Structural Realism: A Critical Appraisal

Peter M. Ainsworth

London School of Economics and Political Science

Ph.D Thesis

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis is my own work.

Peter M. Tinsworth

Abstract

Three principal forms of structural realism are distinguished in this thesis: weak epistemic structural realism (WESR), strong epistemic structural realism (SESR) and ontic structural realism (OSR). In chapter 1, it is argued that the positive argument in favour of WESR, i.e. the claim that it can accommodate the no miracles argument and the pessimistic induction is unconvincing, because (i) the no miracles argument is flawed, so it is no particular virtue of WESR that it can accommodate it and (ii) it is not clear that WESR really can accommodate the pessimistic induction. In chapter 2, it is argued that there are unresolved difficulties in drawing the observable/unobservable distinction (or an appropriate alternative distinction) in a way that is suitable for the WESRist's purposes. In chapter 3, it is argued that the main argument for SESR is unconvincing, because it is based on Russell's principle of acquaintance (or a modern variant of this principle), a principle for which no substantial argument has been given and which has absurd consequences (as shown in appendix 2). In chapter 4, it is argued that neither the WESRist nor the SESRist has provided a convincing response to Newman's objection. In chapter 5, it is argued that, depending on how one interprets the doctrine, OSR is either (i) wholly untenable or (ii) conventional scientific realism (or antirealism) combined with the traditional metaphysical view that objects are bundles of properties and that in the latter case the doctrine has some plausibility, but the arguments that have been adduced in favour of it are inconclusive. The thesis is concluded with a sketch of the sort of position in the scientific realism debate that I find more attractive.

Contents

Introduction

The first fully elaborated statement of structural realism can be found in the work of Russell (1912, 1927) (although Worrall [1989, 1994] argues that the doctrine can also be found in the writings of Duhem [1906] and Poincaré [1903]). Russell himself abandoned the position in the face of Newman's objection (Newman, 1928). The doctrine was revived four decades later by Maxwell (1968, 1970a, 1970b), who both coined the term, "structural realism" and introduced the Ramsey-sentence approach to the doctrine, which is adopted by most contemporary structural realists.[1] There was, however, another lull in the interest in structural realism throughout most of the 1970s and 1980s. Contemporary interest in the doctrine can be traced back to Worrall (1989, 1994) who argued that it allows one to reconcile the two most powerful arguments in the scientific realism debate: the "pessimistic induction" (also known as the pessimistic metainduction) (in favour of scientific antirealism) and the "no miracles argument" (in favour of scientific realism). Over the last ten years or so numerous positions calling themselves "structural realism" have emerged, most notably the "ontic structural realism" first proposed by Ladyman (1998).

The first aim of this thesis is to distinguish the different types of structural realism that have been proposed and provide a clear statement of each.[2] This work is carried out, in outline, in this introduction. The question addressed in the main body of the thesis is whether or not any of these forms of structural realism is tenable.

---

[1] As we shall see, Russell's approach to structural realism is similar to, but distinct from, the Ramsey-sentence approach.

[2] Chakravartty (1998, 2004) has put forward a position he calls "semirealism", which he maintains is a form of structural realism. His position is discussed in appendix 1 (which is best read after chapter 1), where it is argued that although semirealism has some of the characteristics of structural realism it is significantly different from the versions of structural realism that are discussed in the main body of this thesis. Other doctrines that are still more distantly related to structural realism, e.g. "structural empiricism" (see, for example, Bueno, 1999 and van Fraassen, 2006) are not discussed at all in this thesis, as their relation to the doctrines that are discussed herein is considered too distant.

1. ESR and OSR

Structural realism comes in two main forms: epistemic structural realism (ESR) (discussed in chapters 1 to 4 of this thesis) and ontic structural realism (OSR) (discussed in chapter 5 of this thesis).

The ESRist upholds two main theses. On the one hand there is the "realist" thesis:

Mature scientific theories provide us with a substantial amount of knowledge about both the observable and the unobservable world.[3]

On the other hand there is the "structuralist" thesis, which, as a *very rough* first approximation, we may state as follows:

All we *know* of the unobservable world is its structure.

The OSRist upholds the same realist thesis, but a different structuralist thesis, which (again as a very rough first approximation) we may state as follows:

The world *is* a structure.

The ESRist's position will be further explained in section 2 of this introduction. Exactly what the OSRist's position really amounts to is not further discussed until chapter 4.[4]

---

[3] The qualification "mature" is important: the ESRist, like the conventional realist, does not commit himself to realism with respect to all scientific theories. Moreover, the "knowledge" that science is supposed to provide is taken to include claims that are not strictly true, but only approximately true. Spelling out what maturity and approximate truth are may not be easy (although Worrall [1989] has argued that it is straightforward to explicate maturity: he suggests that a theory is mature if and only if it correctly predicts an empirically confirmed result that it was not engineered to yield) but these issues will be held in abeyance throughout this thesis.

2. Ramsey-Sentences and ESR

From a formal point of view, languages are built out of two types of term:[5] logical terms and non-logical terms. In a language of second-order logic these two groups consist of the following:

[1] Logical terms:
(i) logical connectives ("¬", "&", etc.)
(ii) quantifiers ("∀", "∃")
(iii) individual and predicate variables (x_1, x_2,... and X_1, X_2...)
and possibly:
(iv) the identity predicate ("=")

[2] Non-logical terms:
(i) a number of names (a_1, a_2...) denoting objects and
(ii) a number of predicates (P_1, P_2...) denoting properties and relations.

The claim that "all we know of the external world is its structure" might suggest that the ESRist thinks that our knowledge of the external world is purely structural (that it consists of only logical terms). Despite the way they sometimes talk, this is not a view that any serious ESRist has ever adopted (at least, not for very long: as we shall see, Maxwell held this position in 1965 but had abandoned it by 1968) but it is a view sometimes imputed to ESRists by their critics.

---

[4] Both the ESRist and the OSRist presuppose a number of other philosophically controversial theses, e.g. the thesis that the external world is not created by our minds and the thesis that theoretical terms are (at least putatively) referring expressions. However, arguments for and against the realist and structuralist theses have formed the basis for most of the discussion of structural realism in the literature, and will also form the basis for most of the discussion in this thesis.

[5] Sometimes the word "term" is used by logicians as a synonym for "name". Throughout this thesis it is used in a broader sense, as explicated here.

The Ramsey-sentence approach to ESR was first proposed by Maxwell (1968) and is adopted by most (but not all) modern ESRists. Let an observational term be a non-logical term that refers to an observable object, property or relation and a theoretical term be a non-logical term that refers to an unobservable object, property or relation. The Ramsey-sentence of a theory is obtained from a sentence expressing the theory by first replacing the theoretical terms (names and predicates) in the sentence with new variables (using the same variable for each occurrence of the same term, and different variables for different terms). The resulting formula is then turned into a sentence (the theory's Ramsey-sentence) by binding the variables with the appropriate existential quantifiers (placed at the start of the formula, so that every occurrence of the same new variable is in the scope of the same quantifier). Note that, in general (as long as the original sentence contains at least one theoretical predicate) constructing the Ramsey-sentence of a theory will require a language of second-order logic.[6]

Maxwell provides an example (Maxwell, 1970a, p.186). Consider the "theory" expressed by the sentence:

$\forall x([Ax \ \& \ Dx] \rightarrow \exists yCy)$

where "A" and "D" are "theoretical" predicates such that "Ax" means "x is a radium atom" and "Dx" means "x radioactively decays" and "C" is an "observational" predicate such that "Cx" means "x is a click in a suitably located Geiger counter".[7] Its Ramsey-sentence is:

$\exists X \exists Y \forall x([Xx \ \& \ Yx] \rightarrow \exists yCy)$

---

[6] Zahar (2001, p. 236) points out that we can construct an equivalent sentence using a first-order language that contains the predicates "being a set" and "$\in$". However, we would of course have to leave these unRamseyfied, and it is hard to see what justification there could be for that, unless one was prepared to take them as logical predicates, like "=".

[7] Of course this doesn't express any real "theory" and is moreover false: if a radium atom decays on Mars there will not be a click in a suitably located Geiger counter, because there will not be a suitably located Geiger counter, but it will suffice to give the general idea.

The ESRist (if he takes the Ramsey-sentence approach to ESR), elaborates his structuralist thesis as follows:

A theory's Ramsey-sentence is as much as a theory reliably tells us about the world.

Depending on the basis on which they make this assertion, ESRists may argue that either, (i) the Ramsey-sentence of a theory is the true logical form of the theory, so believing in the theory's Ramsey-sentence is believing the theory, or, (ii) the Ramsey-sentence of a theory is not the true logical form of the theory, so believing in the theory's Ramsey-sentence is not the same as believing the theory. These different approaches are considered in more detail in due course.

Notice that, when it is elaborated this way, the ESRist's structuralist thesis has the following corollary:

The knowledge provided by our mature scientific theories consists (in its ultimate form) of statements constructed using only logical and observational terms.

This corollary is a claim the instrumentalist would also endorse. ESR remains distinct from instrumentalism because the instrumentalist claims that our theories provide knowledge only about the observable world, whereas the ESRist maintains that the Ramsey-sentences of our theories provide knowledge about both the observable world and about the structure of the unobservable world.

At first sight, it may seem that as Ramsey-sentences are constructions formed using only logical and observational terms, they could not tell us anything about the unobservable world. This is certainly not true. As van Fraassen (1980, p. 54) notes the claim that "there are entities that are not O_1 and not O_2 and..." where each O_i is an observational predicate and the sentence ascribes to the entities it refers to the negation of every O_i in the language, successfully makes the claim that there are unobservables (which is a claim about the unobservable world), using only logical and

observational vocabulary. However, depending on how the observational term/theoretical term distinction is drawn it may be true that Ramsey-sentences can't tell us anything substantial about the unobservable world (this issue will be discussed in chapter 4).

Given that ESR has been explicated in terms of Ramsey-sentences, one may wonder if ESR is any more "structural" than conventional scientific realism: after all, the Ramseyfied version of a sentence is no more closely linked to the structures that satisfy it than is the original sentence. However, although there isn't any particularly intimate relation between structural realism and structures (in the set-theoretic sense) the term "structural" is not totally inappropriate: the ESRist maintains that theoretical predicates (and the sets that provide the extensional interpretation of these predicates) should not be given an intensional interpretation, but should be treated purely extensionally, i.e. treated in a purely mathematical or "structural" way.

On a related issue, proponents of the "semantic" view of theories may object to the characterisation of ESR given here, which takes for granted that theories are linguistic objects (an assumption that they reject). However, the syntactic view/semantic view debate is orthogonal to the issues discussed in this thesis. One can be a structural realist and advocate either the syntactic or the semantic view of theories. Although structural realism has been explicated in terms of the syntactic view in this thesis the substantive points that will be made about it are independent of this view. There is some further discussion of this issue in chapter 4.

3. WESR and SESR

ESR can be subdivided into two doctrines: weak ESR (WESR) and strong ESR (SESR) (as they will be called in this thesis). SESR presupposes a particular metaphysical and epistemological doctrine called "indirect realism". This is the view that although the external world exists, we do not have direct access to it. It implies that there is a distinction between the "internal world" of our own consciousnesses, to which we have direct

access, and the external world, to which we do not (this internal/external distinction is sometimes called the phenomenal/noumenal distinction or the mental/physical distinction).[8]

It is important to separate the internal/external distinction from the observable/unobservable distinction. Within the indirect realist framework the observable/unobservable distinction may be *roughly* characterised as a distinction between those external objects, properties and relations (e.g. tables, redness[9]) that have a direct counterpart in internal experience and those external objects, properties and relations (e.g. quarks, strangeness) that do not. Because the phrase "direct counterpart" is so vague this characterisation does not succeed in unambiguously drawing the intended distinction, but hopefully the general idea is clear enough. Some attempts that have been made to draw this distinction more clearly are discussed in chapter 2.

