relation that is causally linked with this unique observational relation and the two are isomorphic'. Russellian ESR identifies, even if only up to isomorphism, a concrete physical structure by its causal role in producing perceptual relations in observers, i.e. in producing the concrete observational structures. Finally, I hope to have shown that the idea of being able to uniquely pick out physical relations or their relata rests on a myth and must be given up.

5

HISTORICAL CASE STUDY: THE CALORIC THEORY OF HEAT

1. Introduction

The 1960s marked a turning point for the scientific realism debate. Thomas Kuhn and others undermined the orthodox positivist tradition by showing that a careful study of the historical record speaks against the linear accumulation of scientific knowledge. But, as is so often the case, reaction to the admittedly naïve positivist view was disproportionate and resulted in an equally naïve and diametrically opposite view, namely that there is no significant accumulation whatsoever. Realist philosophers like Richard Boyd and Hilary Putnam were quick to reply that not everything is lost in the wake of a scientific revolution. Successive scientific theories, they claimed, preserve the theoretical relations and referents of earlier theories so long as both belong to a mature science. This attempt to rescue realism did not last long, for in the late seventies and early eighties a more sophisticated anti-realist argument appeared. The pessimistic induction argument, most often associated with Larry Laudan, is now widely considered to be one of the two main obstacles for realism (see, for example, Kitcher (1993: 136), Leplin (1997: 136) and Worrall (1982: 216); the other being the underdetermination of theories by evidence.[145] Put simply, the argument holds that since past predictively successful scientific theories have eventually been discarded, we have inductive evidence that our current theories will also be discarded one day. This landmark attack has stimulated a realist counter-strategy (see, for example, Clyde Hardin and Alexander Rosenberg (1982), Philip Kitcher (1993), Jarrett Leplin (1997), Stathis Psillos (1994) and John Worrall (1989)) that is primarily concerned to show that the historical record provides grounds for optimism. More precisely, it is argued that some theoretical components survive theory change, and that only those are responsible for any success enjoyed by the rejected theories. This strategy is now the mainstream approach for scientific realists.

[145] The pessimistic induction argument can be viewed as a constructive demonstration of underdetermination since the new theory, despite being incompatible with the old theory, entails its correct empirical consequences.

The aim of this chapter will be to address RP2 and SRP2. That is, I will investigate whether theoretical components, structural or not, survive scientific revolutions and, if so, whether they are indeed solely responsible for the success of abandoned theories. Of course, to settle this issue in a satisfactory way is an enormous task, since it would require a detailed analysis of all the relevant historical episodes. For this reason, I have chosen a more tractable problem by limiting my investigation to a single case study, namely the caloric theory of heat. This choice reflects three considerations:

A) The caloric theory appears on Laudan’s list as one of those successful theories that have been abandoned together with their central theoretical terms.

B) It is a case that has not yet been investigated by structural realists.

C) It has received significant attention in recent literature. Therefore, it will be easier to compare any structural realist claims to those of other parties in the scientific realism debate.

The chapter will unfold as follows: I will first present a historical account of the caloric theory of heat. Then, I will proceed to evaluate claims made by realists and anti-realists with regard to the caloric theory and its central theoretical terms. Subsequently, I will examine whether epistemic structural realism can make better sense of the history of the caloric. Even though my investigation and results will be restricted to this case study, whenever reasonable I will try to extrapolate more wide-ranging results about the history of science. Indeed, the final topic in this chapter will be a general assessment of the import of historical arguments in the scientific realism debate.

2. The Rise and Fall of the Caloric Theory of Heat

The Pre-Caloric Era

Chemistry and the study of heat were in their infancy prior to the eighteenth century. The Aristotelian tradition, despite its sterility, was dominant for centuries. According to this tradition, there were four elements or fundamental substances, namely earth, water, air and fire. Unsurprisingly, phenomena of heat were understood with reference to the element of fire. After all, it was common knowledge that objects could be heated by placing them on fire. The Atomists and the Epicureans had roughly the same conception, viewing heat/fire as a substance with weight.[146] Thus, in both traditions 'the materiality of heat hypothesis' was central. Needless to say, the explanations furnished under either tradition were crude and qualitative in nature.

It was only with the rise of alchemy that some limited progress was achieved. Despite the magical underpinnings of their views, the alchemists developed several techniques that contributed to the advance of chemistry. They were, for example, good at distillation and the production of concentrated acids, alcohols, and perfumes. They were also very good at metallurgy, especially on amalgams and on acid/metal reactions.

In the seventeenth century, Robert Boyle attacked the four elements of the Aristotelian tradition and popularised mechanical philosophy. His mechanical world-view took particles as elementary and explained their behaviour through forces such as gravity. Chemistry could not benefit from this world-view, however, since at the level of elementary particles there was not much one could predict about macroscopic phenomena, which constituted the sole domain of chemistry at the time. Nonetheless, some advances in the study of heat phenomena were made, for example, in the studies of adiabatic phenomena and the construction of instruments (see Robert Fox (1971: 41)).

The eighteenth century brought with it an excitement about the study of the nature of air and of combustion. The Aristotelian idea that air is one of four elements was unambiguously rejected. Instead it was conjectured that air is composed of different gases and that heating eventually turns matter into a gaseous state. Stephen Hales introduced methods for collecting and measuring the volume of gases, prompting the development of pneumatic chemistry. The experiments of Joseph Black, Henry Cavendish, Joseph Priestley, and Carl Wilhelm Scheele progressively identified various gases, each with its own set of properties. Black, for example, discovered

[146] For more on the Atomistic and Epicurean conceptions of heat, with an emphasis on Lucretius' work, see Jesús M. Montserrat and Luis Navarro (2000).

some of the properties of carbon dioxide, CO_2, and gave it the name 'fixed air'. Cavendish did the same with hydrogen and called it 'inflammable air'. Thus, one of the main achievements of this period was the idea that 'air' has various manifestations, i.e. it is not just one kind of thing or substance.

At the beginning of the eighteenth century, Georg Ernst Stahl developed the phlogiston theory of combustion. 'Phlogiston', a Greek term which means 'that which is set on fire', was conceived of as the physical manifestation of heat. The materiality of heat hypothesis thus found a new carrier in phlogiston. Phlogiston was thought to reside in all combustible objects and to be released during burning. The more heat given off by an object, the more phlogiston was taken to be contained within that object. Among the theory's main advocates was Joseph Priestley. He managed to isolate oxygen, shortly after Scheele's independent discovery, but also to recognise its central role in combustion. To explain why objects burn more vigorously in the presence of oxygen, Priestley postulated that the gas that we now call 'oxygen' was entirely devoid of phlogiston and was therefore more receptive of the phlogiston present in objects than ordinary air. He thus appropriately named it 'dephlogisticated air'.

The Caloric Theory of Heat

Lavoisier's role was pivotal in discrediting the phlogiston theory of combustion. More important for our aim, however, was Lavoisier's central role in proposing and developing the theory that in one sense replaced it, namely the caloric theory of heat.[147] As its name suggests, the theory's concerns were not confined to combustion.[148] Lavoisier's first steps towards the formulation of the theory can be traced back to 1766. However, a reasonably detailed account of his theory only appears in print about a decade later in the *Mémoires* of the French Academy of

[147] For a detailed account of the relation between Lavoisier's theory and his rejection of the phlogiston theory see Morris (1972: 16-17).

[148] Morris describes the many functions of the theory as follows: "A single theoretical framework accounts for a vast array of heat phenomena including expansion and contraction, changes of state or form, the role of heat as an agent in promoting new chemical combinations, and temperature changes in chemical reactions, especially combustion and respiration" (1972: 35). Fox concentrates on the caloric theory of gases, though he explains that "[f]or the period covered in [his] book it would be quite impossible to separate the study of gases from that of heat, because from the late 1770s to the 1850's acceptance of the static theory [of gases] generally implied belief in the imponderable, highly elastic fluid of heat, or caloric" (1971: 2).

Sciences.[149] Lavoisier coined the term 'calorique' sometime before 1784 – though it was used in print for the first time in 1787 (see Fox (1971: 6) – and he listed caloric on the first table of chemical elements.[150] The theory blossomed in the first half of the 19th century, after the death of Lavoisier, partly due to the development of more precise methods of calculation and measurement. During this period it came to dominate most of the work done in the study of heat.

Somewhat like phlogiston, the caloric is thought of as the physical manifestation of fire or heat, i.e. it is thought of as the substance responsible for phenomena of heat. It is also thought to be an imponderable, i.e. imperceptible or at least hardly perceptible, and *almost* weightless fluid. Given its imperceptibility, one may well ask, how can we make any inferences regarding its presence? The most common answer is that it is intimately related with temperature, the latter of course being a measurable quantity. More precisely, it is thought that the addition of caloric to a body raises its temperature, while its subtraction lowers it. Not all properties ascribed to the caloric were as widely accepted as the ones just mentioned. The following list, therefore, does not reflect a consensus, but rather those properties that were discussed the most at the time: (1) it is an imponderable almost weightless elastic fluid, (2) it is composed of indestructible particles, (3) its particles are mutually repulsive but attracted by ordinary particles, (4) its accumulation in a body is the sole reason for the sensation of heat, (5) its combination with, or release from, ordinary matter is responsible for changes of state, (6) it is a conserved quantity.

Independent developments and ideas facilitated the birth of the caloric theory.[151] First and foremost, the theory subscribes to the materiality of heat hypothesis, an idea that as we have seen predates the caloric theory by at least two millennia. This hypothesis was certainly prevalent all through the eighteenth century, taking centre

---

[149] For a thorough account of the development of the theory see Morris (1972).
[150] Etymologically, the term originates from the Latin 'calor' which means heat. It can, for example, be found in Lucretius' poem *De Rerum Natura*. Lavoisier, together with Guyton de Morveau, Claude Louis Berthollet, and Antoine François de Fourcroy systematised the terminology of chemistry in their *Méthode de Nomenclature Chimique*. His *Traité Élémentaire de Chimie* offered the first table of elements, many of which were to comprise the building blocks of chemistry. Lavoisier also brought about the modern notion of chemical element. Indeed, for this and other contributions he is considered one of the founders of modern chemistry.
[151] For a succinct account of which aspects of the theory were Lavoisier's innovations and which were not see Morris (1972: 30-34).

stage in the works of reputable scientists such as Willem ‘sGravesande, Wilhelm Homberg and Nicolas and Louis Lemery. The most marked theory advocating the materiality of heat, well known by Lavoisier’s time, was that of Herman Boerhaave (1732). According to this theory, fire was composed of weightless particles, whose self-repulsive property resulted in elasticity. These ideas eventually resurface in both the phlogiston and caloric theories.[152] Boerhaave’s theory, however, departed from the caloric theory in one important respect.[153] Fire particles, Boerhaave held, are in constant motion, a motion that is ultimately responsible for phenomena of heat. More precisely, the motion of fire particles caused and sustained the motion of ordinary particles, the latter being directly responsible for phenomena of heat (see Fox (1971: 12-13). Despite the popularity of this picture of heat amongst his predecessors, Lavoisier opted for a static view of caloric particles where the accrual of caloric particles alone would explain the raising of a body’s temperature – property four on the above list.

Fox, justifiably in my opinion, notes that “[t]he strength of eighteenth-century belief in the materiality of fire is undoubtedly one of the most important elements in the background to the emergence of the caloric theory of heat in the 1770s” (14). At the same time, he admits that this “does not constitute the whole story” (14). Another crucial part of the story concerns the development and eventual acceptance of theories of electricity, magnetism, and light in the time leading up to the emergence of the caloric theory. One prominent example is Benjamin Franklin’s theory of electricity. Theories like Franklin’s postulated imponderable fluids, whose properties had much in common with caloric. Aside from the conception of electricity, magnetism, and light as fluids composed of weightless particles, there was also the idea that a mutually repulsive force might dictate their behaviour.

These ideas eventually led fire-theorists to the view that fire itself is a fluid. Two important historical figures in this context were Bryan Higgins, a physician and

---

[152] According to Morris, the majority of chemists in the eighteenth century considered the self-repulsive force between fire particles on par with gravity. Apart from Boerhaave, Morris also cites Pierre Joseph Macquer as a theorist who adopted the view of fire particles having a self-repulsive property.

[153] The idea of motion generating heat was at least partly in line with the caloric theory’s main competitor in the nineteenth century, i.e. the vibratory theory of heat, to be discussed in the next subsection.

chemist, and William Cleghorn, a professor of anatomy. Though working independently, they proposed theories of heat that bear strong resemblances to Lavoisier's theory. Higgins, for example, described fire in 1775 as an elastic fluid, and claimed that its elasticity was due to repulsive forces between its particles. Cleghorn's conception of heat, found in his 1779 dissertation for the degree of MD, bears an even closer resemblance to Lavoisier's theory. He describes fire as a fluid with all the essential properties of the caloric listed above.[154] A case can thus be made that at least Cleghorn had independently invented the caloric theory at roughly the same time as Lavoisier.

The exact relationship between the works of Lavoisier, Higgins and Cleghorn is not entirely clear. Fox claims that Cleghorn was definitely ignorant of Lavoisier's and Higgins' works. But there is no indication whether Higgins and Lavoisier were aware of each other's work or of the work of Cleghorn. What seems certain is that all three had knowledge of the various fluid theories available at the time. As Fox indicates, given the prominence of fluid theories during the latter half of eighteenth century, "it is inconceivable that Cleghorn (or Lavoisier and Higgins, for that matter) would not have been thoroughly familiar with them" (16). Indeed, he points out that Lavoisier had acknowledged the influence of Boerhaave and Franklin. Fox also indicates that Cleghorn was familiar with the work of Boerhaave though he says nothing about Higgins's familiarity. All in all, we can say that the developments in fluid theories as well as Boerhaave's theory of fire facilitated the emergence of the caloric theory.

In addition to the six properties listed above, we also need to consider a few central concepts that had a lasting impact on chemistry. Though not his own invention, Lavoisier distinguished between (chemically) combined caloric, i.e. caloric found "in bodies by affinity or elective attraction, so as to form part of the substance of the body", and free caloric "which is not combined in any manner with any other body"

[154] I could not find any indication whether Cleghorn subscribed to the sixth property, viz. the conservation of fire particles. It is worth pointing out, however, that the second property, i.e. that caloric particles are indestructible, together with the assumption that there is a fixed number of caloric particles, guarantees their conservation.

(1790: 19).[155] In its combined form it was undetectable but became detectable, typically via a thermometer, when it was set free. Presumably then, whenever a body was producing an increase in a thermometer's reading, it was freeing up caloric.

Lavoisier's distinction, which first appears in print in 1772, seems to have been formulated independently of Black's similar distinction postulated about a decade earlier.[156] Black's distinction was prompted by observations that, contrary to common-sense, melting ice maintains the same temperature. To explain this, he distinguished between *latent* and *sensible* forms of heat. According to Black, when ice melts the caloric is converted, i.e. not destroyed, into a state that cannot have an effect on the thermometer, i.e. latent heat.[157] By contrast, *sensible* heat is conceived of as being able to affect the thermometer. Thus, both Black's distinction and Lavoisier's make use of the idea that one state of heat affects the thermometer while the other one does not. One difference between the two conceptions is that, contra Lavoisier's conception, Black seems, at least until 1772, not to have thought of latent heat as heat chemically combined with ordinary matter.[158]

Another important concept that was to have a lasting effect on chemistry is that of heat capacity. As he notes in his manuscripts, Black started thinking about the concept of heat capacity around 1760.[159] Black first noticed that, contrary to the mainstream view, "the quantities of heat which different kinds of matter must receive, to reduce them to an equilibrium with one another, or to raise their temperature by an equal number of degrees, are not in proportion to the quantity of matter in each" (1803: 79). In considering experiments conducted by Boerhaave, Gabriel Daniel Fahrenheit, and George Martine, Black realised that the same mass of

---

[155] Morris indicates that "[t]he idea that heat matter can exist in two distinct states, free or combined, was used by his predecessors and contemporaries in explaining various phenomena" (1972: 31). In a footnote on the same page, he cites these as being Macquer, Gabriel François Venel, and Guillaume François Rouelle. It is also interesting to note that in the posthumously published *Mémoires de Chimie,* written in the years 1792-3, Lavoisier takes the free and combined states of caloric to represent the ends of the spectrum, everything in between being represented by a mixed third state that he calls 'the *adherent* state' (see Morris (1972: 25)).

[156] According to Morris (1972: 27-8), it seems that Lavoisier's distinction was not based on Black's since there is no indication that Lavoisier was familiar with Black's work prior to his postulation of the distinction. Chang mentions also "the independent contribution of Johan Carl Wilcke" to the discovery of latent heat (forthcoming: 2).

[157] Notice that this provides tacit support for the conservation of matter (including the matter of heat).

[158] See Morris (1972: 27-28).

[159] See his posthumous (1803). Niels H. de V. Heathcote and Douglas McKie (1935) provide a comprehensive account of the discovery of the concept.

different substances require different quantities of heat for their temperature to be raised by the same number of degrees. From this he argued that different substances have different capacities for heat. More exactly, for unit masses of any two substances that have unequal capacities for heat, when starting from the same temperature and aiming to increase their temperature by the same degree, the one with the lesser capacity will require less heat. In his own words:

> We must, therefore, conclude that different bodies, although they be of the same size, or even the same weight, when they are reduced to the same temperature or degree of heat, whatever that may be, may contain very different quantities of the matter of heat; which different quantities are necessary to bring them to this level, or equilibrium, with one another (83).

In 1780, Jean Hyacinthe de Magellan coined the term ‘chaleur spécifique’, ‘specific heat’ in English, to convey the concept of heat capacity. Today we distinguish the concept of heat capacity from that of specific heat. The former is defined as the quantity of heat required to raise the temperature of a given substance by one degree. The latter is defined as the heat capacity of the substance per gram, i.e. the quantity of heat needed to raise the temperature of one gram of substance by 1°C. In some contexts, the terms ‘heat capacity’ and ‘specific heat’ are used interchangeably.

The view that caloric could be found in two forms, i.e. latent and sensible, soon became a major point of contention amongst the caloricists. Lavoisier and Pierre Simon de Laplace led those favourable to the distinction, while William Irvine, a student of Black’s in Glasgow who helped him perform experiments on latent heat, and Adair Crawford, a physician who developed Irvine’s views, led those against it. Irvine thought that there was only one state of caloric, and that the amount of caloric in a given object was the product of its absolute temperature, the relative amounts of heat present in equivalent weights of different substances at any particular temperature, and heat capacity. He explained away the state of latent heat by arguing that phenomena associated with it were merely due to variations in the heat capacity of a substance. An analogy that is often employed to make sense of Irvine’s explanation involves a bucket of water: As the bucket widens the level of water goes down and thus more water is required just to keep the water on the same level. Similarly with latent heat phenomena, when ice melts into water, there is an increase in the heat capacity leading to more heat being required just to maintain the same temperature.

Chang (forthcoming: 4) notes that although Irvine's theory of heat capacity did not remain a serious contender beyond the early 19th century, his legacy was reflected in the subsequent debates.[160] One important aspect of this legacy was his work on specifying an exact relationship between heat and temperature capable of generating quantitative, albeit inaccurate, predictions.[161] Prior to this work, quantitative predictions were virtually unheard of in theories of heat. Lavoisier, for example, claimed that the quantity of heat contained in aeriform state was *more than* that contained in liquid state and *even more than* that contained in solid state (see Morris (1972: 15, 34)).

As indicated above, another eminent figure who advocated the caloric theory was Laplace. He favoured a Newtonian foundation to the caloric theory. This was reflected in his demand that caloric particles repelled one another but were attracted by ordinary matter. A universal repulsive force, similar in spirit to Newton's law of gravitation, was conceived between the caloric particles. The following colourful description by Stephen Brush and Gerald Holton gives a few more details of what exactly was involved:

> If heat is applied to a material object, the caloric may be pictured as diffusing rapidly throughout the body and clinging in a shell or atmosphere around every one of the corpuscles. If these corpuscles (whose caloric shells repel one another) are free to move apart, as indeed they are in a gas, they will tend to disperse, and more strongly so the greater their crowding or the heat applied… If, however, a heated object is in solid or liquid form, the mutual attraction among the corpuscles themselves (considered to be a gravitational one) so predominates that the caloric atmospheres can provide mutual repulsion sufficient only for the well-known slight expansion on heating. Or at any rate, the attraction predominates until enough caloric has been supplied for eventual melting or vaporisation ([1952]2001: 235).

Under this Newtonian framework, rough explanations could now be given. For example, as Brush and Holton point out, the expansion of body with temperature could be explained as the result of the repulsive force between caloric particles. In sum, the caloric theory became closely associated with *this* Newtonian picture, an

---

[160] Chang (forthcoming) lists John Dalton, Sir John Leslie, and John Murray as notable Irvinists.
[161] For more on this see Chang (forthcoming: 3-5), Fox (1971:27-39) and Morris (1972: 13)).

association that, as we will shortly see, would mean that the latter's eventual demise would hasten the demise of the former.

The Vibratory Theory of Heat[162]

The idea that heat is due to the motion of particles can be traced to the Atomists and Epicureans. Whether it survived through the centuries or was reinvented is not clear.[163] What we do know is that it started gaining prominence again in the sixteenth century. Francis Bacon, for example, remarked that 'heat itself, its essence and its quiddity, is motion and nothing else'. Galileo, Boyle, Hooke, and Newton all agreed. Daniel Bernoulli had proposed a vibratory theory of gases in 1738, but it went almost completely unnoticed for a century or so. More popular in the eighteenth century were hybrid theories, combining elements from both material and vibratory accounts. These postulated that heat phenomena were due to particles of heat in constant motion. We have already seen an example of this type of theory in Boerhaave.

At the end of the eighteenth century, the main proponent of the vibratory theory is Sir Benjamin Thompson, who is better known as 'Count Rumford'. The central tenet of the vibratory theory is that the motion or vibration of ordinary matter particles produces heat. Yet, the exact nature of the mechanism remains unspecified. Rumford admits as much when he says "I am very far from pretending to know how, or by what means, or mechanical contrivance, that particular kind of motion in bodies, which is supposed to constitute heat, is excited, continued and propagated" (1798: 99). In an effort to justify this lack of knowledge, he cites Newton as being in an analogous situation. That is, not knowing its ultimate mechanism, did not bar Newton from articulating the law of gravity.

Rumford's sustained attack on the caloric theories lasted three decades and was often backed up by experiments. One of his most famous observations came in 1798 while overseeing the boring of cannons in Munich. Rumford noticed that the metal chips

---

[162] The vibratory theory is also referred to as the 'mechanical', 'dynamical', or 'kinetic' theory of heat. As in the case of the caloric theory, there was more than one vibratory theory.

[163] Given the pervasive propagation of ancient Greek and Roman texts throughout the history of western philosophy it seems more probable that the view of heat as motion was merely inherited, i.e. not reinvented, by philosophers of the sixteenth century.

produced in the boring had a high temperature. The explanation according to the caloric theory would have been that caloric was being squeezed out of the cannon in the boring process. That is, the chips were gaining heat from the cannon and so the latter should have been losing heat and, consequently, temperature. Against this explanation, he found that the cannon was also very hot. It thus seemed to him that more caloric was being released than could have been contained. Indeed, Rumford generalised that, in order to sustain friction, the amount of heat produced would have to be inexhaustible. He concluded:

> It is hardly necessary to add, that any thing which any *insulated* body, or system of bodies, can continue to furnish *without limitation*, cannot possibly be a *material substance*: and it appears to me extremely difficult, if not quite impossible, to form any distinct idea of any thing, capable of being excited, and communicated, in the manner the heat was excited and communicated in these experiments, except it be MOTION (1798: 99) [original emphasis].

For the caloric theory to hold, Rumford claimed, a limitless outpouring of caloric would be required. No material body is limitless, and so the cause of heat could not be material. By contrast, if heat was the result of the vibration of particles, then the heating up of both cannon and boring instrument could be more easily accommodated and explained. It was only necessary to assume that the motion was transferred from the borer to the boring instrument and from there on to the cannon.

One problem with Rumford's reasoning is that it is based on the strictly speaking false claim that heat can be produced by friction *inexhaustibly*. Obviously, friction between two or more bodies has to stop after some finite period of time due to the diminution of the bodies involved. This idealisation notwithstanding, Rumford's claim remains threatening to the caloricists, since a great deal of heat would still be produced in such experiments. The question then becomes whether the caloricists could explain the presence of such quantities of heat in a non-ad-hoc manner, a question whose answer remains unclear. Rumford's argument that, contra what a caloricist would expect, both the cannon and the boring instrument were gaining heat seems more effective. Yet, even against this argument, the claim could be made that, for some reason, caloric was squeezed out of both cannon and boring instrument. Whether a claim like this could be explained in a non-ad-hoc way is again unclear. The fact is that Rumford's two arguments were not knockdown arguments.

Unaware of the experiments performed by Rumford, Humphry Davy conducted his own experiments and reached more or less the same conclusion. According to Sir Harold Hartley, a modern historian, Davy was “anxious to decide between the rival theories, that heat consists of an elastic fluid called caloric, or is due to a peculiar motion of the particles of matter” (1971: 94). One of his most famous experiments involved the rubbing of two plates of ice. It was expected that not enough caloric was present to melt the ice yet the friction alone resulted in the production of sufficient heat. In another experiment, he melted wax in a vacuum using heat produced by the friction of a wheel rubbing against some metal. The apparatus sat on top of a block of ice, maintaining a temperature of zero degrees Celsius. He argued that no heat could enter the system and concluded that heat was the result of the vibration of ordinary matter particles.

Davy reported the results in his first publication (1799), resolutely concluding that caloric does not exist. Yet, as numerous authors have pointed out, Davy’s conclusion is a non sequitur.[164] Just like with Rumford’s experiments the evidence did not conclusively refute the caloric theory. The caloricist might not have expected the ice to melt under such circumstances, but it was hardly the kind of evidence that was irreconcilable. All that the caloricist needed to dispute was how well the experimental set up was insulated. In other words, the caloricist could always claim that caloric was leaking into the apparatus. As before, whether this claim could be explained in a non-ad-hoc way is not altogether clear.

Among the many experiments Rumford performed, some dealt with conduction and convection.[165] By conduction he understood the transmission of heat via direct contact between particles, while by convection he understood the transmission of heat via the motion of particles in a fluid. Indeed, Rumford thought that heat in air and water, as well as probably other liquids and gases, was transmitted only via convection, i.e. not via conduction. His justification was that the molecules of liquids and gases are constantly moving, and so would not be able to sustain the

[164] Brush and Holton ([1952]2001: 237). Roller (1950: 86-87) argues that Davy’s experiment of rubbing plates of ice was flawed.
[165] See Chang (forthcoming: 12-14).

transmission of heat via direct contact between molecules. Among the experiments Rumford employed to support his view, one involved the heating of the surface of a vat of water. With this experiment, he showed that there was no detectable increase of temperature below the surface. If conduction could operate in liquids then the heat on the surface would be propagated below via direct contact between the molecules.

More precise experiments revealed a slow conduction of heat in liquids.[166] These results were compatible with the caloric theory, since, for caloricists, conduction just meant the flow of caloric between molecules. In other words, there was nothing in their theory that precluded conduction in liquids. Rumford must have seen the acceptance of conduction in liquids as giving tacit support to the caloric theory. In what probably was an ad-hoc move to avoid this indirect support for the caloric theory, he argued that the phenomena presumed to be due to conduction in liquids were actually due to the radiation of heat.

Phenomena of radiant heat, however, were more of a hazard than a refuge for the vibratory theory. The reason was the perceived relationship between heat and light at the time. Many scientists in the first half of the nineteenth century thought of light and heat as qualitatively identical entities. A consequence of this conception was the view that the nature of heat depended on the nature of light, which, at the time, was generally regarded as particulate. Independent experiments carried out by William Herschel, Macedonio Melloni, and James Forbes suggested that radiant heat exhibits all the properties of light, i.e. properties such as reflection, refraction, and polarization.[167] Through the independent work of caloricists Marc-Auguste Pictet and Pierre Prévost, whose experiments and explanations on radiant heat in the 1790s attained great prominence, the caloric theory gained the upper hand in this domain of phenomena.

---

[166] Chang (forthcoming: 13) notes that the caloricist John Leslie, among others, rejected Rumford's claims.

[167] Brush and Holton indicate that, at the time, there was no distinction between radiant heat and other forms of heat and "it was therefore believed that any conclusion about the nature of radiant heat would be valid for the nature of heat in general" ([1952]2001: 238). By transitivity, since the nature of radiant heat and the nature of light do not seem all that different, it was assumed that the nature of heat and light must not be all that different.

The Demise of the Caloric Theory

The current view in the historiography of science is that Rumford's and Davy's experiments did little to overturn the caloric theory of heat.[168] Historical facts notwithstanding, one needs to consider whether the evidence presented at the turn of the nineteenth century was strong enough to refute the caloric theory or any material theory of heat for that matter. It is certainly true that the caloric theory together with its auxiliary assumptions was in conflict with the friction experiments, since it rejected the idea that heat is merely due to motion.[169] This evidence was not sufficient to refute hybrid theories, like Boerhaave's, since they could employ explanations that ultimately relied on the motion of fire particles. In other words, to the extent that the vibratory theory could explain friction phenomena hybrid theories could do so, too. More importantly, even the caloric theory's inability to square itself with such phenomena did not make a strong case for its abandonment since the vibratory theory also faced several anomalies. For example, the vibratory theory could not yet explain phenomena that involved the conservation of heat in mixtures.[170] This was partly due to the vibratory theory's undeveloped state, which meant that it was unable to gain the upper hand. Overall, we can say that, at least until after the first quarter of the nineteenth century, the reasons to replace the caloric with the vibratory theory were not sufficiently strong.

The question as to why the caloric theory was eventually abandoned has not been answered in a satisfactory way.[171] Various explanations that are neither exclusive nor exhaustive have been put forward over the years. Fox, for example, notes that the rejection of the Laplacian approach to science after 1815, which was based on Newtonian principles and advocated belief in imponderable fluids, was "a major cause of the discrediting of the caloric theory" (1971: 2). Another development that cast doubt on the caloric theory was the electrochemical theory of Jön Jacob Berzelius. His theory explained phenomena of heat that arise in chemical reactions

[168] See, for example, Fox (1971: 4) and Morris (1972: 33). The older view that saw Rumford's experiments as crucial was propagated by Tyndall (1863).
[169] It is worth reminding ourselves here of Duhem's point about the inability of testing theories in isolation. The friction experiments cannot be conclusive, i.e. crucial experiments, against the caloric theory unless we can establish the innocence of all the auxiliaries accompanying it.
[170] See Brush and Holton ([1952]2001: 238).
[171] Fox, whose historical study of the caloric is the most definitive, offers only a patchwork of reasons for the theory's demise.

as having an electrical origin. For the first time, the hegemony of the caloric theory in providing explanations for this domain of phenomena was challenged. Berzelius' theory vied with the caloric theory for the provision of explanations of phenomena involving chemical heat.

As we have seen, an altogether different explanation for the fall of the caloric theory invokes the close relationship attributed to light and heat. When the Newtonian particle theory of light, which took light as a substance, held sway, it was easier to think of heat as a substance, too. Eventually, however, the tables were turned. By the second quarter of the nineteenth century, Fresnel's successful wave theory of light emerged as the victor, replacing the particle theory. A consequence of this change was the abandonment of the view that light was a substance, thereby making it easier to espouse a non-substance theory of heat.[172] This can be seen, for example, in Sadi Carnot's posthumously published notes, where the acceptance of the wave theory of light is taken as evidence in support of the vibratory theory of heat.

Whatever the exact reasons, the outcome was certainly fatal. Fox summarises the attitude of scientists at the end of the caloric theory's life nicely:

> …the result in the 1820s was not a sudden turning towards our modern vibrational theory but a period of generally acknowledged agnosticism with regard to the nature of heat, a period that lasted until the caloric theory was finally abandoned about 1850 (3-4).

During this period, the research potential of the caloric theory continued to decline until it was relegated to a merely pedagogical role. The advent of energy conservation, supplanting heat conservation, dealt the final blow to the caloric theory. Experiments performed by James Prescott Joule confirmed the principle of energy conservation by illustrating the interconvertibility of heat and work. These experiments eventually paved the way for the vibratory theory's coming to

[172] According to Fresnel's theory, light is transmitted through vibrations in an all-pervading elastic medium, the ether. Once this story was accepted it was easier to accept the idea that radiant heat was, in a manner similar, the result of vibrations of ordinary matter particles. One needs to recognise, however, that this similarity between light waves and radiant heat could also be accommodated by a hybrid theory, by supposing that it is the fire particles that are transmitted through the ether. For an in-depth account of the relationship between theories of light and theories of heat during this period see Brush (1970).

dominance in the 1850s. By that time there was no credible resistance offered by the caloric theory.

3. Scientific Realism and the Caloric

Laudan criticises the connection that realists make between success on the one hand and approximate truth and reference on the other. That is, he criticises the inference from the explanatory and predictive success of a theory to its approximate truth and referential success. Laudan believes he has shown that there are plenty of past theories that invalidate that inference.[173] More pertinent to the historical context of this chapter, Laudan argues that the caloric theory, along with other fluid theories of the nineteenth century, is an example of a successful theory whose central theoretical concept turned out to be non-referring (1981: 26-27). In this section, I will examine the realist reactions to Laudan's claim.

Is 'Caloric' a Referential Term?

Attempts to argue that theoretical terms employed in the past can be interpreted as referring to entities posited by current scientific theories originate with Putnam (1975; 1978). These attempts go hand in hand with causal theories of reference. According to the latter, although scientific theories come and go, some theoretical terms latch onto real entities, properties, and relations by virtue of causal chains stretching back to the original dubbing of the object, regardless of whether the descriptions employed were correct.[174] As Boyd (2002) indicates, many, if not most, scientific realists now accept a causal theory of reference that incorporates descriptive elements (see, for example, Kitcher (1993), Papineau (1987), and Psillos (1999)). These hybrid accounts go by the name 'causal-descriptivism'.

Realising that the rampant disregard for descriptions can only lead to trouble, Putnam augmented the causal theory of reference with a principle of charity (a.k.a. the principle of benefit of the doubt). In short, the principle allows us to brand an old theoretical term referential if the descriptions associated with it do not diverge

[173] In fact, he construes the inference as two inferences, i.e. one from explanatory and predictive success to approximate truth, the other from explanatory and predictive success to referential success.

[174] Of course, the dubbing in these cases cannot be performed indexically, since the terms purportedly refer to unobservables. It is assumed, however, that the dubbing can be performed through the effects unobservables have.

unreasonably from those of its modern counterpart. Are realists advocating this principle charitable enough to include the caloric in their list of referential terms? Putnam does not directly answer this question. He could be understood as implying that it is not unreasonable to take ‘caloric’ as a referring term when he says, speaking also on behalf of Boyd, “we do not carry [the principle of the benefit of doubt] so far as to say that ‘phlogiston’ referred” (1978: 25).[175] Also adhering to the principle of charity but being more resolute, Hardin and Rosenberg say that “[i]f one is to draw such a line in chemistry, for example, it would most plausibly come with the publication of Lavoisier’s *Elements of Chemistry* and thus would exclude phlogiston theory as a counterexample” (1982: 612).

Many have pointed out that the approach resulting from the combination of the causal theory and the principle of charity is far-fetched and suffers from several difficulties (see, for instance, Laudan (1984), Cumminskey (1992) and Worrall (1994)).[176] I will here only briefly mention the most serious one. If we allow reference fixing in the above way, many past terms can, with a little help, qualify as referring to entities postulated by current theories. That is, if at least one description associated with a past term is partially correct by our lights, referential success or failure depends on how reasonably close people think that description is from the aggregate of descriptions associated with the current term. Needless to say, opinions vary on what is partially correct and reasonably close. As I pointed out in the previous paragraph, Putnam leaves open the issue whether ‘caloric’ refers, whereas Hardin and Rosenberg are convinced that it does refer. Indeed, even ‘phlogiston’ can be made to refer under some interpretations. Phlogiston was thought of as the cause of combustion, a role afterwards assumed by oxygen. If one takes the description ‘cause of combustion’ as reasonably close to the descriptions associated with the term ‘oxygen’ today, it could be argued that ‘phlogiston’ referred all along to the element oxygen. The problem is that it is certainly a desideratum of an adequate account of reference and scientific theory change to able to provide *unambiguous* answers to questions of referential failure or success.

---

[175] This reading of Putnam presupposes that he classified the phlogiston and caloric theories as belonging to the same science. Only thus, would we be justified to take his comments as implying that the caloric theory was mature, and, consequently, that the ‘caloric’ was a referring term.

[176] Worrall’s article is particularly relevant, for he rejects Hardin and Rosenberg’s claims in the context of defending structural realism.

Given the above specification of the properties ascribed to it by its advocates, I cannot see how one can reasonably maintain that the term ‘caloric’ refers to any currently accepted entity. Not a single one of its essential properties (see the list in subsection ‘The Caloric Theory of Heat’ above) has survived to the present. Unless we help ourselves to an unrealistically charitable understanding of continuity of reference, the caloric must be accepted as a paradigmatic case of reference failure. I do not think existing accounts of reference, including causal-descriptive ones, offer persuasive reasons why the term ‘caloric’ refers to a type of entity that we today call ‘heat energy’. The term ‘caloric’ simply does not seem to refer to anything. Indeed, as we shall soon see, some realists accept the term’s reference failure and seek other ways to protect their turf from Laudan.

Is Caloric Central to the Caloric Theory?

Hardin and Rosenberg (1982) were among the first to argue that realists need not tie the approximate truth of a theory to referential success. More recently, Psillos has given a new twist to this approach by arguing that not all instances of abandoned terms within successful, and presumably approximately true, theories should be alarming to the realist, for at least some of them are simply *not central* to those theories. He thus strives to separate referential success from a theory’s predictive and explanatory success by arguing that the latter need not entail the former. Psillos’ extensive study of caloric (see (1994); (1999: ch.6)) and his attempt to undermine Laudan’s argument makes him a prime target for this subsection.[177]

For Laudan’s argument to have any impact on the realists, holds Psillos, we must examine whether the abandoned theoretical terms were really central to the theories they are customarily associated with. If they were not central, then their eventual abandonment is inconsequential to the preservation commitments of the scientific realist, for their referential failure does not undermine the success, and presumably the truth content, their theories enjoyed. What makes a term central? A term is central, says Psillos, if it satisfies the following three conditions:[178]

[177] The first half of Psillos (1999: ch.6) is simply a shorter and revised version of Psillos (1994).

[178] See his (1999: 129). Notice that the conditions are relative to the time period when the theory under consideration was reigning. This is a contentious issue that I plan to return to later on in this section. Notice also that it is not altogether clear what Psillos means by the term ‘indispensable’.

(CT1) It appeared in a genuinely successful theory.

(CT2) Its descriptions *were* indispensable in the derivation of predictions and explanations of phenomena.

(CT3) It was thought of by the supporters of the theory as denoting a natural kind.

Caloric, argues Psillos, is not a central term, for it fails on account of CT2 and CT3. According to him, though the caloric theory is indeed successful, i.e. CT1 is satisfied, the caloric posit is neither indispensable in the derivation of predictions and explanations, nor thought of as denoting a natural kind by the main advocates of the theory.

To substantiate his claim that caloric is not a central term, Psillos presents a brief history of the transition from the caloric theory to thermodynamics. That condition CT3 is not met by the caloric, argues Psillos, is obvious when one looks at the epistemic attitude of the eminent scientists most closely associated with the caloric theory. As evidence, he cites these admittedly sceptical passages from leading scientific figures:

> It has not, therefore, been proved by any experiment that the weight of bodies is increased by their being heated, or by the presence of heat in them… It must be confessed that the afore-mentioned fact [i.e. that, contra the caloric theory, heating does not bring about an apparent or measurable increase in weight] may be stated as a strong objection against this supposition [i.e. the caloric theory] (Black (1803: 45)).[179]
>
> We will not decide at all between the two foregoing hypotheses [i.e. the caloric theory vs the vibratory theory]. Several phenomena seem favourable to the second [i.e. the vibratory theory], such as the heat produced by the friction of two solid bodies, for example; but there are others which are explained more simply by the other [i.e. the caloric theory] – perhaps they both hold at the same time (Laplace and Lavoisier (1780: 152-3)).[180]
>
> The fundamental law [i.e. that heat is a state function] which we proposed to confirm seems to us however to require new verifications in order to be placed beyond doubt. It is based on the theory of heat as it is understood today [i.e. the caloric theory], and it should be said that this foundation does not appear to be of unquestionable solidity (Carnot (1824: 46/100-101)).

[179] For more on Black's sceptical attitude, see de V. Heathcote and McKie (1935: 27-30).
[180] Psillos also quotes from Lavoisier (1790: 5).

From these passages he concludes that "the scientists of this period were not committed to the truth of the hypothesis that the cause of heat was a material substance" and that "[t]herefore, caloric was not as central a posit as, for instance, Laudan has suggested" (1999: 119).[181]

Kyle Stanford (2003) has rightly criticized Psillos for presenting a biased reading of the history of caloric.[182] To be specific, Stanford holds that the passages Psillos quotes are unrepresentative of their authors' attitudes, since they are drawn from isolated remarks made about the caloric and also the ether.[183] In the case of the caloric, he elaborates, Psillos does not take into account the fact that Black was simply exhibiting a widely shared aversion towards theorising that was prevalent amongst Scots in the eighteenth century. Even more problematic for Psillos' account, argues Stanford, is the fact that Black dismissed the vibratory theory as incoherent on the basis of his own discoveries of latent heat. Thus, Stanford concludes, insofar as Black does take an epistemic stance, it is one of endorsing the caloric conception of heat as at least the more probable of the two.

Fox would no doubt agree with this assessment given his comments on the relationship between Black and the caloric theory:

> There is certainly a danger of being misled by Black's public show of caution into underestimating the closeness of the relationship between his work and the development of the material theory of heat. Black, as I have argued, did a great deal to further the theory, however indirectly or unwittingly; and it is also hard to believe that he himself thought of heat as anything but a substance when he was arriving at and elaborating the concepts of specific and latent heat (1971: 25).[184]

The same bias, says Stanford, is found in the textual evidence Psillos cites from Lavoisier and Laplace. The point of the *Mémoire sur la Chaleur*, he claims, is to present the ice-calorimeter, an instrument measuring the quantity of heat in relation

[181] The vibratory theory of heat, says Psillos, was similarly not fully accepted. Unlike in the case of the caloric theory, Psillos here blames the insufficient development of the vibratory theory.
[182] It is important to note that Stanford does not address Psillos' claims about Carnot.
[183] In the case of the ether, Stanford concedes that Psillos correctly identifies sceptical attitudes towards particular models of the ether. He complains, however, that Psillos ignores the general conviction of scientists that there must be *some* mechanical medium through which light propagates, even though no one was wholeheartedly committed to any particular model.
[184] Fox (1971: 51) indeed makes similar remarks about Pictet, arguing that even though he had doubts about the nature of heat he nonetheless worked on the basis of the caloric theory.

to the weight of ice melted. Since the measurement techniques of the calorimeter were compatible with both the caloric and vibratory theories of heat, it is not surprising, Stanford suggests, that Laplace and Lavoisier attempted to address the widest possible audience by taking a neutral stance on the nature of heat. More convincingly, he indicates that despite their initial remarks in the *Mémoire*, the rest of the book finds Laplace and Lavoisier unequivocal in their commitment to the caloric theory.[185] This is especially true for their explanatory accounts of various phenomena. What is more, Stanford says, Lavoisier's repeated endorsements of the theory and its posit through the years, coupled with his appeal to the caloric for explanations, and his view that it is confirmed by evidence, should be enough to dispel the idea that his attitude towards it was agnostic.

Historians of science, like Fox and Morris, agree that Lavoisier was undoubtedly committed to the caloric conception of heat.[186] Morris, for instance, points out that "[a]lthough in the 1783 joint memoir on heat the authors stated they would avoid a commitment to a particular theory of heat, subsequent explanations of specific phenomena reveal Lavoisier's commitment to the concept of heat as a material substance" (1972: 30-1). Indeed, he goes on to argue that even in those texts where Lavoisier sounds more sceptical, he ultimately employs the material conception of heat in his explanations of various phenomena (31). It is only Laplace's commitment that is genuinely brought into question. Yet, as Fox (1971: 30) notes, even Laplace's scepticism eventually waned, so that by 1803 he was as committed to the caloric conception of heat as Lavoisier had been.[187]

Stanford's critical remarks about the bias in Psillos' textual evidence are reasonable by any standards. Psillos might object that the fact that scientists like Black, Lavoisier, Laplace, and Carnot had any doubts is in and of itself sufficient to undermine the significance of the caloric. It seems to me, however, that this dispute is merely a red herring. Any effort to reconstruct the epistemic attitudes of past scientists often relies on tenuous speculations about the extent to which each of them

[185] Morris agrees with this point, saying that "[i]f Lavoisier wavered in his view, he did so only in the first half of this memoir" (1972: 31n).
[186] Fox (1971: 30) lists a number of other historians who agree on Lavoisier's loyalty to the caloric, namely E.M. Lémeray, G. Bachelard, C.C. Gillispie, and J.R. Partington.
[187] Morris suggests that Lavoisier's material conception of heat may have been due to Laplace (see (1972: 9)).

was committed to a given posit. In his contribution to a symposium on Psillos' book, Redhead rightly complains, "the discussion looks not so much like philosophical analysis, but rather involves peering into the psychology and/or private notebooks to ascertain what scientists really meant by terms like 'ether' or 'phlogiston' " (op.cit: 344). Whether or not a scientific community *sees* a term as central is not important if the term is *really* indispensable for the predictions and explanations of phenomena. After all, the scientific community may well have an epistemic attitude that is inconsistent with the indispensability status of a theoretical term. For example, they may not yet realise its indispensability. Alternatively, they may think it is indispensable in cases where it is not. The main point I am trying to make here is that condition CT3 can be dropped, since it is merely parasitic on condition CT2.[188]

Before I take up the issue of Psillos' evidence for the claim that the caloric does not satisfy condition CT2, I want to make some preliminary remarks on his choice of this condition. While CT3 is a criterion specific to Psillos' defence against Laudan's arguments, CT2 conveys an idea familiar to the realist camp. In order to save scientific realism from prima facie damning historical evidence, the realists seek to drive a wedge between those parts that are responsible or even indispensable for the explanatory and predictive success of theories and those that are not. In other words, realists try to show that we should not expect all components of theories to be preserved through theory change, but only those that are genuinely supported by evidence.

It is worth mentioning, even briefly, two prominent realists who employ this strategy. Kitcher, for example, has argued in favour of drawing a similar distinction between *working posits* and *presuppositional posits*. In his own words:

> Distinguish two kinds of posits introduced within scientific practice, *working posits* (the putative referents of terms that occur in problem-solving schemata) and *presuppositional posits* (those entities that apparently have to exist if the instances of the schemata are to be true) (1993: 149).

[188] A notion might still be central to a theory, despite the fact that scientists have doubts about the theory and the notion itself. This point separates commitment to truth and centrality, a distinction overlooked by Psillos. Leplin raises a similar point arguing that Psillos confuses "the question of what entities scientists believe in with the question of what entities they needed to get predictive success" (2000: 981-2).

The reply to Laudan's argument, holds Kitcher, should be that only the presuppositional posits are suspect. The realist can see these be rejected in the history of science. According to Kitcher, the ether and the caloric were such posits, therefore their rejection should not be seen as grounds for pessimism. After all, large parts of their corresponding theories got preserved. Unfortunately, Kitcher's thoughts on the caloric are limited to these few remarks.

Devitt (1984) has criticised Laudan's arguments from a variety of angles, one of which is rooted on a distinction similar to the one proposed by Kitcher. The realist, he claims, is committed only to 'necessary' posits, thereby implying a category of unnecessary ones. Devitt does not apply this idea to any historical cases. His comments, thus, remain merely suggestive, as he himself notes: "Perhaps many of the historical cases of elimination the radicals [i.e. Laudan, Kuhn, Feyerabend, etc.] produce are of posits which scientists did not really make" (144).