The essential difference between WESR and SESR is that while the WESRist thinks that theoretical terms (i.e. terms referring to unobservable objects, properties and relations) need to be Ramseyfied (leaving Ramsey-sentences containing only logical and observational terms) the SESRist thinks that external terms (i.e. terms referring to external objects, properties and relations) need to be Ramseyfied (leaving Ramsey-sentences containing only logical and internal terms). The version of ESR outlined in the previous section is thus WESR. The SESRist's structuralist thesis can still be stated as:

A theory's Ramsey-sentence is as much as a theory reliably tells us about the world.

---

[8] SESR will not be criticised for relying on indirect realism in this thesis. This is because indirect realism seems to me a plausible view. But it should be noted that for an advocate of direct realism (i.e. the view that we have direct access to the external world) SESR is a non-starter.

[9] In fact, the indirect realist might say that we need to distinguish two "redness" predicates: one that refers to the redness of internal objects (sense-data) and another that refers to the redness of external (physical) objects. This issued will be discussed in more detail in chapter 3.

But because of his different approach to Ramseyfication, the SESRist's structuralist thesis has a somewhat different corollary, viz:

The knowledge provided to us by our mature scientific theories consists (in its ultimate form) of statements constructed using only logical and internal terms.

As we shall see, the main argument for the SESRist's structuralist thesis, which was first advanced by Russell (1912, 1927) is actually only an argument for this corollary (but the thesis does seem highly plausible, given the corollary). In fact, as Russell was writing before the Ramsey-sentence approach to ESR was developed we may take it that, for Russell at least, this "corollary" is the structuralist thesis.

In fact, although the example of Ramseyfication that was given in the previous section is Maxwell's own and it suggests WESR he himself was a SESRist. He makes this quite clear when he states that, "My own view...is that all items should be considered theoretical [meaning that terms referring to them should be Ramseyfied] unless they occur in direct experience; since I reject any form of direct realism, this means that the observable [meaning the things referred to by terms that do not need to be Ramseyfied] is instantiated only in inner events of observers" (Maxwell, 1970a, p. 181).

On the face of it, however, Ramseyfying all terms except those that refer to items of internal experience will result in the Ramsey-sentences of most (if not all) theories being purely formal, and thus entirely devoid of empirical content, because, on the face of it, most (if not all) theories do not deal with items of internal experience at all. Take for example the toy theory given above. Clicks in suitably placed Geiger counters are not items of internal experience, so (taking the SESRist line) the predicates referring to them should also be Ramseyfied away, leading to the following Ramsey-sentence for the theory:

$\exists X \exists Y \exists Z \forall x([Xx \,\&\, Yx] \rightarrow \exists y Zy)$

which is purely formal, and therefore has no empirical content.

The proponent of SESR can avoid this unwelcome conclusion by arguing that although most theories, in themselves, are not directly about internal experience they are nonetheless connected to internal experience by auxiliary theories, which are always implicitly held. For example, we implicitly hold that we will have an experience of a sense-data Geiger counter click only if, (i) there is a (real) Geiger counter click, (ii) we are hallucinating Geiger counter clicks, (iii) we are dreaming about Geiger counter clicks, (iv) somebody is playing a practical joke,...Alongside the theory:

$\forall x([Ax \ \& \ Dx] \rightarrow \exists yCy)$

we would thus also hold:

$\exists x(C'x \ \& \ Eax) \rightarrow (\exists y[Cy] \text{ v } Ha...)$

where a is a constant referring to oneself, "C'x" means "x is a sense-data Geiger counter click", "Exy" means "x experiences y", "Cx" means the same as before and "Hx" means "x is hallucinating Geiger counter clicks". The combined theory we hold is thus expressed by the sentence:

$\forall x([Ax \ \& \ Dx] \rightarrow \exists yCy) \ \& \ (\exists x[C'x \ \& \ Eax] \rightarrow [\exists y(Cy) \text{ v } Ha...])$

Taking the SESRist approach to Ramseyfication one obtains something like the following Ramsey-sentence of the combined theory:

$\exists W\exists X\exists Y\exists Z...\forall x([Wx \ \& \ Xx] \rightarrow \exists yYy)\&(\exists x[C'x \ \& \ Eax] \rightarrow [\exists yYy \text{ v } Za...])$

which does contain some non-logical terms ("C'", "E" and "a") and thus does not fail to make an empirical claim for lack of them.

Chapter 1

Weak Epistemic Structural Realism

As explained in the introduction to this thesis, the difference between WESR and SESR is that while the WESRist thinks that theoretical terms (i.e. terms referring to unobservable objects, properties and relations) need to be Ramseyfied (leaving Ramsey-sentences containing only logical and observational terms) the SESRist thinks that external terms (i.e. terms referring to external objects, properties and relations) need to be Ramseyfied (leaving Ramsey-sentences containing only logical and internal terms).

Given a reasonably charitable reading, Worrall (1989, 1994) appears to be espousing a form of WESR. However, comments like the following (where Worrall discusses the shift from Fresnel's theory of light to Maxwell's):

> There was an important element of continuity in the shift from Fresnel to Maxwell – and this was much more than a simple question of carrying over the successful *empirical* content...the continuity is one of *form* or *structure* not of content. (Worrall, 1989, p. 157, original emphasis)

have led Psillos (1999, chapter 7) to attribute to him a view very different from WESR, the view that the only element of continuity across scientific revolutions (and thus the only aspect of theories about which we should be realists) are *uninterpreted* equations, which are somehow supposed to carry with them empirical content. Such a view would be absurd because (as Psillos goes on to point out) carrying over empirical content requires, of course, much *more* than just carrying over uninterpreted equations (clearly, an uninterpreted equation has no empirical content: the different theories describing the behaviour of pendulums, of masses on springs and of voltages in circuits with a condenser and a solenoid all use the equation $d^2x/dt^2 = -[k/m]x$, whilst clearly having very different empirical contents).

The view ascribed to Worrall here is that the equations are preserved along with the interpretations of some of the terms in the equations. Specifically, the terms that are interpreted as denoting observable objects, properties and relations keep the same interpretations. This view is not only more reasonable but also appears to be better supported by Worrall's case study.[1] For example, consider one of the equations (relating the intensities of incident, reflected and refracted beams of light when light passes from one medium to another) that is preserved in the Fresnel/Maxwell episode:

$R/I = \tan(i-r)/\tan(i+r)$

While the interpretations of I and R vary between Fresnel and Maxwell (being taken as amplitudes of vibration of a [postulated] mechanical ether and a [postulated] non-mechanical electromagnetic field respectively) the interpretations of i and r (as the angles made by the incident and refracted light beams with the normal to the boundary between the media) are the same in both theories. To retain empirical adequacy Maxwell had to retain not only the structure of Fresnel's equations but also the interpretation of the observational terms.

It is also clearly not enough just to retain the equations that link I and R to i and r. Maxwell also needed to retain the equations that link I and R to observed light intensities, i.e. $I^2 = I_I$ and $R^2 = I_R$ (where I_I and I_R are the intensities of light, polarised in the plane of incidence, of the incident and reflected beams respectively) and the interpretations of the "observational" terms (I_I and I_R) in these equations. Maxwell was only free to reinterpret the theoretical terms (I and R). So the case suggests that there is continuity (not just of the equations but also) of the interpretations of some of the terms in the equations (the observationally interpreted terms) across scientific revolutions.

[1] In fact, Worrall has confirmed (in a personal communication) that this is the view he intended.

Worrall's argument for WESR is based on the claim that it can reconcile the two most powerful arguments in the scientific realism debate: the "no miracles argument" (in favour of scientific realism) and the "pessimistic induction" (also known as the pessimistic metainduction) (in favour of scientific antirealism).[2] In section 1 of this chapter the no miracles argument is discussed: it is argued that whilst WESR could accommodate (a form of) this argument, the argument is flawed, so this is no particular virtue of WESR. In section 2 the pessimistic induction is discussed: it is argued that while (a form of) this argument does have some force the WESRist's response to the argument is not particular convincing (no more convincing, in fact, than the conventional realist's response). It is therefore concluded that Worrall's argument for WESR is not compelling.

1. The No Miracles Argument

The no miracles argument has been called the "ultimate" argument for realism (Putnam, 1979, p. 73). The argument runs roughly as follows: some scientific theories enjoy enormous empirical success; if these theories are not even approximately true, this appears to be miraculous; on the other hand the approximate truth of such theories would explain their success; as we don't want to accept miracles, we should thus accept that such theories are approximately true (at least, in the absence of another explanation of their success).[3]

The structural realist's realist thesis was stated in the introduction to this thesis as follows:

---

[2] Although Worrall never puts it this way, we can think of him as using the no miracles argument to support the structural realist's realist thesis and the pessimistic induction to support the WESRist's structuralist thesis.

[3] Note that the argument relies on the notion of approximate truth. It would obviously be too strong if it were phrased in terms of absolute truth. After all, Newton's theory of gravitation was astonishingly successful and so is general relativity, but since the theories are contradictory they can't both be absolutely true (whereas it seems possible that they might both be approximately true).

Mature scientific theories provide us with a substantial amount of knowledge about both the observable and the unobservable world.

If (following Worrall) we define mature scientific theories as those that are successful then this thesis is supported by the no miracles argument, assuming that the totality of the theories that (according to the no miracles argument) we can take to be approximately true can be said to provide us with a substantial amount of knowledge about both the observable and the unobservable world.

However, it may be wondered whether or not this rendering of the no miracles argument is compatible with the ESRist's structuralist thesis. The ESRist's structuralist thesis is that a theory's *Ramsey-sentence* is as much as a theory reliably tells us about the world, i.e. that a theory's *Ramsey-sentence* is all that we can take to be approximately true, whereas the upshot of the no miracles argument is that the theory is itself approximately true. As noted in the introduction to this thesis, the ESRist may think that the Ramsey-sentence of a theory is the true logical form of the theory, or that it is something quite distinct from the theory. In the former case he can accept the no miracles argument in this form (since the Ramsey-sentence of the theory is taken to be the true logical form of the theory to say that a theory's Ramsey-sentence is approximately true is to say that the theory is approximately true). However, if the ESRist holds that the Ramsey-sentence of the theory is something quite distinct from the theory then the no miracles argument in this form is not compatible with his structuralist thesis.

In either case, however, the ESRist could endorse the following variant of the no miracles argument: some scientific theories (e.g. general relativity) appear to enjoy enormous empirical success; if *the Ramsey-sentences of* these theories are not even approximately true this appears to be miraculous; on the other hand if *the Ramsey-sentences of* these theories were approximately true then this would explain their success; as we don't want to accept miracles, we should thus accept that *the Ramsey-sentences of* such theories are approximately true (at least, in the absence of another explanation of their success). This version of the no miracles argument is

definitely compatible with the ESRist's structuralist thesis, whilst it still seems to support his realist thesis: the thesis that mature scientific theories provide us with a substantial amount of knowledge about both the observable and the unobservable world. Although Worrall does not explicitly put forward this version of the no miracles argument, it seems to be the only version of the argument that unambiguously supports the WESRist position, and, since he does take the no miracles argument to support this position, it seems reasonable to attribute this version of the argument to him.[4]

1.1. First Interpretation of the No Miracles Argument

Magnus and Callender (2004), following Howson (2000), suggest the following reconstruction of the no miracles argument:

Premise 1: x is probably a successful theory.
Premise 2: If x is an approximately true theory then it would probably be a successful theory.
Premise 3: If x is not an approximately true theory then it would be a miracle if x is a successful theory.