Having made these initial remarks on the prevalence of CT2 in the realist literature, I return to Psillos' evidence for why he thinks that caloric does not satisfy CT2. This, Psillos claims, can be seen through an investigation of three cases where the caloric theory enjoyed success during this period. These are: 1) the laws of calorimetry, 2) Laplace's prediction of the speed of sound in air, and 3) Carnot's cycle. Let us take each one in turn.

Calorimetry is concerned with measuring changes in the amount of heat. Psillos correctly points out that the laws of calorimetry were developed independently of the caloric conception of the nature of heat. The theoretical basis for these laws consisted of two assumptions closely associated with the caloric conception, namely (1) the conservation of heat in the mixing of two substances, where no chemical reactions occur, and (2) a concept of specific heat. The basis, according to Psillos, was sufficiently minimal to maintain independence between calorimetry and the caloric theory.[189] Thus, Psillos concludes, "since calorimetric laws were independent

[189] Psillos' claim, at least with regard to the principle of the conservation of heat, seems to be vindicated by Laplace and Lavoisier's view that the principle could be upheld without commitment to any theory about the nature of heat.

of considerations about the cause of heat, they could not be used to test either of the theories of the cause of heat." (1999: 118-9).

The second case on the list, argues Psillos, is similarly independent of the caloric theory. Laplace's prediction of the speed of sound in air was both successful and novel, yet it did not rely on the hypothesis of the materiality of heat. In making this prediction, Laplace had corrected Newton's theoretical calculation, which was based on the assumption that the expansions and contractions of a gas occur isothermally, i.e. at a constant temperature. Against this assumption, Laplace suggested that the propagation of sound occurs adiabatically, i.e. without loss or gain of heat. The main point that Psillos tries to raise here is that the adiabatic conception of the transmission of sound was not dependent on the caloric conception.

The third case also supports the idea that the caloric was dispensable in predictions and explanations of phenomena, says Psillos. Before we proceed to Psillos' argument, I need to explain what a Carnot cycle involves. Carnot considers two reservoirs that are kept at constant temperatures T_1 and T_2 where $T_1 > T_2$. A tank with gas contained by a movable piston is subjected to a four-step process:

Step 1: The gas at state A with initial volume V_A is brought in contact with the hot reservoir at constant temperature T_1 where it undergoes an (infinitely slow) isothermal expansion to state B with volume V_B.

Step 2: The reservoir is removed and the gas is left to undergo an adiabatic expansion from state B to a state C with volume V_C, which results in a decrease of temperature from T_1 to T_2.

Step 3: The gas at state C is brought in contact with the cold reservoir at constant temperature T_2 where it undergoes an isothermal compression to state D with volume $V_{D.}$

Step 4: The reservoir is removed and the gas undergoes an adiabatic compression from state D to its initial state A with volume V_A, which results in an increase of temperature from T_2 to T_1.

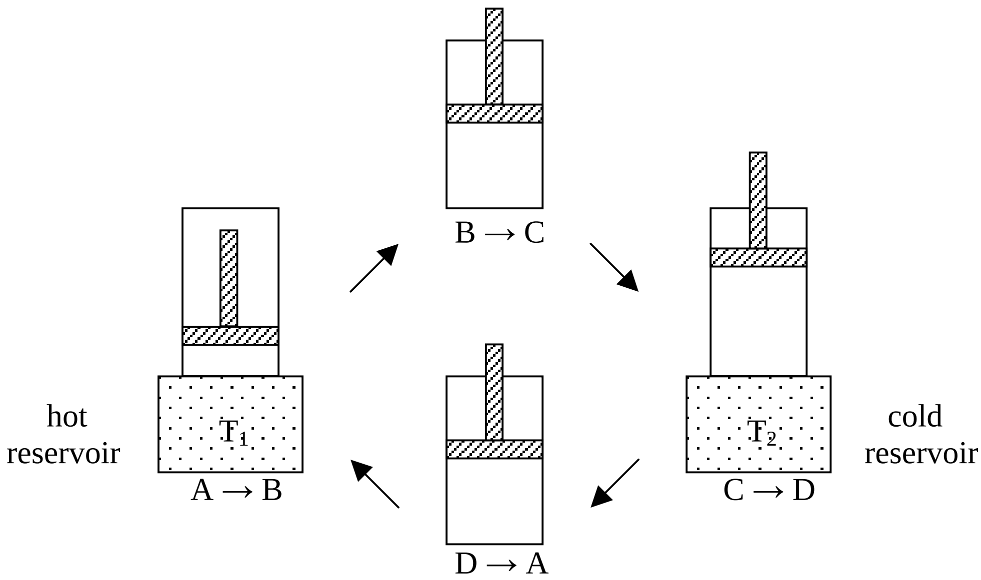

Figure 3: Carnot's Cycle.
Source: Adopted from Kondepudi and Prigogine (1998)

Note that the four-step process is repeatable, i.e. A → B → C → D → A → B → ….

Psillos indicates that although Carnot adopted the principle of the conservation of heat, a principle that was closely associated with the caloric theory but certainly not necessitating it, he purposely avoided reference to the principle in his famous exposition of what is now known as 'the Carnot cycle' (see figure three above).[190] Psillos notes, moreover, that in his description of the cycle as a four step process, Carnot "never explicitly said that the quantity of heat released by body *A* [the hot reservoir] was absorbed by body *B* [the cold reservoir]", a statement that would amount to the conservation of heat (124). Indeed, Psillos cites Clausius (1850: 133-4), who noted that Carnot's theorems did not depend on the assumption that heat is not consumed in a Carnot cycle. Put differently, they did not depend on the assumption of the conservation of heat.

Having presented his case for why the caloric is not a central term of the caloric theory, Psillos turns his attention to Laudan's claims on the relation between evidence and theory. He accuses Laudan of interpreting realists as confirmational

---

[190] The reader will notice that Psillos' comments here are inconsistent with his earlier comments. In discussing the first case, i.e. the case involving the laws of calorimetry, he claims the independence of the principle of conservation of heat from the caloric theory. Here, he emphasises their dependence in order to depict Carnot's avoidance of the principle as a general avoidance of the caloric theory.

holists, i.e. as committed to the idea that the observational evidence for a theory confirms the whole theory equally. According to Psillos, Laudan's view originates from a misreading of Boyd (1981), who claims that evidence confirms not only observational claims but also theoretical ones. Yet, as Psillos asserts, Laudan mistakenly takes this to suggest that *all* theoretical claims – presumably within a given theoretical system or framework – are equally justified by the evidence. Against Laudan, he argues that it is possible to 'localise the relations of evidential support', identifying those claims of the theory that are supported by the evidence and to what extent. More precisely, he argues that "realism requires and suggests a *differentiated attitude to*, and *differentiated degrees of belief in*, the several constituents of a successful and mature scientific theory" (1999: 126).

Naturally, those theoretical claims deemed essential to the successful prediction-making and explaining of phenomena inspire greater degrees of belief in their truth/approximate truth than those deemed inessential. Psillos takes the laws of calorimetry, Laplace's adiabatic process, and Carnot's cycle as inspiring high degrees of belief in the truth/approximate truth of the caloric theory, independently of the caloric posit, which he takes as inspiring no belief at all. He summarises this point in the following passage:

> …the laws of the caloric theory can be deemed to be approximately true independently of the reference failure of 'caloric', i.e. irrespective of the absence of a natural kind as the referent of the term 'caloric'. So, a point worth highlighting is that when the laws established by a theory turn out to be independent of assumptions involving allegedly central theoretical terms, it can still make perfect sense to talk of the approximate truth of this theory (127).

Chang (2003) argues that none of the above three cases lends any support to Psillos' conclusion. More exactly, he argues that there are two major problems affecting the three cases:

(D1) Some predictively and explanatorily successful laws, beliefs, and practices of this era were developed independently of the caloric theory. Hence, any of these that were preserved would not count as supporting the view that parts of the caloric theory were preserved.

(D2) Although certain assumptions about caloric made indispensable contributions to the successes of this era, these assumptions were not subsequently preserved.

According to Chang, the first and third of Psillos' cases suffer from D1 while the second case suffers from D2. With D1, Chang challenges Psillos' view that *certain* parts of the caloric theory were preserved. With D2, Chang challenges the more general view that theoretical parts crucial to the successes of a theory always get retained through scientific revolutions.

Let us look more closely at Chang's analysis of each of Psillos' three cases. The first case, Chang argues, is susceptible to objection D1. He agrees with Psillos that the laws of calorimetry are independent of the caloric conception of heat, but reaches a different conclusion from him. To remind the reader, Psillos concludes that the independence is evidence that the caloric theory properly construed, i.e. devoid of the metaphysical statements about the nature of heat but encompassing such things as the laws of calorimetry, deserves to be credited with the predictive success drawn from the laws of calorimetry, and hence deserves to be considered approximately true. Contra Psillos, Chang argues that the laws of calorimetry cannot be seen as evidence for the preservation of the caloric theory of heat precisely because they are independent of it.

The second case, claims Chang, is susceptible to objection D2. In the years after Laplace's prediction, Poisson derived a general law of adiabatic expansion of gases that Laplacians then used to underpin Laplace's original prediction. Chang points out that Poisson's derivation was carried out by appeal to ontological assumptions central to the caloric theory, namely (a) that the caloric is a discrete fluid, and (b) that it consists of point-like particles that repel one another but attract particles of ordinary matter.[191] Despite the eventual abandonment of these ontological assumptions, Chang argues, they were indispensable (in Psillos' somewhat vague sense of the term) in the derivation of the correct law of adiabatic processes. According to Chang, we have a counterexample to the link between indispensability

[191] In a perplexing manoeuvre, Psillos (1999: 120-1) brings up the issue of Poisson's derivation and its dependence on assumptions about the caloric but does not attempt to square this historical fact with his own account.

and success on the one hand, and preservation and approximate truth on the other. That is, a theoretical component may have been essential for the production of some successful prediction without being preserved or approximately true.

Still in the province of the second case, Chang argues that the ‘explanatory rationale’ behind Laplace’s correction of Newton’s theoretical value for the speed of sound in air rests on assumptions about caloric. More precisely, he argues that Laplace understood adiabatic heating as the result of a mechanical compression leading to the release of caloric. Since at the time no other theory offered a plausible explanation for adiabatic heating, says Chang, it is reasonable to suppose that the caloric-based explanation offered by Laplace was indispensable (again in Psillos’ vague sense of the term) for the success enjoyed by Laplace’s prediction of the speed of sound in air. Just like above, this then serves as a counterexample to the link between indispensability and success on the one hand, and preservation and approximate truth on the other.

Psillos’s third case, Chang argues, also suffers from D1. The first thing to note, he claims, is that Carnot’s work was not central to the caloric tradition. His work came during the twilight of the caloric era, and most of the leading caloricists paid little attention to it. Indeed, according to Chang, Carnot had already forsaken Laplacian microphysics and was working within a ‘macroscopic-phenomenalistic’ framework. This was evident, he claims, in Carnot’s view of the adiabatic law as an empirical regularity, devoid of any reference to Laplacian microphysics.[192] Where Carnot did employ the caloric theory, argues Chang, the results he derived from it were subsequently abandoned or at least considerably revised. As a case in point, he cites Lord Kelvin’s revision of Carnot’s theory, replacing heat conservation with energy conservation. In addition, he cites the abandonment of the picture of the production of mechanical work in a heat engine, where caloric merely gets redistributed amongst the engine’s parts. What got preserved, Chang claims, were elements independent of the caloric theory; elements like the principle of maximum efficiency, which anticipated the second law of thermodynamics.

---

[192] For an exposition of the adiabatic law see section four below.

Three issues arise from the above discussion. The first concerns the independence of a successful theoretical component from a given theory. Though Psillos and Chang agree that the first and third cases are examples of components independent from the caloric, their claims about the extent of this independence are quite different. Psillos takes the independence to be between the laws and the caloric conception of heat as a material substance. Chang, on the contrary, takes the independence to be between these components and the caloric theory as a whole. In effect, they disagree on what the caloric theory really was. They thus draw different conclusions. Psillos presents these cases as supporting the view that an essential part of the caloric theory was approximately true, despite the caloric's referential failure. By contrast, Chang thinks that these cases do not support the caloric theory's approximate truth, presumably because their total independence from the theory means that the theory cannot reap the benefits of their success.

In my opinion, this dispute misses what is of value. No matter whether one chooses to classify the laws of calorimetry and the Carnot cycle as components of the caloric theory, the fact of the matter is that these components have survived into the twenty-first century. That in itself lends some credence to the view that scientific knowledge is to a certain extent a cumulative enterprise.[193] Of course, it all depends on what kind of epistemological status these components have. If they are merely phenomenological in nature, then this does not seem to be of any help to the realist. I will return to this issue in the next section.

The next issue concerns Psillos' second case, namely Laplace's adiabatic conception of the transmission of sound in air. While Psillos wants to maintain the independence of this conception from the caloric conception of heat, Chang wants to show that the conception was very much embedded within the caloric tradition. It is certainly undeniable that (i) Laplace's explanations of the adiabatic propagation of sound and (ii) Poisson's derivation of the general law of adiabatic processes were given on the basis of the caloric conception of heat. Chang then seems to win the argument here.

---

[193] Naturally, the fact that the laws of calorimetry and Carnot's cycle have been preserved is not sufficient to establish that *all* or even *most* successful posits get preserved, but at least they are a step in that direction.

Let us consider these two cases in turn. Allegedly following Psillos' definition of a successful theory or component, as one that gives better explanations when compared to its contemporaries, Chang lists a string of explanations that were, according to him, the real successes of the caloric theory.[194] The most important of these were "the flow of heat toward equilibrium, the expansion of matter by heating, latent heat in changes of state, the elasticity of gases and the fluidity of liquids, the heat released and absorbed in chemical reactions, combustion, the radiation of heat, and the gas laws" (907). These explanations, says Chang, were given on the basis of certain 'essentially contributing' assumptions of the caloric theory. To cite a few: heat is a self-repulsive substance, temperature is simply the density of the caloric, caloric is a chemical substance, etc. Chang makes much of the fact that these assumptions were later abandoned. He suggests this supports the view that even components that are responsible for the success of a theory can be, and often are, in due course discarded.

Do we want to call the explanations present on Chang's list 'the real successes of the caloric theory', just because no alternative explanations were as 'plausible' at the time? All they had going for them was prima facie plausibility as explanations. But even that evaporated in time. It is true that we can call these 'the real successes of the caloric theory' under Psillos' criterion of success, but it is also true that Psillos' account is not adequate. For a scientific statement to be successful it does not suffice to have some plausibility as an explanation. We must demand that these explanations have their own predictive merits, i.e. that they are independently confirmed.[195] Here I side with Psillos. The three cases he considers seem to be the best examples of successful components of the caloric *era*, for they have their own predictive merits.[196] Incidentally, Chang never explicitly denies that Psillos' three examples are real successes of the caloric *era*, though he contests that the first and third examples are successes of the caloric *theory*.[197]

---

[194] I say 'allegedly', for I have not found the relevant passage where Psillos makes such a claim.
[195] I would say that a component has independent confirmation if: (i) it directly contributes to the calculations of values that can then be checked against testing and (ii) the testing involves varying the background conditions and the auxiliaries.
[196] This will become clear in the next section.
[197] By 'caloric era', I mean roughly the chronological period during which the caloric theory was popular.

Although not a view espoused by either Psillos or Chang, it is worth noting that the fact that a law is derivable on the basis of certain premises, does not entail that the success of the law must carry over to the premises. If evidence *e* confirms a hypothesis *H*, it does not necessarily follow that it confirms any theory that entails *H*. I am here thinking of Poisson's derivation of the adiabatic law. Even though there is confirmatory evidence for the adiabatic law, that evidence need not confirm the premises Poisson used to derive it. A similar point can be made without resorting to confirmation theory but simply by way of deductive logic. We have all been taught in elementary logic that a valid argument with a true conclusion need not have a single true premise. In other words, even if Poisson had validly inferred the law, as Chang maintains, the premises employed in the derivation need not be true. Indeed, given their lack of independent confirmation we now think that the premises under consideration are false.

Whether a set of sentences partakes in the confirmation obtained by any of its entailments depends, among other things, on whether that set can be independently confirmed. The premises employed by Poisson in the derivation of the adiabatic gas law, i.e. that the caloric is a discrete fluid and that it consists of point-like particles that repel one another but attract particles of ordinary matter, had no independent confirmation.[198] Today the adiabatic law is derived from different premises. It is derived from the first law of thermodynamics, i.e. that energy is conserved, plus certain other assumptions. The first law, unlike the caloric assumptions, has the benefit of independent confirmation. That, of course, is no guarantee of the truth/approximate truth or even preservation of the first law, but it is certainly good reason to maintain its superiority over the caloric assumptions.

The third issue regards Psillos' requirement of indispensability, as it appears in CT2. Psillos formulates CT2 so that the indispensability of the descriptions attributed to a term are relative to the period when the theory in question reigns. That is, these descriptions must be indispensable to the scientists of that era, though, presumably, they can become dispensable for a later one. Chang shows that the caloric posit and its properties were indispensable, if we follow Psillos' formulation, to Poisson's

[198] Indeed, post-caloric theoreticians of heat would argue for something stronger, namely that these premises have been disconfirmed by evidence.

derivation of the adiabatic law during the era in question, i.e. in the early nineteenth century.

Indispensability must surely be something fixed by the relationship between the theory, the relevant auxiliaries, and the evidence. It must not be something dependent on the whims of scientists of a particular era, as Psillos maintains. Had Psillos opted for a period and scientist-free criterion of indispensability, he would have been able to distinguish between the original and modern derivations of Poisson's law. He could thus reply to Chang that there is more support for the modern premises than for the original ones, since the modern ones enjoy considerable independent confirmation.

4. Structural Realism and Caloric

So much for the history of the caloric theory and the various attempts to reconcile it with scientific realism. It is now time to turn to epistemic structural realism in order to investigate whether it can offer a more plausible account of the components that were responsible for the successes of the caloric *era*.

Let us consider those components that had success. As we have seen, they can, without gross inaccuracies, be identified with Psillos' three cases. To remind the reader these are: 1) the laws of calorimetry, 2) Laplace's prediction of the speed of sound in air, and 3) Carnot's cycle. I presume that by the first one of these, Psillos (but also Chang) means the following two principles:

(L1) $Q = m \cdot c_p \cdot (T_f - T_i)$

(L2) $Q_a = - Q_b$

The first principle states that the heat change Q in a substance that does not undergo any other, especially chemical, reactions will be equal to the product of the mass m, the specific heat at constant pressure c_p, and final T_f minus initial T_i temperature of the substance. The second principle simply states that, on the assumption that the calorimeter is a closed system, an obvious idealisation, the heat lost by object *a* must be gained by object *b*.[199]

Laplace's prediction of the speed of sound in air, i.e. the second case on Psillos' list, was derived from the following formula:

(L3) $v^2 = (c_p/c_v)\, dP/d\rho$

where v is the velocity of sound, c_p the specific heat at constant pressure, c_v the specific heat at constant volume, P the pressure, and ρ the density of air. The general law employed to buttress Laplace's derivation is Poisson's adiabatic gas law typically written as:

(L4) $P \cdot V^\gamma = \text{constant}$

It states that for a fixed mass of gas, which is thermally insulated, the product of its pressure P and volume V will be constant; $\gamma = c_p/c_v$ is the ratio of the two specific heats, i.e. specific heat at constant pressure c_p and specific heat at constant volume c_v, of a gas at a given temperature.

Finally, Carnot's cycle, Psillos' third case, gives rise to what is hailed as 'Carnot's principle' or 'the principle of maximum efficiency':[200]

(L5) $\eta = 1 - T_{low}/T_{high}$

This states that a heat engine operating between two reservoirs at different temperatures, where T_{low} is the low temperature and T_{high} is the high temperature, will have a maximum efficiency η, i.e. given a certain input of heat there is a limit on how much of that heat can be converted into work. Indeed, it states that no actual engine can be one hundred percent, i.e. perfectly, efficient. That is achievable only

---

[199] This statement is idealised in at least two ways: 1) a calorimeter is not a closed system, and 2) some heat will be converted to work in the process of transfer. See below for more on the relationship between heat and work.

[200] The mathematical expression of maximum efficiency is due to Emile Clapeyron's mathematization and reformulation of Carnot's views.

by an *ideal* Carnot engine. The principle of maximum efficiency is the cornerstone of the second law of thermodynamics.

These equations may not all have been postulated within the caloric tradition but they have certainly survived through the scientific revolutions of the nineteenth and twentieth century. They can now be found in modern textbooks on thermodynamics (see, for example, Kondepudi and Prigogine (1998)). What can the structural realist make of them?

Recall that Worrall argues that the preservation of mathematical equations, as opposed to the ontological interpretations underlying them, is to be expected from the structural realist perspective.[201] His case study, as we have seen, involves the survival of Fresnel's equations into Maxwell's theory, ontologically reinterpreted to avoid reference to the elastic-solid ether. The obvious way to extend Worrall's argument here is to simply say that the above equations should be seen as relations devoid of any ontological interpretation. In other words, Worrall would say that the caloricists got something right, namely the structure of some heat processes.[202] Given that these structures have been preserved through two hundred years of theory upheaval, whereas the ontological assumptions about the nature of heat associated with them have not, they provide evidence for structural realism.

The Phenomenological Character of L1-L5

In a certain sense to be specified shortly, L1-L5 are phenomenological laws.[203] All the terms appearing in L1-L5 denote *broadly construed* observable quantities. By this, I do not have in mind that the terms' whole meaning can be captured observationally. Rather, what I have in mind is that we should only be concerned with the observational dimension of their meaning, since that is the only dimension

---

[201] It is worth pointing out that McMullin has expressed views that are broadly sympathetic to the historically motivated version of structural realism, though there is no indication that he accepts the Ramsey-sentence approach. In a discussion on how to overcome pessimistic induction he says: "I would argue that these [i.e. ethers and fluids] were often, though not always, interpretive additions, that is, attempts to specify what 'underlay' the equations of the scientist in a way which the equations (as we now see) did not really sanction" (1984: 17).

[202] Notice, I don't say 'there is something in the caloric *theory*' that was right. I am concentrating on the fact that some successful structural components made their appearance during that *era*, not on whether or not these components were indispensable parts of the caloric theory.

[203] Chang (2003) also stresses the phenomenological character of the principles of calorimetry, i.e. L1 and L2. In conversation, he opposed only the phenomenological character of L5.

that is testable. Each term is tied to one or more observational tests, which fully and completely determine its numerical value. More precisely, the values of these terms are determined either through direct instrument measurements or through a calculation that relies on such measurements.[204] For example, measuring mass and temperature typically involves taking direct measurements from instruments like the triple beam balance and the thermometer. To determine heat change ΔQ in a substance requires first measuring mass, specific heat at constant pressure, and temperature, and second calculating a value from these. In either case, reading the results of measurements is an act of observation, and in this sense the relevant terms should be thought of as broadly construed observational. Consequently, L1-L5 should be thought of as phenomenological laws.

That the terms in L1-L5 denote observable quantities can be demonstrated by looking at the units employed in measuring them. In the International System of Units (SI), mass, thermodynamic temperature, length, time, electric current, amount of substance, and luminous intensity make up the *base quantities*. These are measured in *base units*, namely the meter, the kilogram, the second, the ampere, the kelvin, the mole, and the candela respectively. The base units are the elementary units upon which all other units are defined. Arguably, the base units are phenomenological in character, for they are fixed by purely empirical means, i.e. without reference to theory. For example, the kilogram is equal to the mass of the platinum-iridium cylinder kept at the International Bureau of Weights and Measures at Sèrves, France. That cylinder is, in fact, the international prototype of the kilogram.[205]

My claim that L1-L5 are phenomenological laws is endorsed by the fact that each of the quantities in L1-L5 is either one of the SI base units or defined in terms of these. To be precise, the five equations identify seven quantities, namely velocity, heat, specific heat, pressure, density, volume, and maximum efficiency, that are defined in

---

[204] One may complain that under this definition no term would count as theoretical/unobservable. This complaint misses the point however, since, as I just said, I do not think the whole meaning of these terms can be observationally captured.

[205] The other six base units are similarly fixed by purely empirical means. Admittedly, the empirical fixing of some of the other base units is more complicated, for it does not rely on a physical prototype.

terms of the first four base units, mass, temperature, length, and time.[206] Even if one doubts the phenomenological character of some of the base units, these four seem the most secure.

It is insufficient to state that a quantity is observable just because it is measurable in units that are eventually defined on the base units. If that were true, the critic would complain that many quantities would qualify as observable merely by *stipulating* that they are measurable in terms of the base units. For example, according to Fresnel's theory of light, the distance between two particles of the elastic solid ether under certain well-defined conditions is measurable in base units – essentially in meters. Yet, we consider this quantity as fictional, not observable.

One must simply concede that science can make, and indeed has made, mistakes about what is observable or measurable. Peter Kosso (1998) makes precisely this point with respect to the Michelson-Morley experiment. The experiment was conducted because it was assumed that the ether was observable or measurable in some way. The fact that the expected effects of the ether were not observed is not at all surprising to us, because we now believe that the ether is a fictional posit. The moral of the story is that, at best, we should remain agnostic with respect to quantities that are said to be measurable, yet have not actually been measured. Indeed, not only must they be measured with a fair degree of accuracy and consistency, but also the values obtained must subsequently be rigorously tested. One key way to test them concerns the successful application of a variety of equations that rely on these quantities to predict phenomena. Since the quantities in L1-L5 meet these criteria, it is reasonable to suppose that they are indeed observable.

After that small digression, it is worth considering how the seven quantities at hand are defined in SI:[207] 1) *Velocity* is measured in terms of meters per second. To give a

---

[206] Strictly speaking, it is thermodynamic temperature measured in the Kelvin unit that is taken to be the most basic. However, it is interdefinable with other units of temperature such as the Celsius and the Fahrenheit, which, under SI, are taken to be derived units. Also, I have left γ, the ratio that represents the two specific heats of a substance, out of this list simply because it expresses a relation between items on the list, i.e. two types of specific heat, and hence is defined in terms of them.
[207] To pre-empt any objections similar to those directed at the Logical Positivists, it is not my intention here to present these as *exhaustive definitions* for these quantities, but rather to show that the value of the terms denoting these quantities can be fixed using measurements.

value to (average) velocity, all one needs to do is measure the distance covered and the time elapsed. 2) The units with which we measure the quantity of *heat*, namely the joule, the calorie and the BTU, are interdefinable. Thus, I need only state the relationship expressed by one here: One British Thermal Unit, or BTU, is the quantity of heat required to raise the temperature of one pound of water by one degree Fahrenheit. As before, the quantities in this definition, i.e. temperature and mass are directly measurable. 3) Similarly with *specific heat*, measured as a combination of units of heat, mass, and temperature. One combination of these units sees specific heat as the quantity of heat in calories required to raise the temperature of one gram of some substance by one degree Celsius. The calorie, as I just mentioned, is interdefinable with the BTU. The other two quantities are simply temperature and mass. 4) *Pressure* is measured in the pascal unit, defined as one newton per square meter. The newton is itself defined in more basic units as the force required to produce an acceleration of one meter per second per second to a mass of one kilogram. Acceleration is ultimately reducible to a relation between the measurable quantities of time, direction and length. Notice that all of the terms in the definition of the newton are directly measurable quantities, namely length, time, and mass. 5) The definition is simpler with the units of (mass) *density*, being specified in kilogram per cubic meter. 6) The same level of simplicity follows *volume*. It is measured in units of cubic meter. 7) Finally, *maximum efficiency* is measured in terms of the joule, which, as I have just pointed out, is interdefinable with the BTU.

If the testable aspect of L1-L5 brings out their phenomenogical character, as it is here suggested, how can any version of realism gain support from their preservation? More pertinently, where does this leave the structural realist? Ramsey-style ESR does not seem able to benefit from such laws. Ramseyfication takes theoretical terms and replaces them with variables, existentially quantifying over them. The above equations, as we have construed them here, presumably have no theoretical terms but only observational ones. Hence, there is nothing that needs Ramseyfying. This does not mean that L1-L5 count against Ramsey-style ESR. It does mean, however, that they do not lend any support to it since their preservation seems to be evidence for the accumulation of observational/phenomenal/empirical structure, not of any theoretical structure. Indeed, it is this kind of evidence that van Fraassen uses to support SRP4, namely that there is continuity of structure through theory change, but only of the structure of phenomena not of the structure of unobservables.

One possible reply for the advocate of Ramsey-style ESR would be to dispute the claim that *all* of the above quantities are observable. Obviously, Ramseyfication is no longer redundant once some of the above terms are thought of as theoretical. Another possible reply would be to argue that, if there is no theory, there's no question of theory preservation arising. Yet another possible reply would be to say that it is not the equations that must be Ramseyfied but the whole theory. After all, since the caloric theory makes claims about theoretical entities, for instance by employing theoretical terms like 'caloric', there is something to be Ramseyfied.

How do L1-L5 fit with the Russellian approach to ESR? First, since ESR à la Russell is not expressed via the Ramsey-sentence approach, it avoids the problem of redundancy. Second, far from posing a problem, the phenomenological character of L1-L5 lends credence to it. According to Russell-style ESR, the rational reconstruction of scientific knowledge starts with concrete observational/phenomenal structures. As I pointed out at the beginning of chapter two, laws and other categories of scientific statements can be understood as mathematical structures. Thus, L1-L5 specify phenomenological structures of the kind needed to advance Russell's programme. The next step is to extract the abstract from the phenomenological structures. The last and final step involves an appeal to H-W and MR in order to establish the claim that the physical causes of the phenomena have the same abstract structure as the phenomena.

As an illustration of how the Russellian treatment functions, I will here consider L1 in some more detail. L1 is a concrete observational structure that can be written down as follows: $S_O = (U_O, R_O = (m(x_i, t_i) = m_i, c_p(x_i, t_i) = c_i, T(x_i, t_i) = s_i, Q(m_i, c_i, s_i) = q_i)$, where U_O is the (finite) domain of observable objects and R_O is a class of functions defined on that domain.[208] Function $Q(m_i, c_i, s_i) = q_i$ represents L1. Now, according to the Russellian picture, from S_O we can extract abstract structure S_A =

[208] A function is a special kind of relation. They can thus be represented by sets of ordered n-tuples. For example, a function $f(x)=y$ is a set f of ordered pairs such that whenever $<x,y> \in f$ and $<x, z> \in f$ then y=z. In the above example, I employ functions instead of set-theoretically specified relations for the sake of expediency.

(U_A, R_A), where U_A is the class of all sets equinumerous to U_O, and R_A is the class of all relations isomorphic to the relations in R_O. In other words, S_A is an isomorphism class that contains S_O as one of its members. By appeal to H-W and MR, we infer that there is a unique concrete physical structure whose members stand in causal relations with the members of S_O. We can call this concrete physical structure 'S_P'. The claim is that S_P is isomorphic to S_O. Trivially, S_P is another member of the isomorphism class S_A. The complete Russellian claim then is that there is one concrete physical structure S_P, which is causally responsible for concrete observational structure S_O and can only be specified up to isomorphism, i.e. we can only specify S_A.

The Russellian structural realist can find hope in the preservation of phenomenological laws, since she assumes that relations between observables/phenomena are isomorphic to relations between unobservables. For Russell, if you recall, the unobservables are the external world causes of the observables. Thus, the phenomenological laws signify underlying relations between unobservables that are isomorphic to the relations between the phenomena. Provided one accepts the H-W and MR principles, this result is surely realist in spirit, for it says something about the external world. Of course, the traditional way of understanding the scientific realist debate dictates that phenomenological laws are not supportive of realism in any way. However, the unorthodox way in which the Russellian sets up realism means that this particular type of realism derives its force from phenomenological laws.

5. Structural Realism and the History of Science

Is Everything Structural Preserved?

The answer to this question must be obvious by now. Not all that is structural will be preserved. Examples of structural components that never made it past a scientific revolution are plentiful. In the historical context examined in this chapter, one can cite the following: (a) that the quantity of heat absorbed or freed by a given body is a state function of its properties of pressure, volume, and temperature, (b) the Irvinist equation for the determination of the absolute zero point of temperature,[209] (c) that specific heat is constant under temperature change, etc. These statements are structural, for they state relationships between different quantities, yet they have been discarded as false.

This fact is by no stretch of the imagination lethal to the epistemic structural realist. The preservation claim must simply be qualified to reflect a more realistic picture of the development of science. One suggestion is the following:

(MSSS) Not all structures may survive, but *most predictively successful* elements that do survive are structural.

The ESR-ist can concede that some structures, especially those with little or no predictive power, are left behind in the wake of a scientific revolution. Moreover, the ESR-ist can concede this and still claim that most elements that enjoy genuine predictive success and survive scientific revolutions are structural.[210] In the previous subsection, I indicated that those components that are generally acknowledged as the caloric era's lasting contributions to our knowledge of heat have all been structural.

Does Structure Always Survive Intact?

Worrall's study of Fresnel's theory of light and my study of the caloric theory of heat exemplify cases where structural components survive intact – into classical electromagnetism and thermodynamics respectively. Although both of these theories are still in use today, they co-exist with theories that are considered to be even closer to the truth, namely quantum electrodynamics and statistical mechanics respectively. The relation between the two sets of theories, however, is not as straightforward as that between classical electromagnetism and Fresnel's theory of light or thermodynamics and the caloric theory of heat. Only under certain limits, can one recover approximate versions of classical electromagnetism from quantum electrodynamics and thermodynamics from statistical mechanics.

---

[209] See section six below. For a more thorough account see Chang (forthcoming).
[210] That not all surviving components need be structural will be made clear in the next subsection. It will be pointed out that preservation sometimes happens for reasons other than truth/approximate truth.

Critics of ESR complain that, more often than not, old equations reappear only as limiting cases of new equations. As Worrall points out, this fact can be accommodated by structural realism when appeal is made to the correspondence principle. According to Heinz Post's well-received formulation of the correspondence principle, "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" (1993: 16).[211] What Post has in mind is a correspondence between mathematical structures. Indeed, given that the principle applies solely at the level of mathematics, Worrall notes correctly that its applicability "is not evidence for full-blown realism – but, instead, only for structural realism" (1996: 161).

The challenge is to spell out exactly what this correspondence involves while at the same time avoiding a trivialisation of the relationship between old and new structures. Without entangling myself into a lengthy discussion of the correspondence principle, I would like to offer one familiar type of correspondence that can be spelt out more precisely and that finds support from the historical record. Here's my own rendition of it:

(NC): A structure *L´* that has a predecessor *L* becomes isomorphic to, or approximates, *L* when a parameter in *L´* is neutralised.

The neutralisation of a 'parameter in a structure' just means that the relevant relation in the structure is redefined from an *n*-tuple to an (*n*-1)-tuple, since one of the terms is neutralised. Typically, the neutralisation process in equations involves one of two things: (1) setting the parameter to zero (as, for example, in cases where the value of the parameter is to be added to some other values) or (2) setting the parameter to one

[211] Though the correspondence principle is customarily attributed to Niels Bohr, its spirit has been around at least since Newton pronounced that he had derived his theory from Kepler's laws (see Zahar (2001: 118)).

A recent *festschrift* on Post's contributions to our understanding of the correspondence principle reveals diverse manifestations of these correspondence relations (see French and Kamminga (1993)). Stephan Hartmann (2002) summarises these and argues in favour of a pluralistic view. He raises doubts about the existence of a universal correspondence principle. It is worth noting that nothing in ESR stipulates the need for a universal correspondence principle.

(as, for example, in cases where the value of the parameter is to be multiplied by some other values).

Three examples that feature centrally in their respective theories will authenticate the validity of NC. Take momentum in the special theory of relativity (STR) and in classical mechanics (CM) first.[212] In STR we express momentum with the formula $p = m_0 v \,/\, \sqrt{1 - v^2/c^2}$. In CM we express it as p = $m_0 v$. In the limit, as v $\rightarrow$ 0, the denominator of the STR equation for momentum is neutralised and the classical formula p = $m_0 v$ is recovered.

The second example concerns the relation between Minkowskian and Galilean space-times. Minkowski space-time allows for a non-singular metric which is represented by the matrix diagonal (1, - 1/c^2, - 1/c^2, - 1/c^2), where *c* is the speed of light in a vacuum. Since the metric is non-singular, the above matrix diagonal has an inverse, namely (1, -c^2, -c^2, -c^2). If we let c=∞, that makes 1/c = 0, and the metric becomes singular (1, 0, 0, 0), allowing no inverse. By doing so, relativity of simultaneity disappears and we recover Galilean space-time. As in the case above, neutralising the term 1/c allows the recovery of the old structure.

The third example concerns the relation between the Poisson bracket formulation of classical mechanics and the Moyal bracket formulation of quantum mechanics. The latter introduces non-commutative multiplication for phase space functions. If we set Planck’s constant *h* to zero, thereby neutralising it, commutativity is recovered and so is the Poisson bracket formulation of classical mechanics.

That some old equations can be recovered whole or approximately whole simply by neutralising one of (often many) parameters of new equations cannot be dismissed as mere mathematical trickery. NC-correspondence is quite difficult to meet. To test this, we can use a computer program that generates random pairs of equations. Because a great many different equations are possible, the odds of getting a pair that N-corresponds are very small. This result holds even if we allow for the most liberal

---

[212] This example is taken from Hartmann (2002). The use I make of this example varies from Hartmann’s.

understanding of approximation acceptable in science. In other words, N-correspondence is not only supported by historical evidence but it is also far from trivially satisfiable.

One way to justify the legitimacy of NC as a correspondence principle is to think of neutralisation as a process of idealisation. That is, we can think of *L* as an idealised version of *L´*. When we neutralise a term in *L´* we sacrifice a certain degree of predictive accuracy and, by extension, concreteness. Conversely, the move from *L* to *L´* can be thought of as one of de-idealisation. This view is common amongst scientists. It is also shared by some philosophers of science, notably those of the Poznań School (see, for example, Krajewski (1977)).

To repeat, I am not here suggesting that all cases of correspondence take this form. Rather, given that NC is not easy to come-by, the existence of structures that exhibit NC-correspondence lends some credence to the view that some robust structural continuity exists even where the surviving structures are not intact.

In light of the necessity to employ correspondence principles, I suggest that we modify MSSS thus:

(MSSS´) Not all structures may survive, but *most predictively successful* elements that do survive, either intact or suitably modified (for example, according to NC), are structural. Indeed, most, if not all, predictively successful structures survive.[213]

A measure of vagueness in MSSS´ is unavoidable. I just indicated one way in which new structure can, with reasonable modification, be made to correspond to old structure. That, of course, is not sufficient to fix the meaning of the clause 'suitably modified'. More needs to be done in order to show the overall history of science bears the mark of MSSS´. My intuition is that once a theory gets on the road to mathematization and starts producing accurate and repeatable results, it becomes less

[213] In other words, not only are most predictively successful elements that survive structures, but also most predictively successful structures survive.

likely that non-structural components survive and more likely that those components that do survive, in some form or other, are structural.[214]

The Criterion of the Maturity of Science

Realists who seek to establish some continuity between past and present theories rely on a criterion of the maturity of science. The rationale is to relegate those theories and terms that feed Laudan's cause to immature science, thereby avoiding the need to justify why these theories and terms have been abandoned. This ploy is true even of some structural realists. Worrall's solution to Laudan's challenge incorporates such a criterion. More precisely, his criterion requires that a science be branded 'mature' only when it predicts *novel* types of phenomena. Worrall makes use of this criterion to distinguish between those sciences from which we should expect structural components to be preserved and those from which we should not. He is thus indirectly telling us that structural components from immature theories will, most probably, not be preserved.

This seems like overkill to me. Theories that arise in sciences deemed immature on Worrall's account may still contain structural components that are preserved for good reasons, i.e. for their ability to make accurate predictions, though not obviously for their ability to predict novel types of phenomena. That is, a structure may be able to accurately predict existing types of phenomena due to its having latched onto the world, without being able to make predictions about novel types of phenomena. Worse yet, we can easily imagine a scenario where a postulated structure, which can predict novel types of phenomena, eventually gets thrown away because no one realises that it can. If this sounds fictional, recall that it was not Fresnel who realised that his theory entailed the occurrence of a bright spot at the centre of the shadow of an opaque disc lit from a single slit, but Poisson. This consequence of the theory is not at all obvious. It *could* have been missed altogether. Nothing guarantees that we can see all the consequences of a particular theory, not least because it is unlikely that we have all the relevant auxiliaries required to test a theory at hand. Another reason that could prohibit scientists from realising the potential of a theory for novel

[214] Mathematization on its own is not sufficient. This is evident if one looks, for example, at econometrics, where there is a high level of mathematization and comparatively very little predictive power.

predictions concerns the unavailability of the required technology to build test instruments. We would certainly not want to exclude structures that may otherwise be predictively successful from our evidence of the cumulative development of science. Yet, if we follow Worrall's account, we would have to.

In general, it seems to me that criteria of maturity offer quasi-arbitrary divisions of the history of science that are unable to impute any epistemic benefits to those who use them. The epistemic structural realist can thwart anti-realist claims, i.e. that successful theories have nonetheless been abandoned, by simply appealing directly to the preservation of structure. Whether a structure comes from a mature or immature science or theory, or even from no theory at all, makes absolutely no difference.[215] What makes a difference is whether the structure survives in some recognisable form and is directly responsible for some predictive success.

6. What the History of Science Cannot Teach Us

In the last four decades, historical arguments have played a central role in the scientific realism debate. In one corner, anti-realists argue that no theoretical preservation takes place. That, they claim, should be indicative of the falsity of theories past, and the likely falsity of theories present and future. In the opposite corner, many realists argue that at least some theoretical components get preserved, a detail that, they claim, should be indicative of their approximate truth. Both sides thus agree that the historical record is essential in settling the debate.

Without doubt, the realist needs to provide a rejoinder to the anti-realist's historical arguments. Yet, the expected returns from a realist-friendly interpretation of the history of science have been overestimated. Realists seem to behave as though realism will defeat its foes on the basis of establishing historical continuity. Yet, on a strict reading, that would require belief in the view that the preservation of a component X is a necessary and sufficient condition of X's approximate truth/truth. No realist, I hope, would be happy to adopt such a strong claim. Indeed, it can easily be shown that the preservation of a theoretical component through theory change is

[215] In support of this point, I can point the reader to a recent collection of essays *Models as Mediators* (edited by Morgan and Morrison (1999)), whose authors (for, example, Nancy Cartwright and Margaret Morrison) argue that many structures in science are significantly autonomous from theory.

neither a *necessary* nor a *sufficient* condition for its truth or approximate truth.[216] Realists should not even adopt the weaker, though still strong, claim that preservation is either a necessary *or* a sufficient condition for the approximate truth/truth of what gets preserved.

It is *not a necessary condition* because even though a component may be true/approximately true, its preservation is not guaranteed. Suppose a scientist postulates a law that is actually true or approximately true in its domain of phenomena. Many reasons, quite a few of which are social/cultural, could transpire to make the general scientific community cast the law aside. For example, if the law seems incompatible with well-established theories, there will be no guarantee it gets adopted. This will especially be the case when the predictive accuracy of the law cannot yet be fully tested – as when the instruments to perform such measurements are inexistent, unreliable or inaccurate. An example of, at least temporary, unreliability/inaccuracy in the current historical context involves the Irvinist equation for the determination of the absolute zero point of temperature, $c_ix + L = c_wx$, where c_i is the heat capacity of ice, c_w the heat capacity of water, L the latent heat of fusion, and x the absolute temperature of ice/water at the melting point. The equation was contested at first but the issue could not be settled due to a lack of reliable and accurate measurements. Eventually, the accuracy and reliability of the measurements improved sufficiently to tell against the equation. This fact notwithstanding, the point here is that there is no guarantee that we will always be able to construct instruments that can assess the predictive power of theories. Moreover, even if we do acquire the required instruments, the theory may already have been shelved. For this and other reasons, there is no guarantee that true/approximately true theoretical components will be preserved.

A potential realist reply may take the following form: Had the scientific community tested the law, they would have discovered its wonderful predictive powers, making its rejection difficult, if not completely out of the question. In other words, the predictive success enjoyed by the law should guarantee that scientists, following the canons of rationality, would preserve it for posterity. Though this may largely be

[216] That it is not a sufficient condition is a point that has also been made by Chang (2003).

true, notice that now it is the predictive success of the law that takes centre stage, not its preservation. In fact, the issue of preservation becomes parasitic on the issue of predictive success. Preservation becomes superfluous.

It is *not a sufficient condition* because the mere survival of a given theoretical component does not guarantee that it has latched onto the world. Various reasons may be responsible for a component's survival. It may be a convenient feature of scientific practice, or it may be a useful tool that has no power of representation. Plenty of examples can be drawn from the history of science to make plain that the preservation of a theoretical component is an insufficient condition to its truth/approximate truth. In the case study presented in this chapter, we traced the hypothesis of the materiality of heat for at least two millennia until its wholesale rejection in the middle of the 19th century. Its long preservation guaranteed neither its survival nor its truth/approximate truth.

In the recent history of science, there has been a substantial correlation between preservation and approximate truth. This correlation, however, can be explained by the fact that scientists are more likely to preserve those components that have predictive success and independent confirmation. Since one of science's chief aims is to procure accurate predictions, any preserved elements are likely to have predictive merits. Indeed, as the demands for predictive accuracy increase, it will be reasonable to assume that so does the preservation of predictively powerful, as opposed to merely convenient, elements.

To be fair, very few realists would take preservation as a necessary or sufficient condition for a theory's truth/approximate truth.[217] But then, what role exactly does preservation play in the scientific realism debate? I am not claiming here that we should completely dismiss the importance of history in the scientific realism debate. Given that the pessimistic meta-induction argument is a real threat to the realist, one can employ cases of genuine preservation to defuse anti-realist objections stemming from the history of science. That, however, is as far as the preservation strategy will take the realist. The most telling, though admittedly not conclusive, test for which

[217] Exceptions can always be found. Worrall, for example, takes preservation to be a necessary condition for a theory's truth/approximate truth.

components have latched onto the world is whether they have independent confirmation. Testing this can be done independently of any historical considerations, and, therefore, makes the requirement that a component be preserved superfluous. Realists should thus focus more on elaborating such prediction-based criteria.[218]

7. Conclusion

The primary aim of this chapter was to evaluate how scientific realism and structural realism perform when confronted with the caloric theory of heat.

In assessing scientific realism, two main strategies were considered. The first tries to establish that the caloric is a referential term. Like many others, I argued that this strategy fails to justify in a principled way which terms can be thought of as referential and which cannot. The second strategy, Psillos' own, tries to establish that caloric was not a central term in the theory. Employing, among other things, objections from two critics of Psillos, i.e. Stanford and Chang, I indicated that Psillos' arguments by and large fail. On the basis of Psillos' own definition of what counts as central, the caloric posit was neither doubted by leading figures of the theory as central nor was it entirely dispensable in deriving explanations.

Indeed, what Psillos hoped to achieve by this strategy is unrealistic. First, scientists' epistemic attitudes towards a given theoretical term cannot always be trusted. A glaring example from the given historical context is the trust scientists placed – or, should I say misplaced – in the hypothesis of the materiality of heat. A more reliable factor seems to be whether the term's description is really indispensable in producing predictions and explanations. Second, by being relative to a given epoch, Psillos' criterion of indispensability is vulnerable to Chang's objection that the caloric posit and its properties were indispensable-at-the-time.

In assessing structural realism, I argued that Psillos' three cases of successful components are structural. The three cases involve five equations that I listed as L1-L5. What seemed as prima facie support for structural realism, however, had to be

[218] Of course, if radical underdetermination holds not even prediction-based criteria can save realism.

re-evaluated when the phenomenological character of L1-L5 was revealed. Worrall-style ESR does not seem to benefit from such laws, since they have no theoretical terms to replace with variables and quantify over as Ramseyfication prescribes. Though there are plausible solutions to this problem none were pursued in detail because Ramsey-style ESR faces other difficulties that make it unattractive.

I believe to have shown that Russellian-style ESR can more easily benefit from L1-L5. These laws are the concrete observational structures from which the abstract structure of the physical world is derived. In their capacity as concrete observational structures, they are, of course, supportive of constructive empiricism, too. The crucial difference again depends on whether or not one is willing to accept the principles H-W and MR. These principles allow the believer to cross the boundary from anti-realism to realism.