Conclusion: If x is a successful theory then x is probably an approximately true theory.

Let Ax mean that x is an approximately true theory. Let Sx mean that x is a successful theory. Let P(X) be the probability of X and P(X|Y) be the probability of X conditional on Y. Then the argument runs as follows:

Premise 1: $P(Sx) \approx 1$

[4] This version of the no miracles argument has been put forward explicitly by Cruse and Papineau: "The Ramsey sentence realist says that we should believe in the approximate truth of a successful theory's Ramsey sentence, on the grounds that it would be a miracle that the theory were successful, were its Ramsey sentence not true" (Cruse and Papineau, 2002, p. 179).

Premise 2: $P(Sx|Ax) \approx 1$

Premise 3: $P(Sx|\neg Ax) \approx 0$

Conclusion: $P(Ax|Sx) \approx 1$

This argument is valid. First note that $P(Ax) \approx 1$:

$P(Sx) = P(Sx|Ax).P(Ax) + P(Sx|\neg Ax).P(\neg Ax)$

So, $P(Sx) = P(Sx|Ax).P(Ax) + P(Sx|\neg Ax).(1-P[Ax])$

So, $P(Ax) = (P[Sx] - P[Sx|\neg Ax]) / (P[Sx|Ax] - P[Sx|\neg Ax])$

So, $P(Ax) \approx 1$ (using premises 1-3)

Bayes' theorem states that:

$P(Ax|Sx) = (P[Ax].P[Sx|Ax]) / P(Sx)$

So, $P(Ax|Sx) \approx 1$ (using the result just obtained and premises 1 and 2)

So the argument is valid. However, as Magnus and Callender (2004) point out, it is not sound. Their objection to the argument is (in essence) that premises 1 and 3 will only both seem plausible if one equivocates over the reference class. Suppose that the reference class is the set of all possible theories. Then premise 1 is surely false, because it is surely not the case that a theory selected at random from the set of all possible theories will probably be successful. (Without premise 1 the argument is not valid. A counterexample is illustrated in figure 1a.[5]) So suppose instead the reference

[5] Indeed, even if we were to retain a weakened version of premise 1, to the effect that it would not be a miracle for a given theory to be successful, then the argument would still not be valid. The argument would then run as follows:

Premise 1: $P(Sx) > m$

Premise 2: $P(Sx|Ax) \approx 1$

Premise 3: $P(Sx|\neg Ax) < m$

Conclusion: $P(Ax|Sx) \approx 1$

class is the set of mature theories. If, following Worrall (1989), we define a theory to be mature if and only if it is successful, then premise 1 is certainly true, in fact P(Sx) = 1. But now premise 3 is false, because P(Sx|¬Ax) = 1, because all theories in the reference class are successful, irrespective of whether or not they are approximately true. (Without premise 3 the argument is not valid. A counterexample is illustrated in figure 1b.)

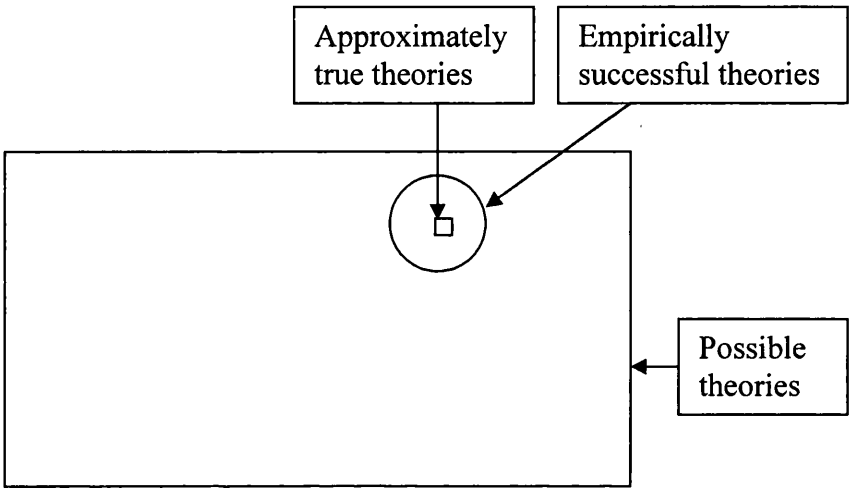

Figure 1a: This illustrates a situation in which, (i) If x is an approximately true theory then it is probably a successful theory: P(Sx|Ax) ≈ 1 (in fact, P[Sx|Ax] = 1), (ii) If x is not an approximately true theory then it would be a miracle if x is a successful theory: P(Sx|¬Ax) ≈ 0, but, (iii) It is not the case that if x is a successful theory then x is probably an approximately true theory: P(Ax|Sx) is not ≈ 1.

(where m is some small probability such that the occurrence of an event is miraculous if and only if the probability of the event is less than m). This is not valid. If we take m to be 0.0010005 then a counterexample is provided when one in a million theories is approximately true, 1001 in a million theories are successful and all approximately true theories are successful. In that case, (i) P(Sx) = 0.0010010 (so P[Sx] > m), (ii) P(Sx|Ax) = 1 (so P[Sx|Ax] ≈ 1), (iii) P(Sx|¬Ax) = 0.0010000 (so P[Sx|¬Ax] < m), but, (iv) P(Ax|Sx) = 0.0009990 (so P[Ax|Sx] is not ≈ 1). Indeed, in this counterexample not only is a successful theory probably not approximately true, it would be a miracle if a successful theory turned out to be approximately true (P[Ax|Sx] < m)!

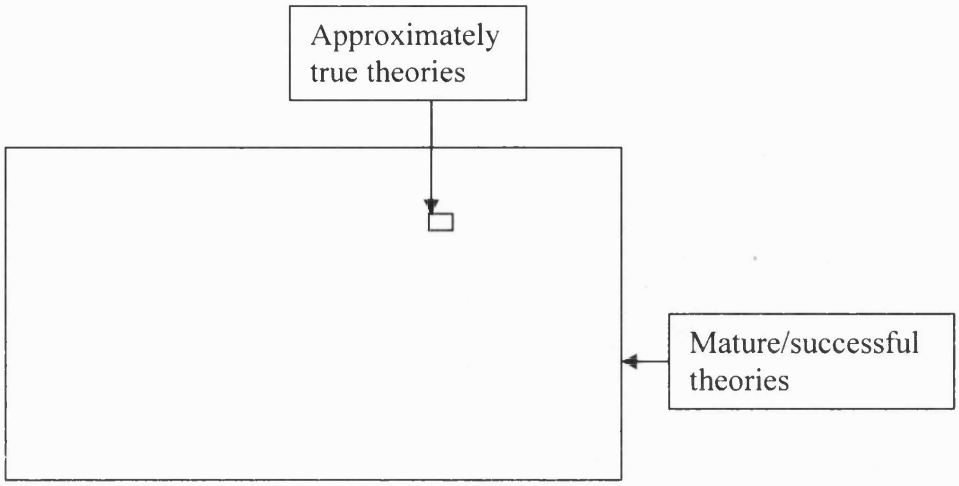

Figure 1b: This illustrates a situation in which, (i) x is probably a successful theory: $P(Sx) \approx 1$ (in fact $P[Sx] = 1$), (ii) If x is an approximately true theory then it is probably a successful theory: $P(Sx|Ax) \approx 1$ (in fact, $P[Sx|Ax] = 1$), but, (iii) It is not the case that if x is a successful theory then x is probably an approximately true theory: $P(Ax|Sx)$ is not ≈ 1.

1.2. Second Interpretation of the No Miracles Argument

Another attempt to reconstruct the no miracles argument probabilistically might run as follows:

Premise 1: If x is an approximately true theory then it would not be a miracle if x is a successful theory.
Premise 2: If x is not an approximately true theory then it would be a miracle if x is a successful theory.

Conclusion: If x is a successful theory then x is relatively likely to be an approximately true theory (i.e. a theory from the set of successful theories is more likely to be approximately true than a theory from the set of all theories).

That is:

Premise 1: $P(Sx|Ax) >> 0$

Premise 2: $P(Sx|\neg Ax) \approx 0$

Conclusion: $P(Ax|Sx) > P(Ax)$

This argument is valid and also (plausibly) sound. Bayes' theorem states:

$P(Ax|Sx) = P(Ax).P(Sx|Ax) / P(Sx)$

So, $P(Ax|Sx) > P(Ax)$ if and only if $P(Sx|Ax) / P(Sx) > 1$

So, $P(Ax|Sx) > P(Ax)$ if and only if $P(Sx|Ax) > P(Sx)$

So, $P(Ax|Sx) > P(Ax)$ if and only if $P(Sx|Ax) > P(Sx|Ax).P(Ax) + P(Sx|\neg Ax).P(\neg Ax)$

So, $P(Ax|Sx) > P(Ax)$ if and only if $P(Sx|Ax) > P(Sx|Ax).P(Ax) + P(Sx|\neg Ax).(1 - P[Ax])$

The second inequality follows from $P(Sx|Ax) > P(Sx|\neg Ax)$ (provided $P[Ax] \neq 1$), which follows from premises 1 and 2 of the argument. So it follows from the premises of the argument that $P(Ax|Sx) > P(Ax)$ (i.e. the argument is valid).

The problem with this attempt to reconstruct the no miracles argument is that the conclusion of the argument is not the conclusion the realist wants. What the scientific realist wants is that it is *likely* that a successful theory is approximately true, i.e. that $P(Ax|Sx) > 0.5$ (i.e. what he wants is not the result that a theory from the set of successful theories is more likely to be approximately true than a theory from the set of all theories but the result that a theory from the set of successful theories is actually likely to be approximately true). But, even if the premises of the argument are true, $P(Ax|Sx)$ can be made arbitrarily small by making $P(Ax)$ small enough. This is apparent from figure 1a. Hence, although this argument may show that success makes approximate truth *relatively* likely it requires an additional argument to show that success makes approximate truth likely, which, so far, the realist has failed to supply.

1.3. Third Interpretation of the No Miracles Argument

The third interpretation of the no miracles argument, which is the most commonly proposed, is that the argument is an instance of the argument form known as "inference to the best explanation" (IBE). IBE is (allegedly) a form of inference that is often used in science. If we have a theory that explains some phenomenon better than any other theory (that we know of) then scientists (allegedly) infer from this that the theory is (approximately) true. So instances of IBE take the following form (in the following p is a description of an established phenomenon and T is a statement of a theory):

Premise 1: p
Premise 2: T is the best explanation of p

Conclusion (via IBE): T is approximately true

We might call this a "ground level" IBE. When the no miracles argument is construed as an instance of IBE, it is a kind of "meta level" IBE: the phenomenon in question is taken to be the success of some scientific theory, T, and the theory that (allegedly) offers the best explanation of this fact is the theory that T is approximately true. I.e. the no miracles argument runs as follows:

Premise 1: T has been successful
Premise 2: The theory that T is approximately true is the best explanation of the fact that T has been successful

Conclusion (via IBE): The theory that T is approximately true is approximately true

Notice that, strictly speaking, the conclusion of the no miracles argument (if it follows the pattern of other instances of IBE) should not be simply that T is approximately true but that it is approximately true that T is approximately true. It is not clear what (if anything) this means. Realists take the conclusion of the no miracles argument to be that T is

approximately true, so they must think that the approximate truth of the approximate truth of T is the same as the approximate truth of T. But this is not obvious: it seems possible that it could be approximately true that T is approximately true even if T is radically false. However, let's leave this point aside, since there are less linguistically frustrating difficulties with this interpretation of the no miracles argument, which are discussed in subsections 1.4-1.6.