It has been pointed out that belief in ESR does not indiscriminately commit one to the belief that all structures will be preserved. Some structures never make it past scientific revolutions. Moreover, when they do survive, structures are not always found intact. More often than not, old structures reappear only as limiting cases of new structures. I suggested, following Worrall and many others, that this fact can be accommodated by ESR when appeal is made to the correspondence principle. I offered one concrete version of the correspondence principle, which I called ‘NC’, arguing that it is corroborated by some well-known cases in the history of modern physics. I believe to have shown that NC is difficult to satisfy, i.e. it is non-trivial, and that it should therefore not be taken lightly. In the end, I conceded that more work needs to be done to establish whether the history of science corresponds to MSSS´, i.e. that “not all structures may survive, but *most predictively successful* elements that do survive, either intact or suitably modified (for, example, according to NC), are structural.”

The criterion of the maturity of science is one of the last issues I took up. My conclusion was that the structural realist does not need to draw quasi-arbitrary distinctions between mature and immature science. Instead, what matters is whether a structure survives in some recognisable form. Indeed, even this last claim is not strictly speaking correct, since preservation seems parasitic on the predictive power of structures. For this reason, I urged the participants in the scientific realism debate to not overestimate what can be achieved by historical continuity.

6

UNDERDETERMINATION OF THEORIES BY EVIDENCE

1. Introduction

So far my discussion of underdetermination was either peripheral as in the case of Newman's objection and the pessimistic meta-induction, or non-traditional, as in the case of the underdetermination of relations by narrowly construed formal properties (see chapter three, section four). In this chapter, I turn to the more traditional forms of underdetermination and discuss some of the central issues surrounding this topic. The aim will be to evaluate to what extent, if at all, the realist, and in particular the structural realist, can overcome the difficulties posed by underdetermination. In pursuing this aim, I will thus be addressing RP1 on the list of challenges any realist position must overcome. To remind the reader, the challenge is to show that from a pool of empirically equivalent theories, at least some are more epistemically warranted than others.

As I have already indicated in chapter one, arguments from the underdetermination of theories by evidence have been criticised as being unable to deliver all that the anti-realist wants to derive from them. Today, the debate over the epistemic significance of underdetermination centres on Laudan and Leplin's seminal article. To repeat, Laudan and Leplin present a two-pronged critique of those underdetermination arguments that rely on the notion of empirical equivalence.[219] On the one hand, they question the view that *all* theories have genuine empirically equivalent rivals. On the other, they argue that, even when theories have such rivals, there are still justifiable ways to choose between them. Let us look at these prongs one at a time.

[219] Though they focus on defeating the inference from empirical equivalence to underdetermination, the tenor of their claims suggests that they want to defeat underdetermination altogether. For example, they say: "The thesis of underdetermination, at least in so far as it is founded on presumptions about the possibility of empirical equivalence for theories – or 'systems of the world' – stands refuted" (1991: 466).

2. Does Every Theory have Empirically Equivalent Rivals?

Laudan and Leplin cite three theses which, when taken together, allegedly "cast doubt on empirical equivalence in general" (1991: 451). These are:[220]

(VRO) What is observable varies through time.

(NAP) The derivation of observable consequences typically requires auxiliaries.

(IAA) Auxiliaries vary through time.

On the basis of these, they argue that the observational consequences of a theory are not fixed but vary over time, and conclude that they are not clearly identifiable and that empirical equivalence is, therefore, defeasible. In their own words:

> … any determination of the empirical consequence class of a theory must be relativized to a particular state of science. We infer that empirical equivalence itself must be so relativised, and, accordingly, that any finding of empirical equivalence is both contextual and defeasible (454).

One immediate reply to this argument is that whether the observational consequences of a theory are fixed or vary over time is a matter independent of whether theories are empirically equivalent. Underdetermination is supposed to hold for *any* given body of observational evidence, giving rise to infinitely many empirically equivalent theories that diverge on their theoretical claims. We can capture the variability of a theory's observational consequences over time by saying that out of a class *Q* which contains *n* sets of observation sentences as members, i.e. O_1, O_2, …, O_n, the theory (plus any auxiliaries) at any time *t* entails just one set, though we may not know which one. Given the definition of UTE-CE, i.e. the underdetermination thesis, any member O_i of class *Q* will be a set entailed by infinitely many theories that diverge on their theoretical claims. Hence, if the underdetermination thesis is correct, whatever the set of observational consequences entailed by a given theory plus auxiliaries, other incompatible theories will also have that set as their set of observational consequences. Naturally, the correctness of the underdetermination thesis cannot be assumed, since it is what is at stake here. Nonetheless, if Laudan and Leplin are to assume its incorrectness in order to prove that their objection holds, then they are simply begging the question. It thus seems

[220] The acronyms are those used by the authors.

that raising doubt about the stability of a theory's observational consequences does not dent the underdetermination thesis.

Laudan and Leplin anticipate this reply and, as a consequence, devote a whole section of their paper trying to undermine it.[221] Their argument consists of two strategies. *First*, to reject the view that there exists an algorithm that can generate empirically equivalent rivals for any given theory. *Second*, to deny that the cases offered as examples of empirically equivalent theories are genuine.

I will begin with the second claim since it can be more easily dismissed. In pursuit of the second strategy, Laudan and Leplin concoct an example, inspired by one of van Fraassen's examples in *The Scientific Image*, and show how it fails to be a case of empirical equivalence. We need not delve into the details. Showing that the two theories under consideration are not empirically equivalent is, of course, a correct step on the path to showing that examples of empirically equivalent theories are not genuine. It is, however, a far cry from showing that neither has *any* empirically equivalent rivals. It is this latter claim that needs to be established in order for Laudan and Leplin's conclusion to go through, namely if there is at least one theory that has no empirically equivalent rivals, then, obviously, not all theories have empirically equivalent rivals.

In pursuit of the first strategy, Laudan and Leplin examine the prospects of algorithms that reduce theories to their instrumentalist counterparts, and claim that these would invariably fail.[222] There are at least two problems with Laudan and Leplin's suggestion. Firstly, such algorithms do not suffice to produce rival theories, for they need to be augmented with a mechanism that expands the instrumentalist counterpart of the original theory into the theoretical vocabulary. The expansion must be conducted in different ways so as to yield inconsistent, i.e. rival, theories. Unsurprisingly, after a short discussion of the limitations of instrumentalist

[221] Here's what they say "[t]he response we anticipate to our argument is a challenge to its assumption that empirical consequence classes must be identified for their empirical equivalence to be established" (455).

[222] Though they do not give any examples of instrumentalist algorithms, they may have something like the Ramsey-sentence in mind. Naturally, quite a few philosophers, including Grover Maxwell and John Worrall would oppose this identification, since they believe that the Ramsey-sentence of a theory *T* is not an instrumental reduction of *T*, but rather is *T* when properly construed.

algorithms, Laudan and Leplin state the obvious, i.e. "what application of an instrumentalist algorithm to a theory produces is manifestly not an alternative *theory*" (456) [original emphasis].

Secondly, after their rejection of instrumentalist algorithms, they hastily conclude: "We know of no algorithm for generating genuine theoretical competitors to a given theory" (457). Laudan and Leplin start this section of their paper by claiming they can defeat the idea that there exist empirically equivalent theories for any given theory. All they end up showing, however, is that they cannot themselves devise an algorithm to produce such theories. As we have seen, the only type of algorithm they consider, viz. instrumentalist algorithm, is not even a suitable candidate, for it produces instrumentalist versions of theories and not genuine rival theories. In general, we can say that our inability to devise an algorithm for producing empirically equivalent theories does not entail that such theories do not exist.[223]

This point can be supported by a lesson learned in the theory of recursion. Two definitions will help us here: (1) We say that a set *A* is *recursive* (or *decidable*) if there exists a Turing machine, i.e. an algorithm, which can determine in a finite number of steps whether or not any given object is a member of *A*. (2) We say that a set *A* is *recursively enumerable* if it can be written as a sequence $A = \{a_1, a_2, a_3, \ldots\}$ which can be generated by means of a Turing machine. All recursive sets are also recursively enumerable but not vice-versa. A halting set, for example, is recursively enumerable but not recursive.[224] Some sets are not even recursively enumerable, namely non-diophantine sets. The moral of the story is that these are non-empty sets that cannot be specified algorithmically. Likewise, in our context, should it turn out that there is no algorithm for producing empirically equivalent rivals, the conclusion need not be that sets containing such rivals are empty.[225]

[223] In the formulation of underdetermination arguments, there is no requirement that empirically equivalent alternatives must be produced first, but rather that such alternatives exist.

[224] A halting set is a recursively enumerable set consisting of all inputs on which a computer program halts.

[225] We can even apply this moral to the issue about the identifiability of a theory's set of observational consequences, which Laudan and Leplin contest. That is, we can think of sets that are purportedly not clearly identifiable as sets that cannot be produced algorithmically. These can nonetheless be non-empty and well-defined sets.

Algorithmically Produced Rivals

Is it really the case that not all theories have empirically equivalent rivals? To evaluate this claim, let us consider some of the obstacles that arise when we try to devise rival theories. A simple algorithm for producing rival theories has the following form: Take an existing theory and add to it a hitherto un-included theoretical claim. Suppose we have a theory X. We add to it theoretical claim T_1, making sure that X does not contain or entail T_1. The result $X\&T_1$ is a new theory rivalling the old one. We can repeat this process indefinitely, each time adding a different theoretical claim that is not included in the original theory.

Obviously, this method is grossly inadequate. One major problem is that it does not guarantee empirical equivalence. To rectify this, we must require that the new theoretical claim does not, when taken together with the original theory, affect its observational consequences. I can think of three different types of theoretical claims that might conceivably satisfy this condition:

(1) Theoretical claims that have no observational consequences whatsoever.
(2) Theoretical claims whose observational consequences are already amongst the consequences of the original theory.
(3) Theoretical claims that have observational consequences only when conjoined to other statements, none of which are entailed by the original theory.

The first type of theoretical claim seems straightforward enough. Are there any examples of such claims? Take Newton's notion of absolute space. According to Newton, "[a]bsolute space, of its own nature without reference to anything external, always remains homogeneous and immovable" ([1726]1999: 410). Many of us would doubt whether the sentence asserting the existence of absolute space has any observational consequences.[226] It is, however, always possible to construct conditional sentences that endow observational consequences to theoretical claims of type one. In the example just mentioned, such a conditional sentence would take the form 'If there is absolute space, then P', where P must be an observation sentence.

[226] Stephen Brush and Gerald Holton go as far as to brand such claims meaningless (see ([1952]2001: 163).

This just means that no theoretical claims of the first type exist, since we can always find additional sentences that when conjoined with any theoretical claim produce observational consequences.

The second type of theoretical claim comprises those that have observational consequences already entailed by the theory in question. This preserves the original set of observational consequences, and thus the empirical equivalence is not violated. It may be objected that if the extra theoretical claim does not contribute any new observational consequences then why append it to the original theory in the first place. In particular, it may be argued that these claims can be rejected on account of parsimony. Though this objection is intuitively sound, it does not eliminate all theoretical claims of this type, since such claims can be desirable for reasons independent of their observational consequences.[227] For example, a theoretical claim of this type may be explanatory in some way that the original theory is not. Given that realists value the explanatory power of theoretical claims, they would find it difficult to dismiss this possibility. Finally, it is worth recalling that the notion of parsimony is notoriously difficult to define and defend.

The third type of theoretical claim finds support from Duhem's thesis, according to which theories or statements cannot be tested in isolation, for they never have observational consequences of their own. Laudan and Leplin would have to accept this type of claim, since their thesis NAP is premised on an almost identical point, i.e. that theories often require auxiliaries in order to entail observable consequences (see (1991: 452)).[228] The point here is that the addition of *certain* theoretical claims T_i' to a theory T yields a theory $T\&T_i'$ that is empirically equivalent to T, since for T_i' to have any observational consequences, they have to be supplemented with additional theoretical claims T_k''. An example of this type of theoretical claim, taken from Duhem, is that light consists of projectiles that are emitted at great speed from luminous sources like the sun. Taken on its own, this central claim of the emission theory of light does not entail anything observable. The projectiles are so small and travel so fast that we cannot directly observe them. To derive observational

[227] I am not claiming here that the reasons are epistemologically significant. They could very well be merely pragmatic.
[228] Unlike Duhem, we can concede that *some*, but not all, theories/theoretical claims can be tested in isolation. Laudan and Leplin also make this point.

consequences we need to assume additional theoretical claims, i.e. that these projectiles penetrate all transparent bodies, that they experience attractions and repulsions, etc. The collection of these propositions allowed the derivation of the observational claim that light travels faster in water than in air. Contrary to this claim, Jean Foucault showed in 1851 that the speed of light in air is greater than in water.

Laudan and Leplin never really elaborate what characteristics a rival theory should have to be considered genuine. They mention the obvious, namely that they do not consider logically equivalent theories as rivals. Their only other remarks on this issue are obscure: "As we do not question the empirical equivalence of logically equivalent theories, we ignore this suggestion and assume henceforth that theories whose empirical equivalence is at issue are logically and conceptually distinct" (1991: 455). I presume that by 'logically and conceptually distinct' they mean 'not logically equivalent'. So far, the focal point of my discussion of producing empirically equivalent theories has been the addition of theoretical claims to existing theories. This allows the production of logically inequivalent rivals but not necessarily the production of incompatible rivals.[229] We thus come to the question: Does every theory have *logically inconsistent* rivals that are nonetheless empirically equivalent?

One obvious way of producing such rivals involves replacing – instead of adding – theoretical claims with claims incompatible with them. Like before, eligible theoretical claims can take one of the following forms:

(1) Theoretical claims that, (a) have observational consequences but only when conjoined to other statements none of which are included in the original theory, or (b) have observational consequences already contained in the main theory. These must be replaced with theoretical claims that are either

[229] Adding a theoretical claim T_1 to a theory T does not always mean that the new theory $T\&T_1$ will be logically inequivalent to the original theory T. This is a simple logical point. The result of conjoining some sentence Q to some sentence P, i.e. P&Q, is not always logically inequivalent to P. Indeed, it is equivalent if and only if P entails Q.

of form (a) or (b), and that result in a theory that is logically incompatible to the original.[230]

(2) Theoretical claims with observational consequences that, if removed, will alter the observational consequence set of a theory must be replaced with theoretical claims that will return the set to its original state and will result in a theory logically incompatible to the original.

There may not exist historical examples for each of the above ways through which replacement can be performed, but, as we saw earlier, the point about underdetermination can be made independently of such examples. The above two methods' logic of constructing empirically equivalent theories that are logically incompatible is quite straightforward: Replace theoretical claims so that 1) the resulting theory is incompatible with the original theory, and 2) the observational consequence set remains untouched.

The realist might object that these are not genuine rivals. Yet, the notion 'genuine rival' remains elusive. As we have seen earlier, Laudan and Leplin's scant remarks were insufficient to pin it down. Until an elaboration and defence of this notion appears, the realist cannot simply rely on the admittedly innocuous-sounding proclamation that some theories have no empirically equivalent rivals. That is not to say that the anti-realist is better off, for the problem of defining the concept of a 'genuine rival' concerns both parties in the debate. In sum, it is not clear where the burden of proof lies.

3. Can we Justifiably Choose between Empirically Equivalent Theories?

This takes us to the next prong of Laudan and Leplin's critique. They claim that even if two theories are empirically equivalent, we can still choose between them. More precisely, they hold that (1) a theory may be supported by evidence that does not form part of its observational consequences, and (2) the observational consequences of a theory need not provide support for it. In other words, they are claiming that

[230] I treat the two types of theoretical claims discussed above jointly, because the empirical equivalence between original and modified theory will still hold even if the theoretical claim replaced is not of the same type.

empirical equivalence does not entail co-confirmation, i.e. two empirically equivalent theories need not be equally confirmed by the evidence.

In support of their second claim, Laudan and Leplin simply point out that the fact that *H* entails *e* does not necessarily mean that *e* confirms *H*. This seems correct, but presents no problems to empirical equivalence cases. One need only restrict a theory's observational consequence set *S* to a subset S_c containing as members all and only those observational statements that have the power to confirm. Notice that this restriction is equivalent to simply taking a theory whose set of observational consequences is S_c to begin with. But there is no reason to suppose that this latter theory has no empirically equivalent rivals.[231] Thus, all that Laudan and Leplin's point achieves is to trade one empirical equivalence class for another.

In support of their first claim they make use of the following argument. Suppose that hypothesis *H* entails evidence *e* and that *e* confirms *H*. Evidence that confirms a hypothesis *H* will also confirm (a) any theory *T* that entails *H* and (b) any other hypothesis H_k entailed by *T*. The point is that while *H* entails *e*, H_k need not. In other words, a hypothesis may be supported by evidence that it does not entail. Here's a reconstruction of the form of argument that Laudan and Leplin sanction (see (464)):

For any *i*, *j*, and *k*:

$H_i \vdash e$	(1) - premise
e confirms H_i	(2) - premise
$T_j \vdash H_i$	(3) - premise
$\therefore$ e confirms T_j	(4) - by CCC (see below)

[231] Recall that, following Laudan and Leplin's article, we assume in this subsection of the chapter that empirical equivalence holds for any given theory. The contested point here is whether there are ways to choose between theories that we have already ascertained as empirically equivalent.

$T_j \vdash H_k$ (where $k \neq i$) (5) - premise

$H_k \nvdash e$ (6) - premise

$\therefore$ e confirms H_k (7) - by SCC (see below)[232]

Samir Okasha (1997) has correctly criticised this form of argument by saying that it relies on two principles that Hempel showed to be incompatible, namely the 'converse consequence condition' (CCC) and the 'special consequence condition' (SCC). According to CCC, if some evidence confirms a statement *S* it also confirms any statement *S′* that entails *S*. According to SCC, if some evidence confirms a statement *S*, it also confirms any statement *S′* that *S* entails. Hempel demonstrated that SCC and CCC *can*, when used together, lead to absurdity. The following argument is an example of how the principles can be used to derive confirmation for *any* theory:

H_i confirms H_i (a) - self-evident[233]

$(T_j \,\&\, H_i) \vdash H_i$ (b) - self-evident

H_i confirms $T_j \,\&\, H_i$ (c) - by CCC

$(T_j \,\&\, H_i) \vdash T_j$ (d) - self-evident

$\therefore$ H_i confirms T_j (e) - by SCC

On the basis of CCC and SCC, this argument shows that anything can confirm anything, an obviously absurd result.

To Okasha's critique I want to add that, even if employed on their own, the two principles can lead to incorrect inferences. It is well known that employing CCC on its own still allows us to derive that evidence which confirms a hypothesis will confirm *any* theory, no matter how ridiculous, that entails the hypothesis.[234]

---

[232] N.B.: Lines 1-4 count as additional premises to the second argument.
[233] This assumes that a statement can confirm itself.
[234] This is a point I briefly raised in section three of chapter three when I was tackling objection PS6. The result is also known as the 'tacking paradox'.

Similarly, employing SCC on its own allows us to derive that evidence which confirms a theory will confirm *any* hypothesis, no matter how ridiculous, that is entailed by the theory.

It is important to note that this does not mean that all inferences made on the basis of either or even both principles will be incorrect. But how would we set about showing which inferences are warranted and which are not? Laudan and Leplin offer no qualification on their use of CCC and SCC, assuming that the logical relations between theories and hypotheses are enough. As we have seen, this leads to the above absurdities. Contra Laudan and Leplin, warranted inferences seem to involve theories and hypotheses that are intimately, not just logically, related.

The sort of consideration referred to by the traditional scientific realist as indicating the intimacy of various parts of theories is *unity*. Together with simplicity, explanatory power, and comprehensiveness, they are frequently cited as the ultimate defence of realist claims (see, for example, Nelson (1996) and Psillos (1999)).[235] Possession of these virtues is viewed as extra-observational evidence that can overcome claims of empirical equivalence.[236] More precisely, that T_1 and T_2 are empirically equivalent does not entail that the evidence equally confirms T_1 and T_2. This is a trivial point once one accepts that the possession of the above virtues counts as (non-observational) evidence.[237] In the Bayesian framework, for example, these theoretical virtues can be reflected in the choice of priors.[238]

Van Fraassen objects that the so-called ‘theoretical virtues’ are nothing but pragmatic features of theories (see, for example, his (1980: 87-89)) with no epistemic significance. In other words, he denies that these virtues have any

[235] Some take these virtues to denote one and the same thing that is merely expressed in different ways.
[236] It is worth noting that Laudan and Leplin try to defeat underdetermination without recourse to extra-observational considerations.
[237] Another realist approach to theoretical virtues is to deny that they are evidential, but assert that they still have a role to play in the truth or approximate truth of a theory. One can thus argue from the platform that they are ‘extra-evidential’ considerations pertinent to the truth content of the theory.
[238] Of course, being a Bayesian does not necessarily mean that one takes theoretical virtues as formal criteria for restricting one’s priors. Thus, in considering the role of the notion of simplicity in Bayesianism, Howson and Urbach (1996) state that “the addition of *any* criterion [including simplicity] for determining prior distributions is unwarranted in a theory which purports to be a theory of consistent degrees of belief, and nothing more” (418) [original emphasis].

evidential status, i.e. that they can confirm or disconfirm a theory. They can, of course, be used as pragmatic criteria for acceptance of a theory, but their role is restricted to just that. The lack of consensus on how to understand theoretical virtues certainly makes things easier for Van Fraassen's view, since it motivates the suspicion that they might be conventional and/or pragmatic features of theory choice. Even more damaging is the insufficiency of evidence that nature is amenable to a unified, comprehensive, simple, and explanatorily powerful account. This view finds some proponents in the realist camp. Nancy Cartwright, for example, argues in favour of the disunity of science.[239]

There is thus a standoff over the epistemic significance of theoretical virtues. In the current context, this means that to the extent that theoretical virtues might be able to substantiate Laudan and Leplin’s claim, i.e. that a theory is supported by evidence that does not form part of its observational consequences, we must reserve our judgment.

Evidential Equivalence

I will not enter into the dispute over the epistemic import of theoretical virtues. Instead, I propose to look at some of the ramifications in case either side wins.

Suppose the anti-realist wins. In that case, we would say that a theory cannot be confirmed or disconfirmed on the basis of theoretical virtues. Does that mean that only observational consequences are evidentially relevant to a theory? The answer to this question depends on whether the realist can find anything other than the aforementioned theoretical virtues that could seriously be considered as evidentially relevant to the confirmation of theories. It must nonetheless be noted that the current state-of-affairs does not provide much hope that other non-observational considerations may eventually succeed where theoretical virtues such as unity and simplicity have failed.

Suppose the realist wins the argument, thereby establishing that theoretical virtues are evidentially relevant. In that case, we would say that a theory can be confirmed

---

[239] That does not mean that she rejects any kind of appeals to unification, let’s say within specific domains, but rather that she rejects overall claims of the unification of the science.

or disconfirmed on a non-observational basis, effectively going beyond the epistemic commitments of constructive empiricism. This said, is the issue of underdetermination settled? The answer is ‘no’. Even if two empirically equivalent theories are evidentially inequivalent, there may still be theories that are evidentially equivalent to each! What needs to be established is whether or not there are evidentially equivalent rivals for any theory, i.e. whether the following sentence is true:

(EVE) $\forall T\ \exists T'\ [(T' \vdash \neg T)\ \&\ T'$ is evidentially equivalent to T].

Even if empirical equivalence claims can be defeated by appeal to theoretical virtues, there remains the issue of evidential equivalence, or EVE for short. If EVE is false, there are at least some theories that the evidence can uniquely identify. If, however, EVE is true, then underdetermination remains rife, albeit in a restricted form that no longer supports constructive empiricism. The difference consists in the extent to which underdetermination can be mollified.

The suggestion that arises here is that realism does not need to defeat underdetermination altogether to be able to defeat constructive empiricism. To achieve the latter, it suffices to provide non-observational constraints to underdetermination. To achieve the former, the following sentence must be shown to be true:

(EVE2) $\neg\ \exists T\ \exists T'\ [(T' \vdash \neg T)\ \&\ T'$ is evidentially equivalent to T].

That is, if we want to defeat underdetermination entirely, we need to show that *no* theory has evidentially equivalent rivals, i.e. all theories are uniquely identified by some set of (observational and non-observational) evidence.