1.4. Laudan's Objections to the No Miracles Argument

Laudan's (1981) main objection to the no miracles argument is based on the pessimistic induction. Roughly speaking, he argues that there have been many successful theories in the history of science that (we now believe) are not even approximately true and that we should infer from this that it is *not* a miracle for a radically false theory to be successful. Hence the no miracles argument is undermined. The pessimistic induction will be examined in more detail in section 2.

Another objection Laudan brings against the no miracles argument is that approximate truth could not in any case be invoked to explain success because there is no real reason to think that the fact that a theory is approximately true implies that the theory will be successful.[6] Laudan objects that proponents of the no miracles argument have not provided an account of approximate truth, so the claim that approximate truth implies success is open to question. He insists that the realist needs, "more than a promissory note that somehow, someday, someone will show that approximately true theories must be successful theories" (Laudan, 1981, p. 1125).

Laudan is certainly right that the connection between the approximate truth of a theory and its success has not been rigorously demonstrated. Indeed, Worrall freely admits (1989, p. 106) that his use of

[6] Laudan accepts, of course, that a *true* theory would be successful. He balks only at the realist's assumption that an *approximately true* theory would also, probably, be successful (Laudan, 1981, p. 1124).

the no miracles argument is based on the *intuition* that there is a connection between the two. However, although this might throw some doubt on the legitimacy of the no miracles argument, it does not completely undermine it, because it does seem intuitively reasonable to suppose there is a connection between the approximate truth of a theory and its success.

1.5. A Related Objection to the No Miracles Argument

Notwithstanding what has been said in subsection 1.4, it seems that the (second) premise of the IBE interpretation of the no miracles argument, which states that, "the theory that T is approximately true is the best explanation of the fact that T has been successful" is false. It seems that even if we put Laudan's worries aside and accept that approximate truth would *imply* success, approximate truth (or even complete truth) would not in any case *explain* success.

Suppose we ask of a given theory, "Why is it so successful?" and we are told that it is because, "It is approximately true" (or something like this). We should not be satisfied with this answer. We think that the success of the theory requires an explanation (presumably) because success is not a characteristic that we would generally expect a theory to display, i.e. because we believe that of all the possible theories that we might have devised, not many would have been successful. The answer given is explanatory only if success is a characteristic we would generally expect an approximately true theory to display, i.e. we believe that all (or at least most) of the approximately true theories that we might have devised would have been successful. But this implies that, of all the possible theories that we might have devised, not many are approximately true. So approximate truth is not a characteristic we would generally expect a theory to display. As it is not as if we had to choose from a list of possible theories, knowing in advance which ones were approximately true, the reply to our question seems only to raise another question, "Why is the theory that we devised approximately true?". Approximate truth can't appropriately *explain* success, because if success needs an explanation then so does approximate

truth: approximate truth is no more a characteristic that we would expect a theory to display than is success.

The problem is highlighted by making a no miracles style argument in a more commonplace setting. Suppose I am watching the national lottery, and the first five balls that are drawn match five of the numbers on my ticket. It would be a miracle if this just happened by chance. On the other hand it would be no miracle (in fact, it would be bound to happen) if I had a winning ticket (i.e. a ticket on which all six numbers match). As we don't want to accept miracles we should thus accept that I have a winning ticket (at least, in the absence of another explanation of why the first five numbers match). In this context it is (I hope) obvious that the proposed explanation is inadequate, because the hypothesis that I have a winning ticket is obviously in need of an explanation at least as much as the fact it is invoked to explain. Likewise, the approximate truth of a theory is at least as much in need of an explanation as its success. (Peter Lipton suggested this comparison at a conference.)

The situation is illustrated in more general terms in figure 2. Suppose we have some outcome, a, with a surprising property, P (surprising in the sense that most outcomes do not have the property). If we ask "Why does a have property P?" and we are told that it is because, "It has property Q" we are being fobbed off, unless we can explain why a has property Q in a non-circular manner (e.g. by pointing out that we selected a from the class of outcomes with this property).

The proponent of the no miracles argument might claim that this shows nothing, since all explanations must stop somewhere and that he is happy to leave approximate truth unexplained. But then it seems that the no miracles argument is question begging, since the antirealist is just someone that is happy to leave success unexplained.

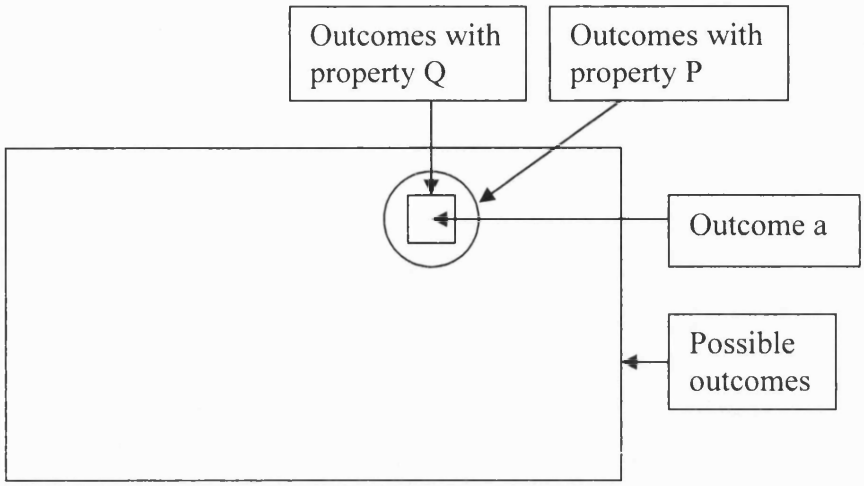

Figure 2: The large rectangle represents the space of possible outcomes (e.g. theories). The circle represents outcomes with property P (e.g. the property of being empirically successful). The square represents outcomes with property Q (e.g. being approximately true).

1.6. The Redundancy of the No Miracles Argument

Instances of IBE are not, of course, deductively valid: the (seemingly) best explanation of some fact may not turn out to be the true explanation of that fact (if indeed there is an explanation of it at all). In fact, antirealists often claim that inferences of this form have no legitimacy at all (in particular, the antirealist denies that the inference to unobservables is legitimate, despite the fact they might offer the best explanation of an observable phenomenon). Consequently, as Fine (1984, pp. 1187-1189) has pointed out, if the no miracles argument is construed as an instance of IBE it appears to be monumentally question begging: it will never convince the antirealist since he doesn't accept that the form of inference it exemplifies is in general legitimate. It is simply preaching to the converted or, at best, the agnostic.[7]

[7] Zahar claims that "Fine's argument is predicated on a contentious assumption, namely that *scientific* realism deserves its name: SR [i.e. scientific realism] is taken to be an empirical hypothesis whose factual basis consists of the successes of various scientific theories." (Zahar, 2001, p. 58, original emphasis). Zahar goes on to argue that scientific realism is not an empirical hypothesis but is rather a metaphysical hypothesis and then concludes that because of this "the whole basis for Fine's argument collapses" (Zahar, 2001, p.59).

Moreover, if someone does accept that IBE is legitimate, the no miracles argument is completely redundant, since he will accept the simpler ground level IBE that can be less problematically used to reach the desired conclusion that our best scientific theories are approximately true.

So the debate really comes down to this: should we accept that IBE is a legitimate form of inference? If yes, then we can use a ground level IBE to infer that our best theories are approximately true. (And if, contrary to what has been suggested, the premises of the no miracles argument are true then we could also use the no miracles argument to reach this conclusion, but it would be unnecessary to do so.) If no, then we can use neither the ground level IBE, nor the meta level IBE of the no miracles argument, and the realist must look elsewhere for an argument to support his position. Let us turn then to consider the arguments for and against the legitimacy of IBE.

1.7. The Underdetermination Objection to IBE

Why does the antirealist (claim to) reject IBE? One reason is suggested by the argument known as the underdetermination of theory by evidence. We invariably base our theories, and test them against, a finite amount of evidence. But any finite amount of evidence could be accommodated by infinitely many theories. The antirealist argues that although some of these theories may be better than others for pragmatic or aesthetic reasons, there is no reason to think that any one of them is more likely to be true than any other. Hence an inference to the best explanation is not very likely to be an inference to a true explanation.

One might take issue with the claim that any finite amount of evidence could be accommodated by infinitely many theories. Or claim that, even if this is so, all the infinitely many theories that accommodate a *substantial* amount evidence are likely to be approximately true, so we can,

---

However, Fine claims only that the antirealist doesn't generally accept the form of inference that the no miracles argument exemplifies and this claim in no way rests on the assumption that scientific realism is an empirical hypothesis, so Zahar seems to be way off beam here.

with reasonable safety, infer the approximate truth of any of them. However, these are not good responses to the underdetermination argument, as is evident if we consider the argument in its "curve-fitting" form. In this form the argument goes as follows. Given *any* finite number of data points (no matter how large) we can draw infinitely many curves that pass through all of them. Hence any finite amount of evidence can be accommodated by infinitely many theories. Moreover, the evidence does not even force on us a curve that approximates the true curve (i.e. the curve that represents the actual relation between the variables represented on the axes) to any reasonable degree: although infinitely many curves will be rather similar to the true curve (in some sense that could be made precise, e.g. – assuming the graph has two axes – in the sense that the area between the curve and the true curve is small or in the sense that for every value of the x co-ordinate both curves have a similar value of the y co-ordinate) infinitely many will be radically different to the true curve (in the same sense) (if we choose a curve somewhere in the "middle" the maximum possible inaccuracy is minimised, in an intuitive sense, but it still has no finite bound, according to either of the measures suggested above). Hence any finite amount of evidence can be accommodated by infinitely many theories that are not even approximately true.

Of course, realists have replied to this argument, often arguing that we have more than just empirical evidence to help us determine the true curve. E.g. it has been argued that we (somehow) know that nature is simple, relative to our way of viewing it, so the simplest curve is most likely to be the true one. But, as noted, the antirealist will say that although there maybe pragmatic and aesthetic reasons to prefer the simplest curve there are no good epistemic reasons: the assumption that nature is simple, relative to our way of viewing it might be pragmatically and aesthetically justified, but it is not epistemically justified.

Applied to the no miracles argument the underdetermination objection runs as follows. Even if the approximate truth of a theory is the best explanation of its success that does not mean that it is particularly likely that it is the (or an approximately) true explanation. After all, there are infinitely many (typically less simple) theories that would have exactly the

same empirical success that any given successful theory enjoys and although infinitely many of these will be approximately true, infinitely many will be radically false. So, unless we are prepared to swallow the view that we know that nature is simple, relative to our way of viewing it, it would not be miraculous if our theory were one of the radically false ones.[8]

1.8. Psillos' Defence of IBE

Psillos (1994, chapter 4) appears to deploy two quite separate arguments in defence of IBE. Firstly, he argues that IBE can be justified. His argument for this runs as follows:

> the best explanation of the instrumental reliability of scientific methodology is that background theories are relevantly approximately true. These background scientific theories have themselves been typically arrived at by abductive reasoning. Hence it is reasonable to believe that abductive reasoning is reliable: it tends to generate approximately true theories. (Psillos, 1999, p. 80)

As Psillos is aware, this argument is circular: IBE is used to infer the conclusion that IBE is reliable. In fact there are two uses of IBE in the argument. The first instance is a generalised form of the no miracles argument (generalised in so far as it discusses scientific theories in general, rather than one particular theory):

Premise 1: Scientific theories are successful.

[8] It has been suggested to me that the realist can reply here that the radically false successful theories are unlikely to remain successful in the future. It is true that infinitely many of these radically false theories will cease to be successful in the future (as long as there is at least some evidence collected in the future). However, as there will only be a finite amount of evidence collected up to any given time in the future there will always be infinitely many radically false theories that remain successful up to that time.

Premise 2: The theory that scientific theories are approximately true is the best explanation of the fact that scientific theories are successful.

Conclusion (via IBE): The theory that scientific theories are approximately true is approximately true.