EVE2 seems too strong a claim, since it requires all theories to lack evidentially equivalent rivals. Realists, it might be argued, can achieve their aim of vanquishing underdetermination with the weaker claim that all *true* theories lack evidentially equivalent rivals. That is, they can achieve their aim if they can show the following sentence to be true:

(EVE3) $\neg \exists T \exists T' [(T' \vdash \neg T) \ \& \ T'$ is evidentially equivalent to T], where *T* is any true theory.[240]

Suppose EVE3 is true. What follows? Once scientific inquiry arrives at a true theory, this theory will have no evidentially equivalent rivals. Since EVE3 refers only to true theories, theories that do not possess this trait may well have evidentially equivalent rivals. To the extent that all of our current theories are at best approximately true, i.e. not true simpliciter, establishing EVE3 is of no immediate help to the issue of underdetermination. In other words, a realist must also require that approximately true theories lack evidentially equivalent rivals:

(EVE4) $\neg \exists T \exists T' [(T' \vdash \neg T) \ \& \ T'$ is evidentially equivalent to T], where *T* is any approximately true theory.

The reference to the concept of approximate truth complicates the evaluation of EVE4's truth conditions. Without a clear understanding of this slippery concept, EVE4 would be hard to establish. This adds complications to those already present due to the concept of evidential equivalence. For, even though we have been supposing that theoretical virtues have evidential status, there is still the issue of how to unpack the concept of evidential equivalence.

I hope to have offered a glimpse into some of the consequences for theory choice should theoretical virtues be more than just pragmatic considerations. One lesson to take away from this sub-section is that theory choice may well remain indeterminate even after a potential defeat of the constructive empiricist brand of underdetermination.

[240] In a sense, EVE3 says something trivial. If *T* is true, then surely *T′* must be false since it contradicts *T*. Why then do we *need* to show that EVE3 is true? Presumably because it is *possible* for two or more theories to be evidentially equivalent and true with respect to all possible observations, yet still have inconsistent components.

4. Structural Realism and Underdetermination

The Inverse Relation of Epistemic Commitments to Underdetermination

To get a firmer grip on the general notion of underdetermination we need to circumscribe it. At one end, we have the most radical form of underdetermination. This arises if no constraints can be given to the content of theories – not even empirical constraints. In short, any theory is as good as any other. At the other end, we have a complete lack of underdetermination. This presumably arises if sufficient constraints can be given to uniquely determine a theory. In the middle lies a spectrum of different degrees of underdetermination, each arising on account of the strength of the different constraints that can be given. Whether or not a set of constraints mitigates underdetermination in an acceptable way depends on whether or not these constraints are warranted.

There is thus an inverse relation between constraints and underdetermination. The stronger the constraints the weaker the impact of underdetermination on theory choice. What kinds of constraints are we talking about? In the above subsection we talked about observational and non-observational constraints. Here we will look at constraints from a slightly different angle, namely from the perspective of the epistemic commitments of positions in the scientific realism debate. We will thus take a position's epistemic commitments to act as the constraints it *claims* to place on underdetermination.

Reformulating the inverse relation in terms of epistemic commitments, we get: The stronger the epistemic commitments of a position in the scientific realism debate, the more rival theories (to any currently successful theory) can *potentially* be eliminated, and, consequently, the more constrained underdetermination becomes. I say 'potentially' because, all things considered, whether rival theories can be eliminated depends on the available evidence. If the epistemic commitments of a position in the scientific realism debate are correct, then one would expect that the most successful theories would tell us something about the world in accordance with those commitments. For example, if structural realism says that we can only know the structure of the physical world, then our most successful theories will at best be able to tell us something about the structure of the physical world. Thus, if structural realism is correct, there are structural constraints that theories must satisfy. Any given theory of optics, for instance, will have to describe the structure of light in a way that more or less approximates the description given by Fresnel's equations.

The four positions considered in this dissertation, i.e. (non-structural) scientific realism, Ramsey-style structural realism, Russellian ESR, and constructive empiricism, can thus be ordered according to their epistemic commitments, and, consequently, according to their potential to eliminate rival theories.

Commitments → / **Viewpoints ↓**	Concrete Observational Structures	Ramsey Sentences	Abstract Structures	Concrete Physical Structures
Constructive Empiricism	●			
Ramsey-style ESR	●	●		○
Russellian ESR	●		●	○
Scientific Realism	●	●	●*	●

● indicates full specification.
○ indicates specification up to isomorphism.
* this holds with certain qualifications; see discussion below.

The table provides a rough classification, for, inevitably, some degree of idealisation is involved. Recall that scientific realists think that we cannot really divide things into observables and unobservables. They allow talk of ‘unobservables’ but only as a façon de parler, for they think that ‘unobservables’ are merely indirectly observable, i.e. observable with the help of instruments. The objects of observation, for scientific realists, are the objects inhabiting the physical world, i.e. electrons, bicycles, molecules, etc. By contrast, empiricists believe that the objects of observation are the phenomena, i.e. some sort of intermediaries between physical objects and perceptions. Van Fraassen, in particular, thinks that the line between observable and unobservable, though hazy, can still be drawn. Comparatively, the structural realists believe, like the scientific realists, that whether something is observed with unaided vision does not affect its status as an observable. Unlike scientific realists and more like constructive empiricists, they draw a line between observables and unobservables. Contrary to both scientific realists and constructive empiricists, the term ‘unobservable’ for structural realists encompasses all physical objects, whereas the term ‘observable’ has as its objects the perceptual contents of our heads. This is not to deny, of course, that these contents have causal origins in the external world.

Despite these discrepancies, there is still enough common ground to uphold the above classification. Take sub-atomic particles for example. These are unobservable according to the constructive empiricist and the structural realist, but indirectly observable according to the scientific realist. They are unobservable to the constructive empiricist and the structural realist for different reasons. The constructive empiricist will say that they are simply beyond the scope of unaided vision. The structural realist will say that subatomic particles populate the external world and as such are unobservable. These differences notwithstanding, all parties agree that photographs of tracks in cloud chambers are observable. They also agree that the content of these photographs can be expressed via observation sentences. This includes any relations or empirical generalisations that might be inferable from the photographs. In other words, all parties undertake an epistemic commitment to concrete observational structures.

Without further ado, let us consider how different types of constraints are supposed to curtail underdetermination. The first line of defence against underdetermination is logical consistency and observational constraints – the latter in the form of concrete observational structures. Innumerable theories can be weeded out as non-rivals solely on the basis of these two criteria. One need only consider that out of the set of *all possible theories*, empirically adequate and internally consistent theories form proper subsets.[241] All abandoned theories together with any empirically equivalent rivals they might have fall under this category simply on account of their empirical failures.[242] Moreover, we can quite easily contrive examples of ineligible rivals by making them entail observationally false consequences, or by making them internally inconsistent.

So even at the level of observational evidence, there are constraints that mitigate the absolute relativisation of theory choice. Naturally, logical consistency and observational constraints are not meant to distinguish between realist positions and constructive empiricism, as these are constraints that all parties in the debate take on board. Indeed, what we have just restricted is an extreme version of

[241] Here the notion of empirical adequacy is employed to mean 'adequate with regard to available observations', i.e. not in Van Fraassen's sense of 'adequate with regard to all observations'.
[242] Current theories also have their own share of failures and thus must also be weeded out. Of course, this is not an argument to abandon current theories without first finding a better alternative.

underdetermination that would probably be uncongenial even to social constructivists like Bruno Latour. What about the more familiar forms of underdetermination, like UTE-CE, that take observational evidence constraints for granted? Is there a way to mitigate these too?

ESR goes one step beyond constructive empiricism in its epistemic commitments. If ESR is correct, there will be structural constraints that theories need to satisfy. In other words, it has the potential to eliminate even more rival theories, i.e. it has the potential to mitigate UTE-CE. In addition to the consistency and observational constraints, ESR endorses the existence of, and epistemic accessibility to, structural constraints. It thereby purports to classify more theories as non-rivals on account of their failure to satisfy the structural constraints. Advocates of Ramsey-style structural realism take these structural constraints to be manifested in the Ramsey-sentence of a successful theory.[243] Advocates of Russellian structural realism take these to be the abstract structures that can be inferred from concrete observational structures. In both cases, the concrete physical structures are specifiable only up to isomorphism.

Arguably, scientific realism takes epistemic commitments even further. In addition to the consistency, observational, and structural constraints, its advocates claim that there are constraints that allow the full specification of concrete physical structures. According to scientific realists, one or more true theories could one day reveal the physical world as it is to us, with all true observation sentences included. By contrast, the ideal case for structural realism would be an isomorphic description of the physical world, plus all true observation sentences. The ideal case for constructive empiricism is just the set of all true observation sentences.

A closer look at the relation between the epistemic commitments of scientific realism on the one hand and each of ESR's two versions on the other can be instructive. The epistemic commitments of scientific realism are unequivocally stronger than those of the Ramsey-style ESR. Recall that the Ramsey-sentence of a theory T, i.e. $R(T)$, is

[243] Of course, if Ramsey-style ESR collapses to some form of empiricism, as some have argued, its structural constraints would not amount to anything beyond the observational level.

logically weaker than *T*. Thus advocates of the truth/approximate truth of *T* are necessarily advocates of the truth/approximate truth of *R*(*T*) but not vice-versa.

Things are not as clear when it comes to the Russellian version of ESR. Since there is nothing about scientific realism that requires adherence to the view that the structure of our perceptions reflects the structure of the external world, commitment to abstract structures seems out of place. Prima facie, then, the set of Russellian ESR-ist epistemic claims is not a proper subset of the set of scientific realist claims. If this is true, the two sets intersect only at the concrete observational structures. Can we still make sense of the view that the epistemic commitments of Russellian ESR are weaker than those of scientific realism?

In so far as scientific realists allow the full specification of concrete physical structures they commit themselves to abstract structures. After all, from the concrete physical structures one can deductively infer the corresponding abstract structures, i.e. the isomorphism classes whose members include the concrete physical structures in question. Of course, if they can fully specify concrete physical structures, scientific realists would probably have no need to infer the corresponding abstract structures. The point, however, is not whether scientific realists have a need to infer abstract structures, but rather whether commitment to concrete physical structures entails commitment to the corresponding abstract structures. I think this is clearly the case, since one can deductively infer the latter from the former and, as we all know, deduction is non-content increasing.

Still, there is a difference in the abstract structures of Russellian ESR and of scientific realism that cannot be ignored. The difference lies in the way each arrives at the abstract structures. According to Russellian ESR, we infer the abstract structures from the concrete observational structures. Against this, the above method of inferring abstract structures in accordance with scientific realism takes concrete physical structures as premises. In view of this difference, the particular abstract structures each position infers need not be the same. Even so, we can say that both positions make a general commitment to abstract structures. In this sense, the epistemic commitments of Russelian ESR are weaker than those of scientific realism.

To avoid confusion, a proviso needs to be made at this point. My discussion of constructive empiricism, structural realism, and scientific realism was somewhat caricaturist. Nothing prevents particular versions of these viewpoints, whose epistemic commitments and potential effects on underdetermination vary significantly. Without radically changing the commitments of these viewpoints, however, any such manifestations will fall within certain limits. For example, on the strongest reading of scientific realism, perhaps one that incorporates belief in a final theory of everything, the constraints defeat underdetermination in its entirety. On a weak reading of epistemic structural realism, even the abstract structure of the unobservable world is underdetermined to some extent.

The upshot of this whole discussion is that differences between the epistemic commitments of the four positions reflect differences in how far each position promises to restrict underdetermination. It is reasonable to assume that as rational agents we want to maximise our rewards. For some, this would mean opting for a strong version of scientific realism, since it offers the highest rewards. Yet, there is a trade-off. High rewards are often tied to high risks. This also holds true here. We can think of the epistemic commitments as risks depending on the extent to which these are warranted by the available evidence and arguments in the scientific realism debate. The question to ask then is: How far should we take our epistemic commitments given the present state of evidence and arguments?

Epistemic Warrant: *Structural Realism vs. Rivals*

In this subsection, I will argue that there is a good case for taking our epistemic commitments as far as ESR but not further. To this effect, I will utilize two groups of considerations that I label ‘historical’ and ‘other’ for the sake of expedience. I take these to be telling with respect to the present state of evidence and arguments in the scientific realism debate. The considerations are drawn from results reached in previous chapters. A number of arguments that up to now received scant attention will be developed further. This subsection hence doubles up as: i) a summary of the main reasons for accepting ESR, and ii) as an *indirect* answer to the question of underdetermination.

- *Historical Considerations*

In so far as the history of science is concerned, scientific realism seems handicapped when compared to structural realism. Worrall’s revival of ESR rests precisely on the idea that it offers the only realist account of the history of science that can accommodate the pessimistic meta-induction argument. In the previous chapter I evaluated realist attempts to overcome the meta-induction argument and argued that these have so far been unsuccessful. By comparison, I offered several reasons why, at least in this department, structural realism fares better than scientific realism. Some of the reasons appeared in chapter three, as arguments against Psillos’ claims that certain non-structural components of Fresnel’s theory were preserved. As I argued there, the components in question are in fact structural, and, hence, encourage the correlation between preservation of structure and predictive success. More crucially, my own case study in chapter five revealed further preservation at the structural level. In particular, I showed that structures postulated during the caloric theory’s reign made their way to modern thermodynamics.

Since components are rarely preserved whole, I have also taken into consideration the correspondence principle. Many, if not most, realists agree that the principle offers the best hope to resist anti-cumulativist arguments and to make sense of intertheoretic relations. This lends credence to structural realism for the principle operates solely at the level of mathematical structures. Although a comprehensive account of the correspondence principle has proved elusive, I offered one concrete version of the principle, what I called ‘NC’, arguing that it is corroborated by some well-known cases in the history of modern physics. I argued that NC is not trivially satisfiable and therefore cannot be easily dismissed.

Admittedly, more research needs to be carried out to establish whether the history of science corresponds to MSSS´, i.e. that “Not all structures may survive, but *most predictively successful* elements that do survive, either intact or suitably modified (for, example, according to NC), are structural.” Even so, from the standpoint of the current state of research, the historical evidence seems to support ESR more strongly than it does scientific realism.

A historical argument distinct from the above considerations, yet potent in its support of ESR, draws its force from the direction of scientific theorising. It is widely acknowledged that the trend in the history of science has been one of increasing mathematization. Explanations in pre-Galilean times were almost invariably devoid of mathematics. Though they were to a certain extent abstract, they would more readily rely on metaphors to explain their subject matter. Since Galileo and especially since Newton, there has been a transformation of scientific explanations, moving away from the narrative-based and ontologically profuse into the more abstract mathematical explanations. Andrew Warwick, in his study of the history of the pedagogy of physics, notes that the changing nature of the subject from the late 17th century onwards led to a change in the way in which it was taught and examined:

> … the oral lectures, catechetical tutorial sessions, guided readings and Latin disputations, through which most university students were taught and examined, were ill suited to imparting the skills of advanced mixed-mathematics. Unlike the more elementary parts of Euclid’s geometry, in which the propositions and demonstrations were expressed in verbal form, the new analysis relied on the mastery and application of several new and highly specialised symbolic languages (2002: 29).

Nowadays, quantum mechanics is the prime example of a theory not at all amenable to the narrative-driven styles of explanation. As most of us would agree, quantum mechanics primarily provides predictions rather than intuitive explanations.[244] Our understanding of it is almost entirely based on the abstract mathematical structures employed. It is interesting to note that one of the first things students of quantum mechanics are taught is to put aside their intuitions about how the world works, and concentrate on understanding the mathematics behind the theory. How else can they come to terms with such unintuitive features as wave-particle duality, non-locality, and the uncertainty relations? Indeed, unintuitive features like these can be found in other highly mathematised theories, and parts thereof. I could thus ask the same question about action at a distance, and a sui generis electromagnetic field.

---

[244] Wesley Salmon makes a similar, but more restricted, point when he remarks: “Because these fine-grained causal explanations are not possible, many philosophers, myself included, have concluded that quantum mechanics does not provide explanations of these correlations [i..e. correlations that exhibit non-locality]” (1998: 76).

Turning away from the issue of increased mathematization, how does an anti-realist position like constructive empiricism account for the preservation found in the history of science? Van Fraassen, if you recall from chapter two, agrees with Worrall that there is a preservation of structure through theory change. Unlike Worrall, however, he argues that it is not the structure of the unobservable that gets preserved but the structure of the observable. I dubbed this 'SRP4' or 'the empiricist structuralist challenge'. Can the challenge be met?

I believe that the Russellian version of ESR can meet this challenge. Assuming, like van Fraassen does, that we can only represent the phenomenal world is assuming that the world of phenomena has come unstuck from the physical world. How do the phenomena arise, if not for the underlying physical causes? Is it reasonable to assume that phenomena encode no information at all about the physical world?

Perhaps, it might be countered, the relation between the physical and the phenomenal worlds is not one of isomorphism. After all, one can accept that there is some relation between the two worlds without knowing what that relation amounts to. In other words, constructive empiricism is compatible with the view that the phenomenal and physical worlds are intimately related, but, owing to the lack of any evidence to the contrary, remains sceptical that the two worlds stand in a relation of isomorphism.

I can think of two reasons why this comeback is not compelling. First, the constructive empiricist is using double standards. He requires that the empirical substructures of theories are isomorphic to the structures in the phenomenal world but dismisses any such the relation between the phenomenal world and the physical world. Granted, the physical world seems more remote from our cognition than the world of phenomena. Does this licence scepticism for the one case but not for the other? The constructive empiricist needs a persuasive argument to show that this is so.

Second, there is an indirect way of testing the correspondence between the physical and the phenomenal world, namely through the correspondence between our *actions in the physical world* and our observations of their effects. Why should there be any such correspondence if there is no correspondence between the physical and the phenomenal worlds? The following example is instructive. When I see the cue ball lined up with the billiard ball I want to hit and the hole I want the ball to end up in, why is it that I know shooting straight will do the trick? It is because the two balls and the hole stand in a certain relation to each other, and if I do not guide my arms in conformity to this relation, I will not be able to shoot straight and experience that I shot straight. Likewise, if I do not want to pot the ball, one way to achieve this is by not shooting straight. In doing so, I can see the effect of my not shooting straight, i.e. I can see the ball bouncing off the sides of the billiard table.

Even though we only have direct access to our experience, all of us, realists and anti-realists alike, assume that our bodies reside in the physical world. Our actions are causes in the physical world and they have effects that we can perceive. This simple idea provides a constant reminder of the correspondence between the physical and the phenomenal worlds.

- *Other Considerations*

Three of ESR’s most forceful arguments are the arguments from the:

(1) structural source of predictive power
(2) limits of mathematical description
(3) linguistic intransmissibility of anything but structure

I have more or less introduced all three arguments in preceding chapters. Here, I will restate the arguments and, where needed, present a more thorough treatment than that given earlier.

The first argument, i.e. *the argument from the structural source of predictive power*, has only been hinted at in previous chapters. In spite of this, it does not need much setting up. The argument rests to a great extent on the simple observation that out of all the features of theories, only the mathematical structures possess sharp predictive power. This is significant because the testing of theories primarily concerns measuring how accurately its predictions match the observable phenomena. In so far as the epistemic warrant of theories is largely decided under such testing, mathematical structures win this contest hands down, for they alone have the ability to produce accurate predictions.

The critic may object, in the spirit of Duhem's thesis, that although mathematical structures are necessary in the process of churning out predictions, they are not sufficient. Other non-structural components are involved, it may be argued, and they surely deserve some of the credit for the predictive success and hence a share of the epistemic warrant. Two things can be said in opposition to this objection. First, we must be wary of so-called 'non-structural components'. As we have seen in section three of chapter three, Psillos' examples of non-structural components turned out to be structural. The onus is on the scientific realist to find legitimate examples. As I argued in that section and elsewhere, the legitimacy of such examples depends on whether they possess independent confirmation. For, even if we find non-structural components being preserved through the history of science, as was the case with the hypothesis of the materiality of heat, we must ask whether these make a direct contribution to the predictive aspects of theories.

The requirement that a component be independently confirmed might prima facie sound at odds with Duhem's thesis, which, after all, is holistic in its approach to confirmation. It should not, for by 'independent confirmation' I include those cases where background conditions and auxiliaries vary.[245] The necessary ingredient in my construal of independent confirmation is that the component contributes to the calculations of values, which can then be checked against testing. Thus, my construal of the notion of independent confirmation does not exclude holistic confirmation.

This brings us to the second point, namely that Duhem's lesson can be completely accommodated within the structural realist framework. By and large, terms in any given equation appear also in other equations. The term for specific heat at constant pressure c_p that appears in three – L1, L3, and L4 – out of the five equations discussed in the previous chapter serves as a good example. When some terms in an equation require for the determination of their value the calculation of other terms, we either turn to equations where the latter terms appear and are given values, or to

[245] My rough definition of independent confirmation can be found in footnote 195. The variance I have in mind does not involve inconsistent auxiliaries.

some experiment that evaluates such terms. The first disjunct, which is what concerns us here, illustrates how equations can act as auxiliaries to other equations. Another way to illustrate how the structural realist framework underwrites holistic confirmation is by noting the following two intuitive observations: 1) Equations, all of whose terms are determined and which agree on the value of any of their shared terms, get a holistic boost in confirmation. 2) Where disagreement crops up, we infer à la Duhem that at least one of the equations is at fault. Thus equations, or, more precisely, mathematical structures, have no problem playing the role of auxiliaries. The structural realist can thus give a holistic picture of confirmation solely in terms of structures.

The argument from the limits of mathematical description, which I presented at the end of the last subsection of chapter three, relies on the close link between structural realism and mathematics. As I have just pointed out, scientists increasingly rely on mathematics to describe the objects of scientific inquiry. Indeed, science employs mathematical objects as surrogates for its own objects. According to the orthodox view of mathematics, whatever can be described in the language of mathematics will be described only up to isomorphism. By transitivity, to the extent that scientific objects are describable only in mathematical terms, they will only be describable up to isomorphism.

Van Fraassen makes use of this argument to support constructive empiricism, claiming that phenomenal objects can only be described up to isomorphism. I see no good reason why the same argument cannot be utilised by the structural realist. The main difference between the two is that while the constructive empiricist takes phenomena to be the only legitimate objects of scientific enquiry, the structural realist extends this list to include physical objects. As I have argued above, the phenomenal world does not float about freely but is anchored in the physical world via causal chains. This should be seen as an argument for the legitimacy of the structural realist extension of the scientific enquiry's list of objects to physical objects.

I briefly presented *the argument from the linguistic intransmissibility of anything but structure* in section three of chapter two, when I pointed out its origins in Poincaré, Russell, Quine, and Carnap. The argument is a more generalised form of the argument from the limits of mathematical description for it extends the structural limits of description to all languages, i.e. not just mathematical ones. The gist of the argument is as follows: Assuming that we only have direct access to our perceptual experience and that this is private, how do we communicate knowledge to others? Though language cannot transmit perceptual experience, it can transmit relations that hold between perceptions.[246] Given that the relata cannot be transmitted, the relations transmitted can only be specified up to isomorphism. Thus, only (abstract) structure can be transmitted through language.

A concrete example will help make this point clear. Suppose two people, let us call them 'A' and 'B', are locked in separate rooms and *A* is presented with an object that she has to truthfully describe to *B* via a phone line. The catch is that *A* cannot mention the object by name or any synonyms it may have. What happens during the telephone conversation is pretty obvious. *A* tries to convey the object to *B* by giving certain characterisations of the object's appearance and function. These characterisations are necessarily structural, for *A* cannot transmit perceptions of the object itself. That is, *A* can only transmit information about the relations the object and any of its parts *appear* to stand in. This typically involves information on its geometrical features, the shade of its colour in comparison to some other shade of colour *A* presumes *B* to be acquainted with, its function with regard to some other object, etc.

Now consider what happens when *A* shows *B* the object. In case *B* did not know the object prior to the whole episode, she would gain new knowledge of it, i.e. knowledge in addition to the structural knowledge acquired during the telephone conversation. The new knowledge would come in the form of perceptions. In case she did know the object already, she would simply not gain any new knowledge, at least not in the sense of perceiving the object for the very first time. What other kind of knowledge does *B* have of the object in this latter scenario? It is the kind of knowledge she could transmit to *A* if the tables were turned, i.e. structural knowledge. Either way, the knowledge that *A* and *B* seem to have at their disposal is

[246] For example, even though I cannot transmit my perceptual experience of two shades of green, I can transmit the relation that one is darker than the other.

of two kinds: perceptions and relations between these. If we accept the correspondence story given above, this knowledge allows one to infer the abstract structure of the (physical) object in question. *A* and *B* thus end up having a third kind of knowledge, one that concerns the physical world.

The following lengthy passage from Quine, which inspired my own example, explicates the structural character of the linguistic transmission of information:

> Send a man into another room and have him come back and report on its contents. He comes back and agitates the air for a while, and in consequence of this agitation we learn about objects in the other room which are very unlike any agitation of the air. Selected traits of objects in that room are coded in traits of this agitation in the air. The manner of the coding, called language, is complicated and far-fetched, but it works; and clearly it is purely structural, at least in the privative sense of depending on no qualitative resemblances between the objects and the agitation. Also the man's internal state, neural or whatever, in which his knowledge of the objects in that room consists, presumably bears none but structural relations to those objects; structural in the privative sense of there being no qualitative resemblances between the objects and the man's internal state, but only some sort of coding, and, of course, causation. And the same applies to our own knowledge of the objects, as gained from the man's testimony (1968: 161).

In all stages of this process, the information is structural through and through. The private acquisition of knowledge of the objects is structural. The linguistic transmission of this knowledge is structural. The acquisition of knowledge by the testimony of others is structural. Even the sound waves produced in speech encode only structural information about the world.

Taking Stock

To the extent that historical and other considerations support ESR more strongly than any of its rivals, we can say that its epistemic commitments are more warranted. Given the inverse relation between epistemic commitments and underdetermination, it can be argued that we have at least some warrant to believe that the impact of underdetermination on theory choice *can* be restricted in accordance with the epistemic commitments of ESR. Of course, just because we have reason to believe that epistemic structural realism is correct, this does not mean that *any* given scientific theory will not be underdetermined with respect to structure. Whether a particular theory satisfies certain constraints is something that must be judged on an individual basis. Thus, it is only those theories for which we have evidence to believe that they have latched on to the structure of the world that will not be underdetermined with respect to structure.

It is not argued here that this indirect strategy to restrict the impact of underdetermination *solves* the problem. Rather, it is argued that there is some warrant for the view that we can appeal to structural constraints on matters of theory choice. Naturally, if the constructive empiricist version of the underdetermination thesis is indeed correct, then no realist epistemic commitments will be correct, and, correspondingly, the only constraints that theories will be able to satisfy will be observational. Having said that, it is not clear that the constructive empiricist version of the underdetermination thesis holds, for as I pointed out at the end of section two, the notion of a 'genuine rival' remains elusive for both parties in the debate. Although for each body of observational evidence there are infinitely many empirically equivalent theories that diverge on their theoretical claims, it is not clear how many of these, if any, are genuine rivals to existing well-confirmed theories. Similarly with the idea of evidential equivalence, it is not clear whether any genuine competitor theories can be evidentially equivalent.

5. Conclusion

In this chapter, I have addressed some central themes in the debate concerning arguments from the underdetermination of theories by evidence. The starting point was Laudan and Leplin's influential article that criticises underdetermination arguments on two fronts.

First, they question the view that all theories have genuine empirically equivalent rivals. Yet, their main claim, i.e. no algorithms exist that can produce empirically equivalent rivals for any given theory, remains unsubstantiated. To this effect, I offered some concrete algorithms that seem to do the job. Even if no genuine algorithms can be constructed, I argued, it does not follow that some theories have no empirically equivalent rivals. The analogy from recursion theory suggests that such rivals may exist despite our inability to produce them algorithmically.

Second, they argue that, even when theories have such rivals, there are still ways to justifiably choose between them. In support of this argument, they offer two reasons:

(a) that the observational consequences of a theory need not be evidential, and (b) that evidential considerations go beyond the observational consequences of a theory. I argued that even though (a) is correct, it fails to defeat claims of empirical equivalence, since all that it achieves is to trade one empirical equivalence class for another. Similarly, Leplin and Laudan's appeal to the converse consequence condition and the special consequence condition in support of (b) fails to show that evidential considerations go beyond the observational consequences of a theory.

A more promising avenue, not entertained by Leplin and Laudan, towards the claim that the observational consequences of a theory form a proper subset of the evidence for or against it is the appeal to theoretical virtues such as unity, simplicity, and explanatory power. Suppose that it can be shown that these virtues have evidential status. Even then, if two theories are empirically equivalent it does not mean that they will be equally unified, simple, or explanatorily powerful. Unsurprisingly, whether theoretical virtues have such evidential status is a disputed matter. In particular, the anti-realist denies the epistemic significance of these virtues, emphasising instead their pragmatic character.

Avoiding a full-scale discussion of the evidential merits of theoretical virtues, I concentrated instead on the possible ramifications of a 'yes' or 'no' answer. What I discovered is that even if the underdetermination that comes with empirical equivalence claims is defeated by appeal to theoretical virtues, another type of equivalence claims remains. Evidential equivalence is the idea that there is at least one evidentially equivalent rival theory for any given theory. If EVE is true then the realist must come to terms with the idea that theory choice will never be completely resolved, not even in the ideal limit. Indeed, to defeat underdetermination altogether, the realist must sustain a thesis that is logically stronger than the negation of EVE. To this effect, I offered several stronger alternatives to the negation of EVE, i.e. EVE2-4, pointing out some of the difficulties involved in trying to come up with the right formulation.

A suggestion that emerged from the discussion on evidential equivalence was that there is an inverse relation between epistemic commitments and underdetermination. The greater the epistemic commitments of a position in the scientific realism debate, the more rival theories it can potentially eliminate, and, consequently, the more underdetermination can be constrained. In this sense, one can order constructive empiricism, structural realism, and scientific realism on the basis of the increasing constraints they impose on underdetermination. At the end of the day, of course, what matters is the extent to which the epistemic commitments of a viewpoint are justified.

Summarising the results offered in this dissertation, I argued that there are good grounds to choose ESR, and in particular the Russellian variety, over its rivals. To be more precise, I offered two kinds of considerations that seem to give structural realism the edge: (i) historical, and (ii) other.

With respect to historical considerations, I reasoned that structural realism provides a better account of scientific theory change than that offered by the scientific realist. This is evident in the structural continuity we witness in the history of science. Contra van Fraassen’s claim that the continuity of structure takes place only at the level of phenomena, I argued that it is unreasonable to assume that the structure of the phenomena encodes no information about the structure of the physical world. In particular, I argued that the structural correspondence between our actions in the physical world and their observable effects indicates a correspondence between the physical and the phenomenal worlds.

With respect to other considerations, I utilised three arguments that go the longest way in support of structural realism: 1) The argument from the structural source of predictive power rested on the idea that only mathematical structures possess the sharp predictive power necessary to test the epistemic warrant of theories, and thus only they should take the credit. 2) The argument from the limits of mathematical description relied on the fact that scientists employ mathematical objects as surrogates for the objects of physics. Since the latter can only be described up to isomorphism, the former presumably inherit that trait. 3) The argument from the linguistic intransmissibility of non-structure made use of the fact that linguistic communication cannot transmit perceptions but only structure.

Provided that the historical and other considerations support ESR more strongly than any of its rivals, as I believe to have made a case for here, there are reasons to believe that underdetermination can be reduced in accordance with its epistemic commitments. At any rate, there seems to be no convincing reason to think that the constructive empiricist version of the underdetermination thesis holds, and, hence, no convincing reason that ESR is false.

7

SOME PROMISING AVENUES FOR FUTURE RESEARCH

1. Introduction

Over the last six chapters I have sought to evaluate the plausibility of the structural realist answer to the epistemological question set out at the beginning: ‘What kind of knowledge, if any, does science reveal of the physical world?’ I began my evaluation with an analysis of the debate and an outline of the main challenges facing scientific realism. I then proceeded, in chapter two, to a historico-analytical account of structural realism in its various guises. Concentrating on the epistemic variety, I distinguished between two prominent versions, Ramsey-style and Russellian ESR. I completed that chapter by identifying the main challenges to ESR. The next four chapters were then spent addressing these challenges, namely those unique to ESR and those affecting realist positions more generally. Some were answered to my, and I hope the reader’s, satisfaction. Others were left only partly answered, hopefully laying the groundwork for further research on this topic. Others still were explicitly bracketed from the onset.

In this short chapter, I will outline some promising avenues for future research on ESR. The plan is to divide the workload into two manageable sections. First, I will consider certain avenues of research arising from challenges that were taken up in chapters three to six. Second, I will briefly consider ways to develop answers to the challenges logged in the first chapter but not addressed.

2. Outstanding Issues

From Within

Psillos’ remonstration of ESR, as it is reflected in his seven objections listed in chapters two and three, has been shown to be largely unfounded. Having said that, at least one issue in chapter three demands further attention. Russell’s principle that relations between percepts have the same mathematical properties as relations between their non-perceptual causes, needs to be placed on a firmer footing. You may recall, from section four of chapter three, that Russell’s own justification of the principle was muddled and inadequate. I tried to rectify this by offering some reasons why we should accept MR, both in the aforementioned section as well as in section four of chapter six.

Specifically, in chapter three I argued that it is a general commitment of epistemological realism that there be some sort of correspondence between the mental or the linguistic on the one hand and the physical on the other. Since correspondence is not very informative unless it preserves (the mathematical properties of) relations, MR has at least some prima facie plausibility. The trouble is that this correspondence, even if relation-preserving, seems to be merely postulated to exist between the two realms. If we only have direct epistemic access to the perceptual realm, how exactly can we verify the sort of relations it bears to the physical realm?

I made an effort to redress this issue in section four of chapter six. There I argued that the isomorphic correspondence between the perceptual and physical world can be tested in an indirect way. Instrumental to this task is the more familiar correspondence between actions and their observable effects. Since actions are nothing but causes in the physical world, the correspondence between actions (or indeed inactions) and their observable effects is a special instance of the correspondence between causes in the physical world and their observable effects.[247] Although the link between the two correspondences is intuitive, more needs to be said about the isomorphic nature of the correspondence between actions and their observable effects. To this end, one would do well to draw upon the wealth of empirical and theoretical results from the fields of psychology of perception, neuroscience, cognitive science, AI, etc. A project along these lines would throw more light on the trilateral relation between actions, mind, and world.

The discussion of the Newman objection in chapter four raised a couple of issues that I want to pursue further here. One such issue concerns the call made by Worrall

[247] There seems to be no good reason to expect that our actions are different from other causes in the physical world, at least not in a way that is relevant to the current context.

and Zahar, as well as by Ladyman and French though for different reasons, for a new relation-based semantics. According to standard semantics, the interpreted terms of first-order sentences uniquely pick out individuals. ESR advocates cannot accept this, for they hold that we cannot uniquely pick out individuals. One way out for the epistemic structural realist is to resort to notions that work without contravening standard semantics. Notions like abstract structure. Another way out, the one advocated by Worrall and Zahar, is to replace standard semantics with a semantics that takes relations, instead of relata, as primitive. The challenge then for those sympathetic to this proposal is to come up with an independent justification why we should opt for such a radical approach, as opposed to sticking with the more familiar notion of isomorphism. More importantly, the challenge is to deliver a relation-based semantics that is as efficacious and successful as standard semantics. The advent of such semantics would not only solve ESR's conceptual difficulties, that some take it to have, but also potentially revolutionise philosophy.

One of the dissertation's most promising suggestions concerns the relationship between Quine's ideas and structural realism. As we saw in chapter four, Quine advocated a form of structuralism that he thought to be intimately related to the theses of underdetermination and of indeterminacy of reference. I drew parallels between these ideas and structural realism. One implicit parallel takes (in)determinacy of reference to be circumscribed in accordance with the limits of knowledge. From the structural realist perspective, this amounts to the idea that reference can be fixed only up to isomorphism. When I said that objects cannot be uniquely identified but only up to isomorphism, it was implied that *referents* are so identified too. Given the widespread concern realists have about issues of reference, a systematic study of these issues from a structural realist perspective should certainly be high on a list of priorities for future research.

Chapter five addressed the general challenge that realism needs to give a plausible explanation of the upheavals in the historical record of science. In particular, it requires that at least some components of theories, other than observational consequences, survive scientific revolutions, and, furthermore, that only those that survive are responsible for the success of a given theory. Structural realism, as we have seen, tries to answer this challenge in terms of the continuity of structure. This chapter also confronts the more specific challenge that there is insufficient historical evidence for structural continuity through theory change. In reply to this latter challenge, I concluded, following Worrall's suggestion, that appeal must be made to the correspondence principle.

Alas, no all-embracing account of the correspondence principle has yet been formulated. Many people in the debate now believe such an account is unlikely to be found. A plurality of correspondence relations is the order of the day. To this end, I offered 'NC-correspondence', as one such correspondence relation, according to which a structure L' becomes isomorphic to, or approximates, a predecessor L when a parameter in L' is neutralised. I also refined the challenge, stating that the ESR-ist must develop and defend a list of correspondence relations that substantiate the view that not all structures may survive, but *most predictively successful* elements that do survive, either intact or suitably modified (for, example, according to NC), are structural. A list of this kind would mean that more historical cases can be subsumed under a structural realist explanation, and would, therefore, count more decisively in its favour.

In chapter six, I looked at one of the most intractable problems in the scientific realism debate, the underdetermination problem. Formulated in terms of a challenge to the realist, RP1 demands that we are able to choose between empirically equivalent theories. In evaluating the most prominent attempt to overcome RP1 in recent years, that of Laudan and Leplin, I considered equivalence types other than empirical equivalence. I pointed out that even if RP1's demands could be met, that would not necessarily vanquish underdetermination, for other types of equivalences still lurk in the background. Evidential equivalence, for example, is the idea that there is at least one evidentially equivalent rival theory for any given theory, where how well a theory stands up to the evidence is not taken to be merely a question of *entailing* true observational sentences but also involves, roughly speaking, the *way* the theory entails these sentences – an issue that brings in the 'theoretical virtues'.

Even though it does not mean complete victory against underdetermination, having the ability to overcome empirical equivalence would be tantamount to constructive empiricism's defeat. The fact that theoretical virtues can act as a lever against the deadlock of empirical equivalence was briefly mentioned in section three of the same chapter. A lot of research is currently being carried out to show why theoretical virtues should be evidentially relevant in matters of theory choice.[248] The challenge for the structural realist is to show why they should be evidentially relevant *only* to structures. In other words, assuming that theoretical virtues carry evidential weight in matters of theory choice, why should they favour structural realism over traditional scientific realism?

From Without

Some topics deep at the heart of the scientific realism debate have had very little or no exposure in this dissertation. This is true of two of realism's central obstacles, briefly mentioned in the first chapter and explicitly bracketed thereafter. To remind the reader these are:

(RP3) It must be shown, or at least it is preferable to show, why the success of science needs explaining and, furthermore, why scientific realism provides a better explanation than any alternative position.

(RP4) The notions of approximate truth, truthlikeness and verisimilitude need to be given rigorous characterisations. If no adequate formal treatments can be given, as indeed conceded by some realists, more robust informal accounts as well as the reasons why such accounts would work need to be clearly explained.

A lot can be said about the relation of either to ESR. Suitably adapted, the second part of RP3 takes the form 'Why does structural realism provide a better explanation than any alternative position?' To the extent that the success of science needs explaining – an assumption that thereby answers the first part of RP3 positively – structural realism has a pretty strong case. I see no good reason why the same arguments I utilised in the last subsection of section four in chapter six to support ESR cannot also be used here. For example, given that we take historical preservation to be a rough indicator of a theory's success and that the kind of

[248] Elliot Sober, for example, has been spearheading a very influential programme in the philosophy of science that takes the Akaike theorem, named after the statistician who invented it, to be giving us a good estimate of a theory's simplicity based on the way it deals with the evidence.

preservation we witness is essentially structural, structural realism can be said to offer a better historical explanation of the success of science. Similarly, the argument from the structural source of predictive power can be utilised to explain the predictive success of science on the basis of the view that mathematical structures have latched on to the structure of the world.

An issue related to RP3 is inference to the best explanation. This notion has a lot of currency in the practice of science, where inferences are seldom straightforwardly deductive or inductive. Some form of IBE thus seems indispensable. I am inclined to think that IBE can be structuralised as some sort of inference to the best causal structure. That is, the best explanation for a given set of data would have to give a causal account of the underlying structure. In a sense, I have been making tacit use of IBE all through the dissertation. When I claimed in section three of chapter two that 'observational data falls into certain patterns allowing us to discover/postulate relations between observables' what I had in mind was something along these lines. The test for the structural realist is to conceive of a version of IBE that delegates the explanatory power of theories to structures and nothing but structures.

As a final point, I suspect that a characterisation of the notions of approximate truth, truthlikeness, and verisimilitude finds more fertile ground in structural realism than in any other type of realism. For one thing, the position's focus on the mathematical notion of abstract structure makes it easier to provide a formal treatment to measures of truth. Yet, even if the venture of formalisation fails to take off, the fact that the seat of predictive power seems to reside in structures cannot but bolster the belief that informal accounts of the notions of approximate truth, truthlikeness, and verisimilitude will also adopt the structural point of view. I have already talked loosely of structures approximating other structures in my discussion of NC-correspondence. I can only hope that a methodical enquiry into the link between approximate truth, truthlikeness and verisimilitude on the one hand, and structural realism on the other, will eventually bear fruit.

3. Conclusion

In light of the above assessment, I want to identify four topics that appear to me the most alluring as avenues for future research. Each of these is accompanied by a few sample questions that I hope will whet the reader's appetite.

(1) Theory of Reference: Does structural realism dictate any particular theory of reference? If reference can only be fixed up to isomorphism, what does this entail for the conditions of successful reference?

(2) The Correspondence Principle: If there is no all-embracing correspondence principle, what are the conditions of adequacy for each individual type of correspondence? How can we justify the claim that correspondence relations are neither trivially satisfiable nor socially constructed? Are all types of correspondence supportive of epistemic structural realism?

(3) Inference to the Best Explanation: Can a structural version of IBE be offered? Something in the ballpark of inference to the best structural explanation, for instance. If so, would this structural version of IBE perform without loss of vital inferential powers to the practice of science?

(4) Approximate Truth: Can we get a handle on the distance between theory and truth solely in virtue of structures? How are we to understand claims about one structure approximating another?

I trust that research into these four topics will provide ample data to peruse, in addition to the data provided in this dissertation, thereby facilitating a more informed judgment about the merits and shortcomings of epistemic structural realism. If I may be forgiven for the potential irreverence to the ghost of William of Occam, I will finish with an apt adjustment of his dictum that is in line with structural realism:

Do not multiply entities beyond the limits of structural isomorphism!

Bibliography

Achinstein, P. (1994) ‘Explanation v. Prediction: Which Carries More Weight?’, in D. Hull, M. Forbes and R.M. Burian (eds.) *PSA 1994*, vol. 2, East Lansing, MI: Philosophy of Science Association, pp. 156-164.

Aronson, J. L. and Harré, R. and Way, E. (1994) *Realism Rescued*: *How Scientific Progress is Possible*, London: Duckworth.

Aune, B. (1967) *Knowledge*, *Mind*, *and Nature*: *An Introduction to a Theory of Knowledge and the Philosophy of Mind*, New York: Random House.

Bain, J. (1998) *Representations of Spacetime: Formalism and Ontological Commitment*, unpublished Thesis, University of Pittsburgh.

Beloff, J. (1962) *The Existence of Mind*, London : MacGibbon & Kee.

Black, J. (1803) *Lectures on the Elements of Chemistry*, published from his manuscripts by John Robinson, Edinburgh: Mundell and Son.

Boerhaave, H. (1732) *Elementa Chemiae*, Lugduni Batavorum.

Bogen, J. and Woodward, J. (1988) ‘Saving the Phenomena’, *The Philosophical Review*, vol. 97(3): 303-352.

Boghossian, P. and Peacocke, C. (eds.) (2001) *New Essays on the a Priori*, Oxford: Oxford University Press.

Boyd, R. (1981) ‘Scientific Realism and Naturalistic Epistemology’, in P.D. Asquith and T. Nickles (eds.) *PSA 1980*, vol. 2, East Lansing, MI: Philosophy of Science Association.

Boyd, R. (1984) ‘The Current Status of Scientific Realism’ in J. Leplin (ed.) *Scientific Realism*, Berkeley: University of California Press, pp. 41-82.

Boyd, R. (1990) *Realism, Approximate Truth, and Philosophical Method in Scientific Theories*, Minneapolis: University of Minnesota Press.

Boyd, R. (2002) ‘Scientific Realism’, *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/ archives/sum2002/entries/scientific-realism/.

Bradie, M.P. (1977) ‘The Development of Russell’s Structural Postulates’, *Philosophy of Science*, vol. 44: 441-463.

Braithwaite, R.B. (1940) ‘The Philosophy of Physical Science’, *Mind*, vol. 49, no. 196: 455-466.

Brown, J.R. (1994) *Smoke and Mirrors*: *How Science Reflects Reality*, London: Routledge.

Brush, S. (1970) ‘The Wave Theory of Heat: A Forgotten Stage in the Transition from the Caloric Theory to Thermodynamics’, *British Journal for the History of Science*, vol. 5: 145-167.

Brush, S. and Holton, G. ([1952]2001) *Physics, The Human Adventure*: *From Copernicus to Einstein and Beyond*, 3rd edition, New Brunswick, NJ: Rutgers University Press.

Bueno, O. (1997) ‘Empirical Adequacy: A Partial Structures Approach’, *Studies in History and Philosophy of Science*, 28A(4): 585-610.

Bueno, O. (1999) ‘What Is Structural Empiricism? Scientific Change in an Empiricist Setting’, *Erkenntnis*, 50(1), 59-85.

Bueno, O. (2000) ‘Empiricism, Scientific Change and Mathematical Change’, *Studies in History and Philosophy of Science*, 31A(2): 269-296.

Cao, T.Y. (2003a) ‘Structural Realism and the Interpretation of Quantum Field Theory’, *Synthese*, vol. 136(1): 3-30.

Cao, T.Y. (2003b) ‘Can We Dissolve Physical Entities into Mathematical Structures?’, reply to French and Ladyman, *Synthese*, vol. 136(1): 57-71.

Carnap, R. (1928) *Der Logisches Aufbau der Welt*, Berlin: Schlachtensee Weltkreis-Verlag.

Carnap, R. (1956) ‘The Methodological Character of Theoretical Concepts’, in H. Feigl and M. Scriven (eds.) *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minnesota Studies in the Philosophy of Science, vol. 1, Minneapolis: University of Minnesota Press.

Carnot, S. (1824) ‘Reflections on the Motive Power of Fire’, in E. Mendoza (ed.) *Reflections on the Motive Power of Fire by Sadi Carnot and other Papers on the Second Law of Thermodynamics by E. Clapeyron and R. Clausius*, New York: Dover Publications, 1960.

Cartwright, N. (1980) ‘Do the Laws of Physics State the Facts?’, reprinted in Curd, M. and Cover, J.A. (eds.) *Philosophy of Science*: *The Central Issues*, New York: W.V. Norton & Company, 1998, pp. 865-77.

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

Cartwright, N. (1998) ‘How Theories Relate: Takeovers or Partnerships?’, *Philosophia Naturalis* 35, pp. 23-34.

Cassirer, E. (1936) *Determinism and Indeterminism in Modern Physics*, Yale University Press.

Cassirer, E. (1944) ‘Group Concept and Perception Theory’, *Philosophy and Phenomenological Research*, vol. 5: 1-36.

Chakravartty, A. (1998) ‘Semirealism’, *Studies in History and Philosophy of Science*, vol. 29A(3): 391-408.

Chakravarrty, A. (2003) ‘The Structuralist Conception of Objects’, *Philosophy of Science*, vol. 70(5): 867-878.

Chang, H. (2003) ‘Preservative Realism and Its Discontents: Revisiting Caloric’, *Philosophy of Science*, vol. 70(5): 902-912.

Chang, H. (forthcoming) ‘Thermal Physics Before Thermodynamics’, *Enciclopedia Italiana, Storia Della Scienza*, vol. VI-VII: Science in the Nineteenth Century.

Churchland, P.M. (1982) ‘The Anti-Realist Epistemology of van Fraassen’s “The Scientific Image” ’, *Pacific Philosophical Quarterly*, vol. 63: 226-235.