The (abbreviated) conclusion of the generalised no miracles argument is then used as a premise in a second instance of IBE:

Premise 1: Scientific theories are approximately true.
Premise 2: The theory that scientific theories are arrived at via a reliable rule of inference is the best explanation of the fact that scientific theories are approximately true.

Conclusion (via IBE): The theory that scientific theories are arrived at via a reliable rule of inference is approximately true.

As scientific theories are (allegedly) arrived at via IBE, this implies that it is (approximately) true that IBE is a reliable rule of inference: which is (approximately) the conclusion that Psillos desires.

Psillos argues that although circular this justification of IBE is not viciously circular. He distinguishes between "premise-circular" arguments, where the conclusion of the argument is among the premises of the argument and "rule-circular" arguments, where the conclusion of the argument is that the form of inference deployed in the argument is legitimate. Perhaps an example will help to clarify the difference. The following is a premise-circular justification of modus ponens:

Premise: Modus ponens is legitimate.

Conclusion: Modus ponens is legitimate.

This is premise-circular because the conclusion is the same as the premise. The following is a rule-circular justification of modus ponens:

Premise 1: If the absolute is perfect then modus ponens is legitimate.
Premise 2: The absolute is perfect

Conclusion (via modus ponens): Modus ponens is legitimate.

This is rule-circular because the conclusion of the argument is that modus ponens is legitimate and modus ponens is used to obtain the conclusion from the premises.

The foregoing justification of IBE is rule-circular, since IBE is used to reach the conclusion (that IBE is reliable) from the premises but it is not premise-circular (since no premise in the argument states that IBE is reliable) and Psillos claims that rule-circular arguments (unlike premise-circular arguments) are not viciously circular.

However, it is hard to see why rule-circular arguments should be thought of as any less vicious than premise-circular arguments, particularly as rules of inferences can often be converted into premises, and vice versa. For example, in some systems of first-order logic there is a rule of "universal generalisation": if, in a deduction, one has obtained the formula α then (with some restrcitions) one may also infer $\forall x\alpha$ from the premises of the deduction (cf. Suppes, 1957, p. 99). On the other hand, in other systems there is no such rule, rather all the formulas of the form $\alpha \rightarrow \forall x\alpha$ (with some restrictions) are taken as premises that may be used at any time in any deduction, and one has the rule of modus ponens, allowing one to achieve the same results (cf. Enderton, 2001, p. 112).

In any case, whether or not we call the above argument viciously circular, it is clear that it will not convince anyone who does not already accept IBE that IBE is reliable (because it uses IBE) and as such it does not appear to answer the antirealist. Psillos goes on to claim that IBE is a basic form of inference that humans deploy and as such we should not expect to be able to provide a justification of IBE that is independent of IBE (he adds that a justification that depends on IBE – such as the one that he provides – is not insignificant, in so far as there is no a priori guarantee that such a

justification exists). He draws a parallel here to justifications of deductive rules of inference. In fact he claims that, "The situation is totally analogous to the defence of deductive reasoning. There is no way in which one can persuade a *deductively blind* person of the soundness or rationality of deductive arguments" (Psillos, 1999, p. 88, original emphasis).

There are two points to note here. Firstly, Psillos appears to be conceding that his argument will not convince the antirealist: as Lipton (2001, p. 349) notes, to deem a form of inference basic is tantamount to conceding that a sceptic with regard to that form of inference cannot be answered. Secondly, as Lipton also notes, Psillos' analogy between justifying IBE and justifying deductive inferences is not exactly apposite, because one is very hard pressed to find a deductive sceptic, whereas antirealists who profess to be sceptics with regard to IBE are relatively common. As Lipton puts it, "it does not appear that our inductive practices, as the realist construes them, are in fact basic for all of us" (Lipton, 2001, p. 349). In a sense then, it is more problematic to deem IBE basic than it is to deem various forms of deductive inference basic, simply because there are apparently more IBE sceptics than deductive sceptics. (Of course that does not mean that IBE isn't basic: it just means that if we deem it basic there are lots of people whom, we have to concede, we cannot convert.)

Psillos' second defence of IBE is to suggest that it needs no justification. He claims that the correct attitude towards rules of inference is not "guilty until proven innocent" but rather "innocent until proven guilty" (cf. Psillos, 2001b, p. 368) so that even if IBE cannot be justified and is no more basic than any other rule of inference it is nonetheless legitimate to use it:

> If one knew that a rule of inference was unreliable, one would be foolish to use it. This does not imply that one should first be able to prove that the rule is reliable before one uses it. All that is required is that one should have no reason to doubt the reliability of the rule (Psillos, 1999, p. 85)

This seems to suggest that no rule of inference is more basic than another; that all rules of inference are born equal. However, a blanket innocent until proven guilty policy is unworkable. Suppose we start with the assumption that both induction and counter-induction (cf. Salmon, 1966, pp. 12-17) are innocent. There are then two ways in which we can proceed: either we can prove counter-induction guilty using an inductive argument:

Premise: Counter-induction has been unreliable in the past

Conclusion (via induction): Counter-induction will be unreliable in the future

or we can prove induction guilty using a counter-inductive argument:

Premise: Induction has been reliable in the past

Conclusion (via counter-induction): Induction will be unreliable in the future

If we proceed by the latter route we will be left believing that counter-induction, but not induction, is a reliable form of inference (and, in the absence of induction, it seems there will be no possible way in which counter-induction could ever be found guilty). But surely we don't want to be left believing this. Yet, without pre-judging the case, there seems to be nothing to choose between these two routes. So it seems that we must pre-judge the case, i.e. we must hold a selective policy, whereby some rules are guilty until proven innocent and others innocent until proven guilty (so, after all, we must take some rules of inference as more basic than others).[9] In that case, however, it remains open as to whether or not IBE is one of the former or the latter, so Psillos can't well argue from the premise that there is no

[9] We can't hold *all* rules to be guilty until proven innocent because if one did so one would never be able to prove anything innocent (i.e. one would never formulate any rules of

reason to doubt IBE to the conclusion that IBE is legitimate, because the antirealist may believe that IBE is guilty until proven innocent. It seems then that Psillos' second argument in favour of IBE will also not convince the antirealist that IBE is legitimate.

1.9. The "Double Standards" Defence of IBE

Inferring any universal law – even a simple empirical generalisation – from a finite number of observations, requires an application of IBE. For example, we infer from a finite number of observations of green emeralds that all emeralds are green. But that is not the only hypothesis that would explain the observations. Another such hypothesis is that all emeralds are grue, or that all emeralds are gred, etc.[10] We infer that all emeralds are green because that seems to us the best of these hypotheses (for some reason that is, admittedly, obscure). Or, to take a less contrived example (discussed by Poincaré [1903] and Worrall [forthcoming]), we infer from a finite number of observations of the positions of the planets that that the planets orbit the Sun in (approximate) ellipses. But the finite observations of the planets could also be explained by a number of hypotheses suggesting that the planets orbit the Sun in much more devious ways, and just happen to lie on (approximate) ellipses whenever we make an observation.

Even antirealists (at least in the scientific realism debate) are not usually sceptics with regard to inferences to simple empirical generalizations like these. Hence, despite what they might claim, antirealists are not *total* sceptics with regard to IBE. But (so the "double standards" defence of IBE goes) unless they can say why some instances of IBE (inferences to simple empirical generalizations) are legitimate whereas others are not there seems to be no justification for being a *selective* IBE sceptic, and antirealists seem to have an unacceptable double standard. In

---

inference) because a proof of innocence must require at least one rule of inference to go through.

[10] Where an object is grue if and only if it is green upto a particular future time, t, and blue after t and an object is gred if and only if it is green upto t and red after t (cf. Goodman, 2006, pp. 74-75).

the absence of a justification for this selective scepticism the antirealist ought either to become a thoroughgoing IBE sceptic, and therefore an inductive sceptic, or accept IBE in other cases. If he takes the former route his position is unassailable. But no one really would take that route. So he must take the latter route. But if he accepts IBE is legitimate in general, then he will accept at least the ground level IBE discussed above (if not the no miracles argument itself), which already suggests that our best scientific theories are approximately true.

Van Fraassen (1989) has a response to this "double standards" argument. To understand his response we must distinguish two conceptions of rationality. According to one, "what is rational to believe is exactly what one is rationally compelled to believe" (van Fraassen, 1989, p. 170) whereas according to the other (which van Fraassen favours), "what is rational to believe includes anything that one is not rationally compelled to disbelieve" (van Fraassen, 1989, pp. 170-171).

For van Fraassen IBE is better seen as a *tool* of inference than as a *rule* of inference. He accepts that it is rational to use IBE to form beliefs (as long as these beliefs are logically consistent) but he does not accept that it is irrational to refrain from using IBE in some or (presumably) all cases. So he thinks it is perfectly rational for the antirealist to use IBE to arrive at beliefs about simple empirical regularities, but to refrain from using IBE to arrive at beliefs involving unobservables. He presumably also thinks that it is perfectly rational for the realist to use it in both cases (as long as the beliefs he forms are logically consistent) and for the inductive sceptic to use it in neither. For van Fraassen rationality *permits* us to use IBE (within certain bounds) but it never *obliges* us to use it.

The antirealist's so-called double standard only seems irrational if one thinks of IBE as a rule of inference, whose dictates must be obeyed (on pain of irrationality). If it is rather a tool of inference that we are free to use it as liberally or parsimoniously as we please then the antirealist's double standard is not irrational.

1.10. Section Summary

It has been argued that, (i) the first interpretation of the no miracles argument is valid but not sound, (ii) the second interpretation of the no miracles argument is (plausibly) sound, but that it does not deliver the conclusion that the realist needs, and, (iii) the third interpretation of the no miracles argument (as an instance of IBE) is probably not sound (even assuming that IBE is legitimate), because the second premise is false, but even if it is a sound instance of IBE it is completely redundant, since the conclusion of the argument is arrived at by a simpler, ground level, IBE.

The more important issue in the scientific realism debate seems to be the question of whether or not the antirealist can rationally use IBE in some cases but refrain from its use in others and it has been suggested that if van Fraassen's approach to rationality is the right one then this is possible.

Consequently, these arguments do not appear to be compelling, and so Worrall's attempt to form a position that accommodates these arguments (as well as the antirealist's pessimistic induction) seems unnecessary. Moreover, in the next section it will be argued that the claim that the WESRist can accommodate the pessimistic induction is unconvincing.

2. The Pessimistic Induction

The pessimistic induction can be stated roughly as follows: many previously successful theories are (by our current lights) radically false; we should therefore infer that many of our currently successful theories will also appear to be radically false in the future; we thus have no reason to believe that any theory is even approximately true.[11]

[11] It should be born in mind that the antirealist's scepticism is (usually) restricted to only the theoretical parts of our theories, so when he claims that "many previously successful theories are (by our current lights) radically false" he means that they are radically false *with respect to the claims they make about the unobservable world.* Likewise what he recommends that we should infer is that many of our currently successful theories will also appear to be radically false in the future, *with respect to the claims they make about the unobservable world.*

The strength of the pessimistic induction depends on the amount of evidence in its favour, i.e. on the number of theories that were successful but are (by our current lights) radically false. Laudan (1981) has claimed that the evidence can be amassed ad nauseam and he cites a dozen or so examples of successful theories whose claims we now take to be decidedly false. However, as Worrall points out, he seems to be working with a loose notion of success. Many of his examples are of theories that merely accommodated previously known empirical results. Worrall argues that a theory is genuinely empirically successful only if it predicts a result that it was not engineered to yield.[12]

Worrall concedes, however, that there is at least one theory Laudan cites that does seem to have enjoyed the right sort of empirical success – Fresnel's wave theory of light. This theory was based on the idea that light is a wave transmitted by an all-pervading mechanical medium, the "luminiferous ether". It correctly predicted various results that it was not engineered to accommodate, most famously the existence of a white spot at the centre of the shadow cast by an opaque disc held in light diverging from a point source. Much of the discussion of the pessimistic induction in the literature has since focussed on this case study.

There have been two types of response to the pessimistic induction from the realist camp. The first response is to deny the cogency of the argument. This response is discussed in subsections 2.1 and 2.2. It is argued that although the standard version of the pessimistic induction might be undermined by (one of) these criticisms a non-standard form of the pessimistic induction is unaffected. The second response is to accept an attenuated version of the argument and move to a form of "selective" realism: to concede that many previously successful theories do appear to be radically false in some respect(s) and that there is no reason to think that our

[12] However, he does not insist that such a result need be (temporally) novel. So the prediction (or rather, retrodiction) of the advance of the perihelion of Mercury can count as a genuine empirical success of general relativity, because the theory was not engineered to yield the result, despite the fact that the result was known when Einstein devised the theory. Zahar (1973) has also emphasised the importance such predictions and retrodictions, calling them "heuristically novel".

current theories are any better off, but to argue that these previously successful theories nonetheless appear to be approximately true in some other respect(s) and so we can still reasonably hold that our current theories are also approximately true in this (these) latter respect(s). Variants of this response are discussed in subsections 2.3-2.5. Amongst these is Worrall's response to the pessimistic induction (discussed in subsection 2.5), which also constitutes his argument for moving from conventional realism to WESR. It is argued that none of these responses is particularly convincing. Finally, in subsection 2.6, a novel response to the pessimistic induction is proposed.

2.1. Lewis' Objection to the Pessimistic Induction

Lewis (2001) reconstructs the pessimistic induction as follows:

Premise: Most previously successful theories are radically false.[13]

Conclusion: Success is not a reliable indicator of approximate truth (where a reliable indicator is one that produces a low proportion of false positives and a low proportion of false negatives): it is not the case that both $P(Sx|\neg Ax) \approx 0$ and $P(\neg Sx|Ax) \approx 0$.

As Lewis notes, this argument is not valid. The fact that previously successful theories that are radically false outnumber previously successful theories that are approximately true is consistent with there being both a low proportion of false positives and a low proportion of false negatives, as long as there have been many more false theories than true theories. This is illustrated in figure 3.

[13] Note that this premise seems to be somewhat stronger than the premise of the standard pessimistic induction (as it has been presented in this thesis), which only claims that *many* previously successful theories are radically false. However, as Lewis does not object to the premise of the argument, this seems to be irrelevant.

However, given Lewis' interpretation of the pessimistic induction, the argument is not clearly inductive, nor does it clearly give grounds for pessimism. It seems that the following interpretation is more faithful:

Premise: Most previously successful theories are radically false.

Conclusion: Most currently successful theories are (probably) radically false.

However, one can also extrapolate from what Lewis says (p. 377) an argument that also seems to undermine this interpretation of the pessimistic induction. The argument runs as follows: because it does not follow from the premise that success is not a reliable indicator of approximate truth we may assume (as seems intuitive) that success *is* a reliable indicator of approximate truth. But given that success is a reliable indicator of approximate truth and that *most current theories are successful* then it follows that most currently successful theories are approximately true. (The only reason most previously successful theories were radically false is because most previous theories were unsuccessful: as is apparent from figure 3.)

As Lewis is aware, this response to this form of the pessimistic induction rests on the following contention: most previous theories were unsuccessful, and most current theories are successful.[14] As Saatsi (2005) points out, this contention is problematic. The problem is not that the claim appears to be false but that it is not at all clear how we could possibly judge whether it is true or false. For example, which theories are we supposed to consider when we try to decide whether or not most previous theories were unsuccessful?

> Should we count in only the theory proposals made by eminent scientists, or perhaps all the proposals actually

[14] Cf. "this commits the convergent realist to the empirical claim that successful theories were rare in the past and are common today" (Lewis, 2001, p. 377).

> published in scientific journals, or what? It is easy to imagine a variety of sociological factors, say, yielding scores of unsuccessful and false theories, directly affecting the notion of reliability at stake. But why should we care about *those* theories? It just seems that the debate...does not involve unsuccessful and false theories (or true yet unsuccessful ones, for that matter) in anything like the way Lewis projects. (Saatsi, 2005, p. 1096, original emphasis)

This suggests that before one can even tentatively endorse Lewis' objection to the pessimistic induction one needs a clearer idea as to what Lewis' crucial contention amounts to.

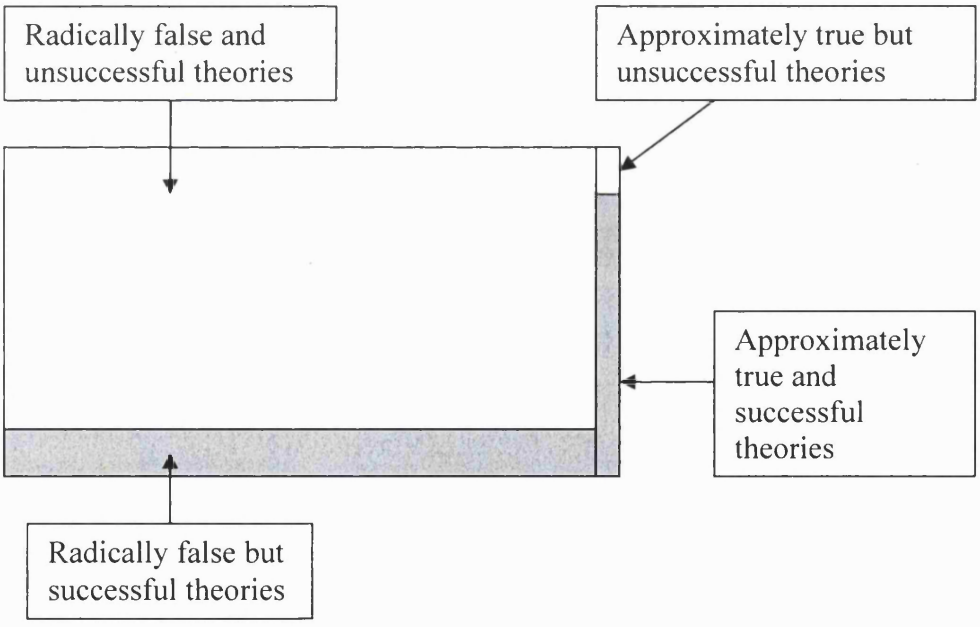

Figure 3: This illustrates a situation in which, (i) Most successful theories are radically false, but, (ii) Success is a reliable indicator of approximate truth ($P[Sx|\neg Ax] \approx 0$ and $P[\neg Sx|Ax] \approx 0$).

2.2. Lange's Objection to the Pessimistic Induction

Lange summarises his objection to the pessimistic induction as follows:

> For the sake of simplicity, suppose that at each moment in the history of science, the number of theories that are justly receiving wide acceptance remains the same (though the identities of these theories may, of course, differ from moment to moment). Let that constant number be N. (Admittedly, it is difficult to see how to count theories precisely, but since the pessimistic induction itself refers to "*most* theories adopted in the past", I too shall presume that theories have some criterion of individuation.) Obviously, theories that were accepted at some earlier moment, and are also accepted now, were never rejected in the meantime. Therefore, since theories currently believed false have generally experienced more rapid turnover than theories currently accepted, the past contains more room for theories currently believed false than for theories currently accepted. It would be very easy for there to be more than N theories that were accepted sometime in the past but have now been rejected as false. If, for each theory currently accepted, there were (for example) two predecessors that were once accepted but have since been discarded, then fully two-thirds of the theories accepted at some time or other are currently believed false (Lange, 2002, p. 283, original emphasis)

Lange's point can be made vivid with a little science fiction. Consider a species of alien that started developing mature scientific theories 10,000 years ago. Assume that over the last 10,000 years they have had at any given time 100 theories. Suppose that for the last 9,900 years their theories haven't changed at all (and that, despite rigorous testing, they have not found any empirical evidence to undermine any of their theories). Suppose that for the first 100 years after the dawn of their mature science the average lifespan of a theory was 20 years. Then there have been a total of 600 theories since the dawn of their science, most of which (we can assume) look radically false

(by their current lights). But it would obviously be silly for them to infer that the currently accepted 100 theories, that have survived 9,900 years will go the way of the other 500 theories, which lasted, on average, 20 years each.

Lange notes that the following non-standard form of the pessimistic induction would be immune to his objection:

Premise: At most past moments, most successful theories that were held at the time were radically false.

Conclusion: At the current moment, most successful theories that we hold are (probably) radically false.

This evades the objection because, "a current theory that has long been accepted will count not just once in the cumulative grand total, but rather once with regard to each of the past moments at which it was receiving wide acceptance." (Lange, 2002, p. 284). However, as Lange points out, it seems very doubtful that the premise of this non-standard pessimistic induction is true.

Is it possible to construct a version of the pessimistic induction that has a plausible premise and that avoids Lange's response to the standard pessimistic induction? Call the following argument the pessimistic induction*:

Premise: At most past moments, most successful theories that were held at the time and had been around for less than T were radically false.

Conclusion: At the current moment, most successful theories that we hold and have been around for less than T are (probably) radically false.

It seems that, if T is chosen appropriately (as perhaps about 100 years) then the pessimistic induction* (i) is immune to Lange's objection to the standard

pessimistic induction (ii) has a plausible premise and (iii) is interesting.[15] In this case we restrict our pessimism to those theories that are relatively recent, not extending it to those that have been around for a long time. This seems intuitively reasonable: we are very confident that the theory that the heart pumps blood around the body, or the theory that water is composed of hydrogen and oxygen in the ratio of 2:1 will not be overthrown. It seems that this is a version of the pessimistic induction that the realist needs to worry about.

2.3. Psillos' Response to the Pessimistic Induction

Psillos accepts that the general form of pessimistic induction is legitimate, but he argues that an attenuated version of the argument can be accommodated by moving to a "selective" realism. Psillos makes a distinction between the "idle constituents" of a theory and the "essentially contributing constituents". However, this terminology is somewhat misleading, since Psillos acknowledges that so-called "idle constituents" may have at least some role to play in the derivation of empirical predictions. The distinction between "idle constituents" of theories and "essentially contributing constituents" of theories is that the scientists who devise and use the theory accord a low degree of belief to the former and a high degree of belief to the latter. In this regard Psillos distinguishes between models and theories:

> A model of X comprises a set of assumptions which are not yet believed to describe X. On the contrary, a theory of X is the end product of scientific theorising. When something is advocated as a theory of X, the degree of belief that it correctly describes/explains X is, generally, high. (Psillos, 1999, p. 143)

[15] If T is chosen to be 0 seconds, then the premise is trivially true, but so is the conclusion, so the argument is not interesting.

Psillos claims that although assumptions about the ether were of some importance in deriving empirical predictions, the scientists who made these assumptions did not believe them to be true: the mechanical ether was just a "model". The sensible realist, Psillos thinks, will believe in theories (or at least in the approximate truth of theories) but not in models (perhaps the terms "modelling assumptions" and "theoretical commitments" would be more appropriate than "idle constituents" and "essentially contributing constituents").

The sensible realist could thus accommodate the following attenuated form of the pessimistic induction: *the "idle constituents" of* many previously successful theories are (by our current lights) radically false; we should therefore infer that *the "idle constituents" of* many of our currently successful theories will also appear to be radically false in the future; we thus have no reason to believe that *the "idle constituents" of* any theory are even approximately true.