Churchland, P.M. and Hooker, C.A. (eds.) (1985) *Images of Science*, Chicago: University of Chicago Press.

Clausius, R. (1850) ‘On the Motive Power of Heat, and the Laws which can be Deduced from it for the Theory of Heat’, in E. Mendoza (ed.) *Reflections on the Motive Power of Fire by Sadi Carnot and other Papers on the Second Law of Thermodynamics by E. Clapeyron and R. Clausius*, New York: Dover Publications, 1960.

Creath, R. (1998) ‘Carnap, Rudolf’ in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*, London: Routledge. Retrieved from http://www.rep.routledge.com/ article/DD012SECT1.

Cumminskey, D. (1992) ‘Reference Failure and Scientific Realism: A Response to the Meta-Induction’, British Journal for the Philosophy of Science, vol. 43: 21-40.

Curd, M. and Cover, J.A. (eds.) (1998) *Philosophy of Science*: *The Central Issues,* New York: W.V. Norton & Company.

Da Costa, N. C. A. and French, S. (1990) ‘The Model-Theoretic Approach in the Philosophy of Science’, *Philosophy of Science*, vol. 57: 248–65.

David, M. (2002) ‘The Correspondence Theory of Truth’, *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/sum2002/entries/truth-correspondence/>.

Davy, H. (1799) ‘An Essay on Heat, Light, and the Combinations of Light’, *The Collected Works of H. Davy*, vol. 2: 1-86, repr. New York: Johnson Reprint Corporation, 1972.

Demopoulos, W. and Friedman M. (1985) ‘Critical Notice: Bertrand Russell’s *The Analysis of Matter*: Its Historical Context and Contemporary Interest’, *Philosophy of Science* 52: 621-639.

Devitt, M. (1984) *Realism and Truth*, Oxford: Blackwell.

Duhem, P. ([1914] 1991) *The Aim and Structure of Physical Theory*, Princeton (NJ): Princeton University Press.

Eddington, A.S. (1939) *The Philosophy of Physical Science*, Cambridge University Press.

Eddington, A.S. (1941) ‘Group Structure in Physical Science’, *Mind*, vol. 50, no. 199: 268-279.

English, J. (1973) ‘Underdetermination: Craig and Ramsey’, *Journal of Philosophy* 70: 453-462.

Feyerabend, P.K. (1962) ‘Explanation, Reduction and Empiricism’, in H. Feigl and G. Maxwell (eds.) *Scientific Explanation, Space, and Time*, vol. 3, *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, 28–97.

Feyerabend, P.K. (1965) ‘Problems of Empiricism’, in R.G. Colodny (ed.) *Beyond the Edge of Certainty*, Englewood Cliffs, NJ: Prentice Hall, pp. 145–260.

Fine, A. (1984) ‘The Natural Ontological Attitude’, in J. Leplin (ed.) *Scientific Realism*, Berkeley: University of California Press, pp. 83-107.

Fine, A. (1998) ‘Scientific realism and antirealism’, in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*, London: Routledge. Retrieved from http://www.rep.routledge.com/article/Q094SECT2.

Fox, R. (1971) *The Caloric Theory of Gases*: *From Lavoisier to Regnault*, Oxford: Clarendon Press.

French, S. (1998) ‘On the Withering Away of Physical Objects’, in E. Castellani (ed.), *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, Princeton University Press, pp. 93-113.

French, S. (1999) ‘Models and Mathematics in Physics: The Role of Group Theory’, in J. Butterfield and C. Pagonis (eds), *From Physics to Philosophy*, Cambridge University Press, pp. 187-207.

French, S. (2003) ‘Scribbling on the Blank Sheet: Eddington’s Structuralist Conception of Objects’, *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, vol. 34(2): 227-259.

French, S. and Kamminga, H. (1993) *Correspondence*, *Invariance and Heuristics*: *Essays in Honour of Heinz Post*, Boston Studies in the Philosophy of Science, vol. 148, Dordrecht: Kluwer Academic Press.

French, S. and Ladyman, J. (2003a) ‘Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure’, *Synthese*, vol. 136: 31-56.

French, S. and Ladyman, J. (2003b) ‘The Dissolution of Objects: Between Platonism and Phenomenalism’, *Synthese,* vol. 136: 73-77.

Friedman, M. (2001) ‘Transcendental Philosophy and A Priori Knowledge: A Neo-Kantian Perspective’, Boghossian et al., pp. 367-383.

Gettier, E. (1963) ‘Is Justified True Belief Knowledge?’, *Analysis*, vol. 23: 121-123.

Giere, R. N. (1988) *Explaining Science*: *A Cognitive Approach*, Chicago: University of Chicago Press.

Giedymin, J. (1982) *Science and Convention: Essays on Henri Poincaré's Philosophy of Science and the Conventionalist Tradition*, Oxford: Pergamon.

Gillies, D. (1993) *Philosophy of Science in the Twentieth Century*, Oxford: Blackwell Publishers.

Glymour, C. (1980) *Theory and Evidence*, Princeton: Princeton University Press.

Goodman, N. (1965) *Fact, Fiction and Forecast*, Indianapolis, IN: Bobbs-Merrill.

Gower, B.S. (2000) ‘Cassirer, Schlick and “Structural” Realism: The Philosophy of the Exact Sciences in the Background to Early Logical Empiricism’, *British Journal for the History of Philosophy*, vol. 8(1): 71-106.

Gutting, G. (1983) ‘Scientific Realism vs. Constructive Empiricism: A Dialogue’, *Monist*, vol. 65: 336-349.

Guyton de Morveau, L.B., Lavoisier, A., Berthollet, C.L., and Fourcroy, A.F. (1787) *Méthode de Nomenclature Chimique*, Paris, translated by James St John, 1788.

Hacking, I. (1982) 'Experimentation and Scientific Realism', reprinted in Curd, M. and Cover, J.A. (eds.) *Philosophy of Science*: *The Central Issues*, New York: W.V. Norton & Company, 1998, pp. 1153-68.

Hanson, N.R. (1958) *Patterns of Discovery*: *An Inquiry into the Conceptual Foundations of Science*, Cambridge: Cambridge University Press.

Hardin, C.L. and Rosenberg, A. (1982) 'In Defence of Convergent Realism', *Philosophy of Science*, vol. 49: 604-615.

Harman, G. (1965) 'The Inference to the Best Explanation', *Philosophical Review*, vol. 74: 88-95.

Harré, R. (1988) *Varieties of Realism*: *A Rationale for the Natural Sciences*, Oxford: Blackwell.

Hartley, H. (1971) *Studies in the History of Chemistry*, Oxford: Clarendon Press.

Hartmann, S. (2002) Essay Review 'On Correspondence', *Studies in History and Philosophy of Modern Physics*, vol.33: 79-94.

Hertz, H. (1893) 'Maxwell's theory is the system of Maxwell's equations' in *Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity through Space*, translated by D. E. Jones, New York: Dover, 1962.

Hilbert, D. (1998) 'Colour, theories of', in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved February 10, 2004, from http://www.rep.routledge.com/article/Q013SECT1.

Hilpinen, R. (1976) ‘Approximate Truth and Truthlikeness’, in M. Przelecki, et al (eds.) *Formal Methods in the Methodology of the Empirical Sciences*, Dordrecht: Reidel.

Hoefer, C. and Rosenberg, A. (1994) ‘Empirical Equivalence, Underdetermination, and Systems of the World’, *Philosophy of Science*, vol. 61(4): 592-607.

Howson, C. (2001) *Induction: Hume's problem*, Oxford : Clarendon.

Howson, C. and Urbach, P. (1996) *Scientific Reasoning*: *The Bayesian Approach*, 2nd edition, Chicaco and La Salle (IL): Open Court.

Hume, D. ([1739]1975) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press.

Ketland, J. (forthcoming) ‘Empirical Adequacy and Ramsification’, *British Journal for the Philosophy of Science*.

Kirkham, R.L. (1992) *Theories of Truth: A Critical Introduction*, Cambridge, Mass: MIT Press.

Kitcher, P. (1993) *The Advancement of Science*, Oxford: OUP.

Kitcher, P. (2001) ‘Real Realism: The Galilean Strategy’, *Philosophical Review*, vol. 110(2): 151-197.

Kondepudi, D. and Prigogine, I. (1998) *Modern Thermodynamics*: *From Heat Engines to Dissipative Structures*, Chichester: Wiley.

Kosso, P. (1998) ‘Observation’, in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*, London: Routledge. Retrieved from: http://www.rep.routledge.com/article/Q076

Krajewski, W. (1977) *Correspondence Principle and Growth of Science*, Dordrecht: D. Reidel Pub. Co.

Kuhn, T. ([1962]1996) *The Structure of Scientific Revolutions*, third edition, Chicago: University of Chicago Press.

Ladyman, J. (1998) 'What is Structural Realism?', *Studies in History and Philosophy of Science*, vol. 29: 409-424.

Lakatos, I. (1970) 'Falsificationism and the Methodology of Scientific Research Programmes', in I. Lakatos and A.E. Musgrave (eds.) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.

Laudan, L. (1977) *Progress and its Problems*: *Toward a Theory of Scientific Growth*, Berkeley: University of California Press.

Laudan, L. (1981) 'A Confutation of Convergent Realism', *Philosophy of Science*, vol. 48(1): 19-49.

Laudan, L. (1984) 'Realism without the Real', *Philosophy of Science*, vol. 51: 156-162.

Laudan, L. and Leplin, J. (1991) 'Empirical Equivalence and Underdetermination', Journal of Philosophy, vol. 88: 449–72.

Lavoisier, A. (1783) 'Réflexions sur le Phlogistique pour Servir de Suite à la Théorie de la Combustion et de la Calcination', *Mémoires de l'Académie des Sciences*, Paris, repr. in D. McKie, *Antoine Lavoisier*, New York: Collier Books, 1962.

Lavoisier, A. (1789) *Traité Élémentaire de Chimie*, translated by R. Kerr, as *Elements of Chemistry*, repr. New York : Dover Publications, 1965.

Lavoisier, A. (1805) *Mémoires de Chimie*, Mme Lavoisier (ed.), Paris.

Lavoisier, A. and Laplace, P. S. (1780) 'Mémoire sur la Chaleur', *Mémoires de mathématique et de physique, tirés des registres de l'Académie des Sciences*, Paris, pp. 355-408, translated by H. Guerlac, as *Memoir on Heat*, 1982.

Leplin, J. (ed.) (1984) *Scientific Realism*, Berkeley: University of California Press.

Leplin, J. (1997) *A Novel Defence of Scientific Realism*, Oxford: Oxford University Press.

Leplin, J. (2000) Review of Stathis Psillos' *Scientific Realism*: *How Science Tracks Truth*, *Mind*, vol. 109 (436): 980-983.

Lipton, P. (1991) *Inference to the Best Explanation*, London: Routledge.

Lyons, T.D. (2002) 'Scientific Realism and the Pessimistic Meta-Modus Tollens', in S. Clarke and T.D. Lyons (eds.) *Recent Themes in the Philosophy of Science*, Dordrecht: Kluwer Academic Publishers.

Mandelbaum, M. (1964) *Philosophy, Science, and Sense Perception*: *Historical and Critical Studies*, Baltimore: John Hopkins Press.

Mayo, D. (1991) 'Novel Evidence and Severe Tests', *Philosophy of Science*, 58: 523-552.

McKie, D. (1962) *Antoine Lavoisier*, New York: Collier Books.

McKie, D. and de V. Heathcote, N.H. (1935) *The Discovery of Specific and Latent Heats*, London: E. Arnold & co.

McLendon, H.J. (1955) 'Uses of Similarity of Structure in Contemporary Philosophy', *Mind*, vol. 64: 79-95.

McMullin, E. (1984) 'A Case for Scientific Realism', in J. Leplin (ed.) *Scientific Realism*, Berkeley: University of California Press.

Maxwell, G. (1962) ‘The Ontological Status of Theoretical Entities’, in H. Feigl and G. Maxwell (eds.) *Scientific Explanation, Space, and Time*, vol. 3, *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, 3-15.

Maxwell, G. (1968) ‘Scientific Methodology and the Causal Theory of Perception’, in I. Lakatos and A. Musgrave (eds.) *Problems in the Philosophy of Science*, Amsterdam: North-Holland Publishing Company.

Maxwell, G. (1970a) ‘Structural Realism and the Meaning of Theoretical Terms’, in S. Winokur and M. Radner (eds.) *Analyses of Theories, and Methods of Physics and Psychology*, Minneapolis: University of Minnesota Press, 181-192.

Maxwell, G. (1970b) ‘Theories, Perception and Structural Realism’ in R. Colodny (ed.) *Nature and Function of Scientific Theories*, Pittsburgh: University of Pittsburgh Press, 3-34.

Miller, D. (1974) ‘Popper’s Qualitative Theory of Verisimilitude’, *British Journal for the Philosophy of Science*, vol. 25: 166-177.

Montserrat J.M. and Navarro L. (2000) ‘The Atomistic View of Heat in Lucretius’, *Centaurus*, vol. 42(1): 1-20.

Morgan, M. and Morrison, M. (eds.) (1999) *Models as Mediators*: *Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press.

Morganti, M. (forthcoming) ‘On the Preferability of Epistemic Structural Realism’, *Synthese*.

Morris, R. (1972) ‘Lavoisier and the Caloric Theory’, *British Journal for the History of Science*, vol. 6: 1-38.

Musgrave, A. (1985) ‘Realism Versus Constructive Empiricism’ in P.M. Churchland and C.A. Hooker (eds.) *Images of Science*, Chicago: University of Chicago Press.

Nelson, D. (1996) 'Confirmation, Explanation, and Logical Strength', British Journal for the Philosophy of Science, 47(3): 399-413.

Newman, M.H.A. (1928) ‘MR. Russell’s “Causal Theory of Perception” ’, *Mind*, vol. 37: 137-148.

Newton, I. ([1726]1999) *Mathematical Principles of Natural Philosophy*, trans. by I. B. Cohen and A. Whitman, Berkeley: University of California Press.

Newton-Smith, W.H. (1981) *The Rationality of Science*, London: Routledge and Kegan Paul.

Newton-Smith, W. H. (1989) ‘Modest Realism’ in A. Fine & J. Leplin (eds.) *PSA 1988*, Vol.2, East Lansing: Philosophy of Science Association.

Niiniluoto, I. (1987) *Truthlikeness*, Dordrecht: Reidel.

Niiniluoto, I. (1999) *Critical Scientific Realism*. Oxford: Oxford University Press.

Oddie, G. (1986) *Likeness to Truth*, Dordrecht: Reidel.

Okasha, S. (1997) ‘Laudan and Leplin on Empirical Equivalence’, *British Journal for the Philosophy of Science*, vol. 48(2): 251-256.

Papineau, D. (1987) *Reality and Representation*, Oxford: Blackwell Publishers.

Pepper, S.C. (1967) *Concept and Quality*: *A World Hypothesis*, La Salle, Illinois: Open Court.

Poincaré, H. ([1905]1952) *Science and Hypothesis*, New York: Dover.

Poincaré, H. ([1913]1946) *The Value of Scienc*e, translated by George B. Halsted, in H. Poincaré, *The Foundations of Science: Science and Hypothesis, The Value of Science, and Science and* Method, Lancaster, PA: The Science Press.

Popper, K. (1963) *Conjectures and Refutations*: *The Growth of Scientific Knowledge*, London: Routledge and Keagan Paul.

Post, H.R. (1971) 'Correspondence, Invariance and Heuristics', *Studies in the* History and Philosophy of Science, vol. 2(3), pp. 213-55.

Psillos, S. (1994) 'A Philosophical Study of the Transition from the Caloric Theory of Heat to Thermodynamics: Resisting the Pessimistic Meta-Induction', *Studies in the History and Philosophy of Science*, vol. 25: 159-190.

Psillos, S. (1995) 'Is Structural Realism the Best of Both Worlds?' *Dialectica*, 49: 15-46.

Psillos, S. (1996) '*Scientific Realism and the Pessimistic Induction*', PSA 1996, vol. 3: S306-S314.

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

Psillos, S. (2000a) 'Carnap, the Ramsey-Sentence and Realistic Empiricism' *Erkenntnis*, vol. 52(2): 253-279.

Psillos, S. (2000b) 'The Present State of the Scientific Realism Debate', *British Journal for the Philosophy of Science*, vol. 51 (Supp): 705-728.

Psillos, S. (2001a) 'Is Structural Realism Possible?', *Philosophy of Science* (Supplement), vol. 68: S13-24.

Psillos, S. (2001b) Author's Response to Symposium on *Scientific Realism: How Science Tracks Truth*, *Metascience*, vol. 10(3): 366-371.

Putnam, H. (1975) *Mathematics, Matter and Method*, vol. 1, Cambridge: Cambridge University Press.

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

Purtill, R. (1967) Kuhn on Scientific Revolutions, *Philosophy of Science*, vol. 34(1): 53-58.

Quine, W.V.O. (1951) 'Two Dogmas of Empiricism', *Philosophical Review*, vol. 60: 20-43.

Quine, W.V.O. (1968) 'Comment' in the Discussion section of Maxwell's 'Scientific Methodology and the Causal Theory of Perception', in I. Lakatos and A. Musgrave (eds.) *Problems in the Philosophy of Science*, Amsterdam: North-Holland Publishing Company.

Quine, W.V.O. (1969), *Ontological Relativity and Other Essays*, New York: Columbia University Press.

Quine, W.V.O. (1975) 'On Empirically Equivalent Systems of the World', *Erkenntnis*, vol. 9: 313-328.

Quine, W.V.O. (1981) *Theories and Things*, Cambridge, Mass.: Harvard University Press.

Quine, W.V.O. (1992) 'Structure and Nature', *Journal of Philosophy*, vol. 89(1): 5-9.

Quine, W.V.O. ([1995] 1998) *From Stimulus to Science*, second edition, Cambridge, Mass.: Harvard University Press.

Redhead, M.L.G. (2001a) 'The Intelligibility of the Universe', in A.O'Hear (ed.) *Philosophy at the New Millennium*, Cambridge: Cambridge University Press, pp. 73-90.

Redhead, M.L.G. (2001b) ‘Quests of a Realist’, Review article of Stathis Psillos's Scientific Realism: How Science Tracks Truth, Metascience, vol. 10(3): 341-347.

Resnik, M.D. (1981) ‘Mathematics as a Science of Patterns: Ontology and Reference’, *Nous*, vol 15(4): 529-550.

Roller, D. (1950) *The Early Development of the Concepts of Temperature and Heat*: *The Rise and Decline of the Caloric Theory*, Cambridge: Harvard University Press.

Russell, B. (1912) *The Problems of Philosophy*, Oxford: Oxford University Press.

Russell, B. (1927) *The Analysis of Matter*, London: George Allen & Unwin.

Russell, B. (1940) *An Inquiry into Meaning and Truth*, London: Allen & Unwin.

Russell, B. (1948) *Human Knowledge: Its Scope and Limits*, London: George Allen & Unwin.

Russell, B. (1968) *The Autobiography of Bertrand Russell*, vol. 2, London: George Allen & Unwin.

Salmon, W. (1994) ‘Comment: Carnap on Realism’, in W. Salmon and G. Wolters (eds.) *Logic, Language and the Structure of Scientific Theories*, Pittsburgh: University of Pittsburgh Press.

Salmon, W. (1998) *Causality and Explanation*, Oxford: Oxford University Press.

Schlick, M. (1925) *General Theory of Knowledge*, translated by A.E. Blumberg and H. Feigl, New York: Springer-Verlag.

Shapere, D. (1964) Book Review: *The Structure of Scientific Revolutions*, *Philosophical Review*, vol. 73: 383-394.

Smart, J.J.C. (1963) *Philosophy and Scientific Realism*, New York: Humanities Press.

Smith, P. (1998) ‘Approximate Truth and Dynamical Theories’, *British Journal for the Philosophy of Science*, vol. 49: 253-277.

Solomon, G. (1989) ‘An Addendum to Demopoulos and Friedman (1985)’, *Philosophy of Science*, vol.56: 497-501.

Stanford, K. (2003) ‘No Refuge for Realism: Selective Confirmation and the History of Science’, *Philosophy of Science*, vol. 70(5): 913-925.

Stump, D. (1989) ‘Henri Poincaré’s Philosophy of Science’, *Studies in History and Philosophy of Science*, vol. 20(3): 335-363.

Thomson, B. (1798) ‘An Inquiry Concerning the Source of the Heat which is Excited by Friction’, *Philosophical Transactions of the Royal Society of London*, vol. 88: 80-102.

Tichý, P. (1974) ‘On Popper’s Definitions of Verisimilitude’, *British Journal for the Philosophy of Science*, vol. 25: 155-160.

Tichý, P. (1978) ‘Verisimilitude Revisited’, *Synthese*, vol. 38: 175-196.

Tyndall, J. (1863) *Heat Considered as a Mode of Motion*, London: Longman Green.

Van Fraassen, B. C. (1980) *The Scientific Image*, Oxford: Clarendon Press.

Van Fraassen, B. C. (1985) ‘Empiricism in Philosophy of Science’ in P.M. Churchland & C.A. Hooker (eds.) *Images of Science*, Chicago: The University of Chicago Press.

Van Fraassen, B.C. (1997) 'Structure and Perspective: Philosophical Perplexity and Paradox' in M.L. Dalla Chiara et al. (eds.) *Logic and Scientific Methods*, Dordrecht: Kluwer Academic Press, 1997.

Van Fraassen, B.C. (1999) 'Structure: Its Shadow and Substance', *PhilSci Archive*, at http://philsci-archive.pitt.edu/archive/00000631/

Votsis, I. (2003a) Review of Ilkka Niiniluoto's *Critical Scientific Realism*, *Philosophy of Science*, vol.70(2): 444-447.

Votsis, I. (2003b) 'Is Structure Not Enough?', *Philosophy of Science*, vol. 70(5): 879-890.

Warwick, A. (2002) *Masters of Theory*: C*ambridge and the Rise of Mathematical Physics*, Chicago: University of Chicago Press.

Weyl, H. (1963) *Philosophy of Mathematics and Natural Science*, NY: Atheneum.

Worrall, J. (1982) 'Scientific Realism and Scientific Change', *Philosophical Quarterly*, vol. 32(128): 201-231.

Worrall, J. (1984) 'An Unreal Image', *British Journal for the Philosophy of Science*, vol. 35: 65-80.

Worrall, J. (1985) 'Scientific Discovery and Theory-Confirmation' in *Change and Progress in Modern Science*, ed. Joseph C. Pitt, Dordrecht, Netherlands: D. Riedel.

Worrall, J. (1989) 'Structural Realism: The Best of Both Worlds?' in Papineau, D. (ed.) *The Philosophy of Science*, Oxford: Oxford University Press, 1996.

Worrall, J. (1994) ‘How to Remain (Reasonably) Optimistic: Scientific Realism and the “Luminiferous Ether” ’, in D. Hull, M. Forbes and R.M. Burian (eds.) *PSA 1994*, vol. 1, East Lansing, MI: Philosophy of Science Association, pp. 334-342.

Worrall, J. (2000) ‘Miracles and Models: Saving Structural Realism?’, paper given to the Annual Meeting of the British Society for the Philosophy of Science, Sheffield, 2000.

Worrall, J. (2002) ‘What Evidence in Evidence-Based Medicine?’, *Philosophy of Science*, vol. 69(3): S316-S330.

Worrall, J. and Zahar, E. (2001) ‘Ramseyfication and Structural Realism’, Appendix IV in Zahar E. *Poincaré*’*s Philosophy*: *From Conventionalism to Phenomenology*, Chicago and La Salle (IL): Open Court.

Zahar, E. (1973) ‘Why did Einstein's Programme Supersede Lorentz’s?’, *British Journal for the Philosophy of Science,* vol. 24: 95-123.

Zahar, E. (1996) ‘Poincaré’s Structural Realism and his Logic of Discovery’, in Jean-Louis Greffe, Gerhard Heinzmann and Kuno Lorenz (eds.) *Henri Poincaré: Science and Philosophy*, Berlin: Academie Verlag and Paris: Albert Blanchard.

Zahar, E. (2001) *Poincaré’s Philosophy*: *From Conventionalism to Phenomenology*, Chicago and La Salle (IL): Open Court.