Elsamahi (2005) objects to Psillos' approach by pointing out that even in cases where we can determine what degrees of belief the scientists who develop and use a theory assign to its various assumptions, "It does not seem possible to affirm with confidence that the degree of certainty or conviction expressed by a scientist toward one of his or her assumptions reflects the degree of epistemic significance of that assumption." (Elsamahi, 2005, p. 1358). This suggests that we ought to be sceptical of Psillos' approach even if it seems to be supported by the history of science. In any case, however, it does not seem to be so supported: the history of science seems to suggest that the "essentially contributing" constituents of theories are often later deemed false. The claim that Fresnel and others did not really believe in the existence of a mechanical ether can be questioned. Indeed, Laudan (1981, p. 1121) notes that Maxwell thought that "the aether was better confirmed than any other theoretical entity in natural philosophy". And Worrall says that:

> It can't, I think, sensibly be denied that Fresnel *believed* in the ether as a real, material medium. He refers to such a medium explicitly (and in explicitly "realist" terms) at

> various points in his scientific work. For example, in his famous (1818) "Prize Memoir" on diffraction, he characterises the general problem of diffraction as follows: "Given the intensities and the relative positions of any number of systems of light waves of the same wavelength, propagating in the same direction, to determine the intensity of the vibrations resulting from the concourse of these different systems, that is to say, the velocity of the oscillations of the molecules of the ether" (248). In his (1822, 136) he stated that the properties of polarised light are simply explained on his theory "by supposing that, in light waves, the oscillations of the molecules of the ether are executed at right angles to the rays". (Worrall, 1994, p. 336, original emphasis)

Moreover, as Psillos himself notes (1999, p. 143), the difference between models and theories, or between "idle constituents" of theories and "essentially contributing" constituents is only a matter of degree (scientists will generally have some degree of belief in a "model" and less than complete confidence even in a "theory"). But this doesn't seem to cohere with the view that Psillos in advancing. It suggests that there are *not* well-defined portions of theories that we can identify as those that it is rational to believe to be (approximately) true and others that we can identify as those that it is rational to be more sceptical about. Rather, it suggests that there is a continuum from theoretical statements that it is rational to assign a very high degree of belief through to theoretical statements that it is rational to assign a very low degree of belief.

2.4. Kitcher's Response to the Pessimistic Induction

Kitcher (1993, pp. 133-149) responds to the pessimistic induction in a similar way to Psillos. He argues that, "The history of science does not reveal to us that we are fallible in some undifferentiated way. Some kinds of

claim endure, other kinds are likely to be discarded as inaccurate." (Kitcher, 1993, p. 138) He distinguishes between "working posits" and "presuppositional posits". The former are theoretical claims that are used to derive empirical predictions. The latter are theoretical claims that play no role in the derivation of empirical predictions. His claim is that the sensible realist should only hold that the former are (approximately) true.[16] The difference between Psillos' "idle constituents" and Kitcher's presuppositional posits is that, (i) Psillos' "idle constituents" may play some role in the derivation of empirical predictions, and, (ii) Psillos claims that the "idle constituents" of theories are actually identified in advance by the scientists who develop and use the theories (the scientists place a lower degree of belief in "idle constituents" than in "essentially contributing" constituents).

Thus, according to Kitcher, the sensible realist could accommodate the following attenuated form of the pessimistic induction: *the presuppositonal posits of* many previously successful theories are (by our current lights) radically false; we should therefore infer that *the presuppositonal posits of* many of our currently successful theories will also appear to be radically false in the future; we thus have no reason to believe that *the presuppositonal posits of* any theory are even approximately true. Clearly, however, Kitcher's response can only work if the history of science supports this attenuated version of the pessimistic induction, rather than the standard version.

Let's return to the Fresnel/Maxwell case. Is the existence of the ether merely a presuppositional posit of Fresnel's theory? Psillos' (1999, pp. 130-145) study of nineteenth century optics suggests that the concept of a mechanical ether was at least heuristically useful in deriving empirical predictions. Worrall is also at odds with Kitcher here:

[16] In a similar vein, Chakravartty (1998, 2004) makes a distinction between "detection properties" and "auxiliary properties" as the basis of his "semirealism": his position is further discussed in appendix 1.

> the suggestion that the material ether was an idle component in Fresnel's system is significantly misleading. Whittaker [1951] claims, for example, that "geometrical reasoning" led the way in Fresnel's development of the wave surface in birefringent crystals. However, this "geometrical reasoning" itself did not spring from nowhere, but was based on Hooke's law, Huygens's principle, the principle of superposition ("coexistence of small movements") and other assumptions of a *general* mechanical kind. Moreover, although Fresnel's extension to cover all crystals of Huygen's famous sphere/spheroid construction for the two refracted beams can be characterised mathematically as a process of putting two equations together (by introducing three parameters for Huygens's two), that process was in turn undoubtedly guided by Fresnel's "realist" belief that there could only be one light-carrying medium and the "natural" assumption that, in the general case, the coefficients of elasticity of that medium in the three orthogonal directions in space will be different. Fresnel *did* get some important heuristic mileage out of certain general mechanical-dynamical ideas concerning some sort of mechanical medium with some sort of vibrating parts. (Worrall, 1994, pp. 336-337, original emphasis)

Studies of other historical cases seem to further undermine Kitcher's approach. Elsamahi claims that, "There are many...examples in the history of science for [sic] assumptions that led to the predictive success of their theories but were subsequently considered false" (Elsamahi, 2005, p. 1355). He gives two such examples. Firstly, he argues (Elsamahi, 2005, pp. 1353-1355) (against Psillos, 1999, pp. 119-121) that the (subsequently rejected) assumption of the caloric theory of heat that heat consists of material particles was non-idle (he argues that Laplace used this assumption to predict the speed of sound in air). Secondly, he notes that:

> Slater's theory of oxidative phosphorylation, which was proposed and accepted in the 1950s, postulated that a high-energy intermediate compound transfers energy from the oxidation of fuel substances (e.g., sugars and fatty acids) to the reaction that leads to forming ATP molecules in mitochondria. The latter molecules become the direct source of energy for cellular reactions. This assumption of an intermediate high-energy molecule was responsible for successful predictions in the areas of muscle fiber and pumping sodium and potassium across cell membranes. Yet this assumption was declared false around twenty years later because credible experimental data showed that no such intermediate molecule exists. (Elsamahi, 2005, p. 1355)

It thus seems very doubtful that Kitcher's response to the pessimistic induction is supported by the historical evidence.

2.5. Worrall's Response to the Pessimistic Induction

It has been argued that (i) (a form of) the pessimistic induction has some force and (ii) the responses conventional realists have made to the pessimistic induction are unconvincing. It thus seems that it would be a notable virtue of WESR if it could accommodate the pessimistic induction successfully.

Worrall argues that although much of the theoretical machinery of Fresnel's theory was abandoned in the transition from Fresnel's to Maxwell's theories of light, there was nonetheless continuity of more than just empirical content: Fresnel's equations were retained, and with them, the structure of Fresnel's theory. Hence, what this case supports is not the standard form of the pessimistic induction but rather something more like the following attenuated pessimistic induction: *in their unRamseyfied forms*, many previously successful theories are (by our current lights) radically

false; we should therefore infer that *in their unRamseyfied forms* many of our currently successful theories will also appear to be radically false in the future; we thus have no reason to believe that *in its unRamseyfied form* any theory is even approximately true.[17] This form of the pessimistic induction is compatible with the WESRist's realist thesis (because it is compatible with the view that the Ramsey-sentences of mature theories are approximately true). Hence, Worrall contends that WESR is a form of realism that allows one to accommodate the pessimistic induction (as well as the no miracles argument).

A possible objection to this argument is that although the Fresnel/Maxwell episode supports WESR other episodes of scientific theory change do not, since it is unusual for a theory to adopt the equations of its predecessor unaltered. Worrall concedes that complete structural continuity across scientific revolutions of the sort exemplified by the Fresnel/Maxwell case is the exception but claims that a substantial degree of structural continuity is the norm:

> This particular example is in fact unrepresentative in at least one important respect: Fresnel's equations are taken over completely intact into the superseding theory – reappearing there newly interpreted but, as mathematical equations, entirely unchanged. The much more common pattern is that old equations reappear as *limiting cases* of the new – that is, the old and new equations are strictly inconsistent, but the new tend to the old as some quantity tends to some limit. (Worrall, 1989, p. 160, original emphasis)

[17] Although this is clearly the form of the pessimistic induction that the WESRist needs it should be noted that Worrall doesn't mention Ramsey-sentences at all in the papers in which he discusses these issues (1989, 1994). Rather, he says that the equations of a theory survive ([partially] reinterpreted). However, taking the view that the equations of a theory survive (as opposed to the theory's Ramsey-sentence) seems hard to justify philosophically because it seems to be only little more than a contingent fact that parts of a theory happen to be expressed mathematically (as equations) and other parts do not.

Since the WESRist only claims that a mature theory's Ramsey-sentence is typically *approximately* true, a substantial degree of structural continuity is all the WESRist would expect, so theory changes along these lines seem to be compatible with WESR.

However, Mark Newman (not to be confused with Max Newman, another critic of structural realism) objects that:

> this account is not sufficiently clear. It is far from obvious that we can successfully compare the equations of quantum mechanics with those of classical dynamics. In the former case we are dealing with operators operating on rays in Hilbert space, in the latter we are talking of continuous real valued functions. In what ways and to what degree can these equations be said to be similar? (Newman, 2005, p. 1378)

Redhead (2001a) has answered this question on Worrall's behalf. Discussing the shift from Galilean space-time to Minkowski space-time brought about by special relativity and the shift from classical to quantum mechanics he notes that although:

> we do not have continuous transformation of structure as we move away from the classical structure characterised by $1/c = 0$ in the case of relativity, and $h = 0$ in the case of quantum mechanics to the new structures with non-zero values of $1/c$ or h (Redhead, 2001a, p. 88)

nonetheless:

> 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 structures do seem to arise in a natural, if not inescapable, way out of the old structures. (Redhead, 2001a, p. 88)

There is, nonetheless, a serious difficulty facing Worrall's argument for WESR, which is that, despite first appearances, the sort of continuity of structure displayed in the Fresnel/Maxwell case does not support Worrall's attenuated pessimistic induction. This is because not all the structure of Fresnel's theory is (even approximately) preserved in Maxwell's theory.

Worrall acknowledges that Fresnel was committed to the existence of a "luminiferous ether" with various properties: it was taken to be all pervading, elastic, solid etc. Worrall further argues that this ether was not an idle posit but reasoning about it enabled Fresnel to arrive at his equations (the equations that reappear in Maxwell's theory). So it seems that Worrall would agree that (an important) claim of Fresnel's theory was that, "The ether is an all-pervading, elastic, solid and is the medium through which light is transmitted" but would contend that this claim is just the sort of claim that gets left behind in scientific revolutions.

There is, however, a serious problem with this contention, which is that *any* (contingent) claim carries with it a structural commitment (in that it will not be possible to satisfy it in every structure and thus, if the claim is part of a theory, it puts a restriction on the possible structures that satisfy the theory) so there is no such thing as a (contingent) claim that expresses *only* the sort of content that the WESRist is happy to see abandoned. Consider, for example, the belief that the ether was a solid. If the ether was thought to be a solid then it was presumably thought that any finite part of the ether would consist of finitely many constituent particles. In so far as the electromagnetic field can be said to have constituent parts at all they are space-time points, and a finite part of the electromagnetic field consists of continuously many. This suggests that the ether and the electromagnetic field are structurally very dissimilar.

All we seem to have in the Fresnel/Maxwell case is structural continuity with respect to the parts of the theory (Fresnel's equations) that

happen (as a matter of fact) to be expressed mathematically, not (even approximate) continuity with respect to the entire theory structure, which is what the WESRist would expect.

In the light of this consideration, the WESRist might be tempted to argue that one should be a selective structural realist – a realist not with respect to the structure of the whole of a successful theory but only with respect to the structure of certain parts of a successful theory. This response is similar to the responses of Psillos and Kitcher, discussed in the previous two subsections. Perhaps a response along these lines can be made to work for the WESRist (and perhaps also for the conventional realist) but, as we have seen, there are problems that make this response unconvincing. In particular, it seems to be difficult to provide an independent criterion that identifies the parts of theories about which we should be (structural) realists, so the move looks rather ad hoc.

2.6. A Novel Response to the Pessimistic Induction

It will be argued in this subsection that, despite the fact that the conventional realist's and WESRist's responses to the pessimistic induction seem unconvincing, there is still room for a degree of optimism in the face of the pessimistic induction. A little more science fiction will help explain why. Consider a species of alien that started developing mature (i.e. successful) scientific theories 1,000 years ago. Assume that over the last 1,000 years they have had, at any given time, 100 theories. Let's suppose that, of all possible successful theories they could devise, only 10% are approximately true. But let's also assume that, if a successful theory is radically false, then there is a 50% chance that, over the course of 100 years it will be "found out" and replaced with another successful theory, whereas, if a successful theory is approximately true, it is bound to be retained.

At the dawn of their science they have 100 successful theories and (typically) 10 of these are approximately true, while 90 are radically false. After 100 years they will have (typically) replaced 45 (50% of 90) of the radically false theories with 45 new successful theories, of which (typically) 4.5 will be approximately true. They will have retained the 10

approximately true original theories (along with 45 of the radically false original theories). So after 100 years their science will consist of 100 theories, of which typically 14.5 will be approximately true and 85.5 will be radically false. After another 100 years they will have (typically) replaced 42.75 (50% of 85.5) of their radically false theories with 42.75 new successful theories, of which (typically) 4.275 will be approximately true. They will have retained the 14.5 approximately true old theories (along with 42.75 of the radically false old theories). So after 200 years their science will consist of 100 theories, of which typically 18.775 will be approximately true and 81.225 will be radically false.

The change in the percentage of approximately true theories held by the aliens over their 1,000 year history is illustrated in figure 4a. In this case both the premise and the conclusion of the pessimistic induction* are true.[18] It is true that:

At most past moments, most successful theories that were held at the time and had been around for less than 100 years were radically false.

because, at any past moment, typically 90% of the successful theories that have been around for less than 100 years are radically false. It is also true that:

At the current moment, most successful theories that we hold and have been around for less than 100 years are (probably) radically false.

because, at the current moment, typically 90% of the successful theories that have been around for less than 100 years are radically false.

But, despite the fact that the conclusion of the pessimistic induction* is true of the aliens' science, the aliens have reason for optimism about their science, because there is definitely progress: a greater percentage

[18] This response would also work as a response to the standard pessimistic induction. Attention is focussed on the pessimistic induction* here, because it has been argued that this is more worrisome for the realist than the standard pessimistic induction.

of it is becoming approximately true as time goes by. Indeed, in 4,000 years time (typically) over 90% of the theories that they hold will be approximately true, as illustrated in figure 4b.

What this shows is that, even if the pessimistic induction* is sound, and even if one is not convinced by the conventional realist and WESRist responses to the pessimistic induction,[19] then, although we cannot uphold conventional realism (or WESR), which holds that it is rational to believe that (the Ramsey-sentences of) our mature theories are approximately true (because, in the situation envisaged most mature successful theories are not approximately true, nor need their Ramsey-sentences be) there could still be room for a degree of optimism: it would still be possible that science is progressing, even with regard to its claims about the unobservable world. So one could still hold a position that was essentially realist in spirit.

2.7. Section Summary

It has been argued that Lewis' critique of the pessimistic induction rests on a highly questionable assumption and that although Lange's critique does undermine the argument as it is usually stated, one can formulate a version of the pessimistic induction that is immune to Lange's critique of the original. It has also been argued that the conventional realist's and the WESRist's responses to the pessimistic induction are unconvincing. This seems to be a problem for WESR (and conventional realism). (However, it has been argued that one could still uphold a position that was broadly realist in spirit.)

3. Chapter Summary

[19] If one of these responses to the standard pessimistic induction could be made convincing there is no reason to suppose that it would not work as a response to the pessimistic induction*.

It has been argued that the no miracles argument is not compelling and so that WESR's ability to accommodate the argument is no particular virtue of the position. It has been further argued that the WESRist's response to the pessimistic induction is unconvincing. This suggests that Worrall's arguments for WESR do not provide sufficient motivation for one to adopt the position.

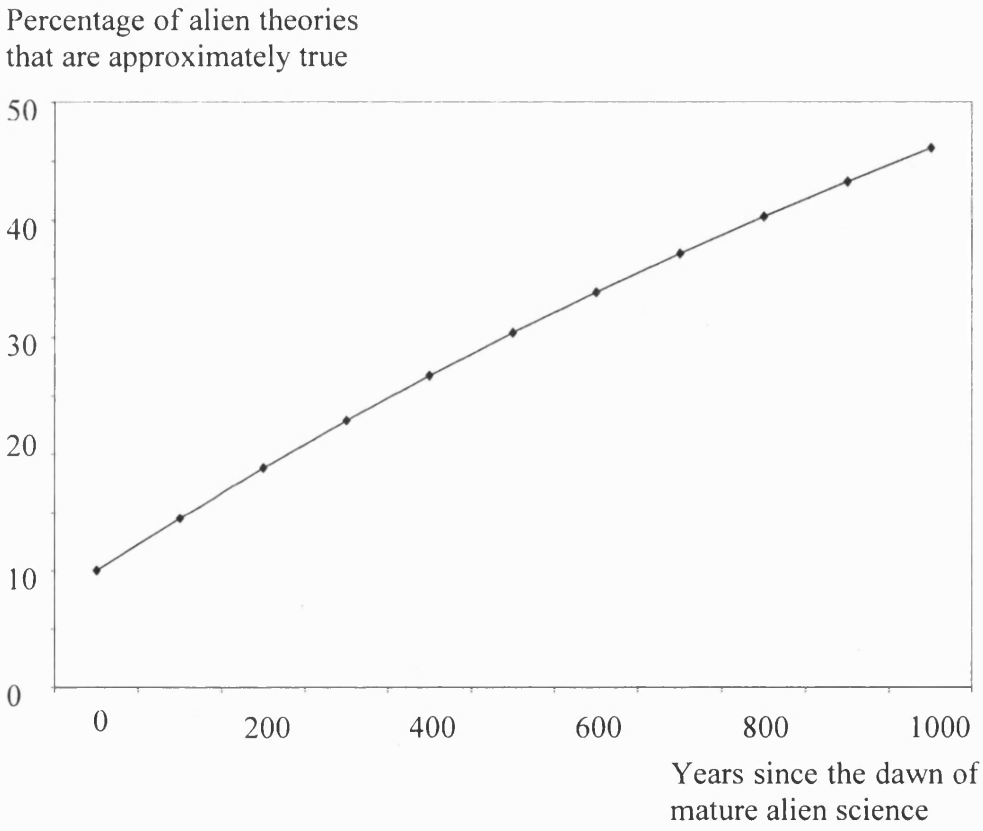

Figure 4a: The development of alien science over 1,000 years, assuming 10% of successful theories are approximately true and there is a 50% chance that a radically false successful theory will be "found out" in any 100 year period.

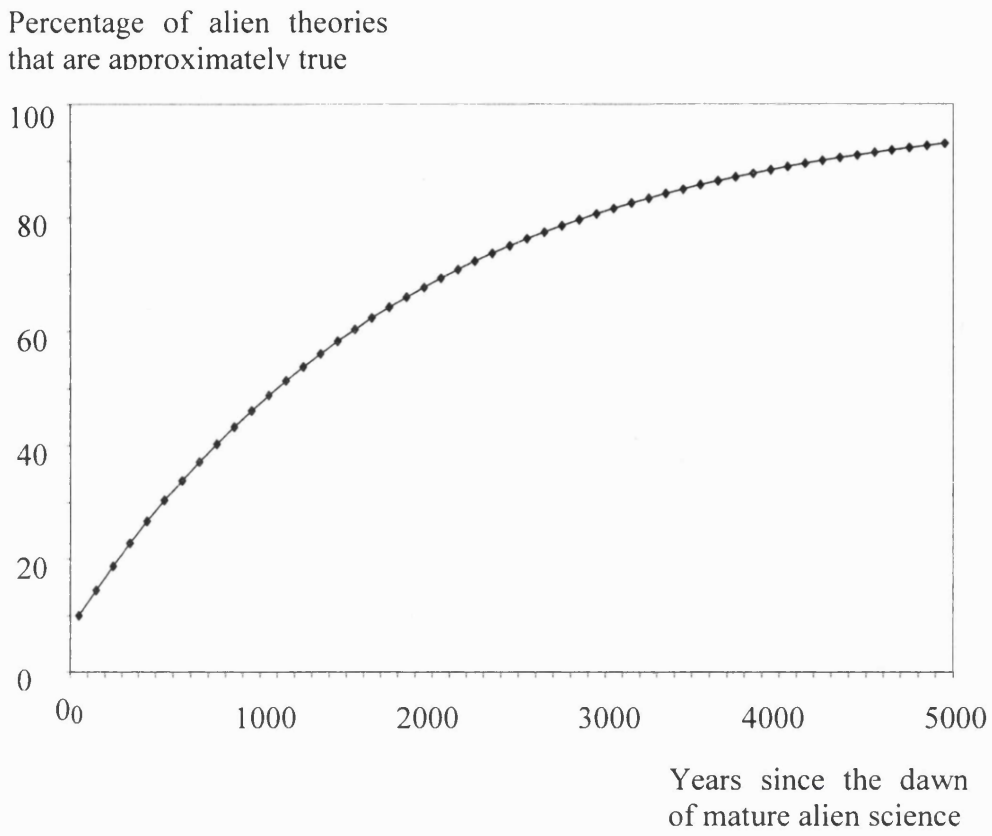

Figure 4b: The development of alien science over 5,000 years, assuming 10% of successful theories are approximately true and there is a 50% chance that a radically false successful theory will be "found out" in any 100 year period.

Chapter 2

An Objection to Weak Epistemic Structural Realism

In the previous chapter it was argued that Worrall's argument in favour of WESR is not compelling. That does not of course mean that WESR is not the correct position in the scientific realism debate. In this chapter an argument that suggests that WESR is probably not a tenable position is put forward. (A discussion of arguably the most devastating criticism of WESR, Newman's objection [Newman, 1928] is postponed for consideration in chapter 4, since it also affects SESR, which is the subject of chapter 3.)

One of the conventional realist's objections to instrumentalism is that it is impossible to draw a distinction between theoretical terms and observational terms that is both sharp and principled. (The instrumentalist seems to need such a distinction, since he claims that there is a significant difference in kind between these terms: observational terms refer, whereas theoretical terms do not.) This objection seems to work equally well against the WESRist: if there is no sharp and principled distinction between theoretical and observational terms then there seems to be no principled way to decide whether or not a given term ought to be Ramseyfied. In this chapter the arguments for the claim that there is no sharp and principled distinction between theoretical terms and observational terms are outlined, as are various attempts that have been made to defeat or evade these arguments. The discussion concentrates on the consequences for WESR.

1. The Observational Term/Theoretical Term Distinction

1.1. Putnam's Critique of the Distinction

Putnam (1962) argues that it is not possible to distinguish observational terms from theoretical terms as follows. If observational terms are terms that can only be used to refer to (sets of [tuples of]) observable objects, then

there are none. Even a paradigm observation term, like "bigger than", can perfectly meaningfully be used to talk about unobservable objects, as one might state that atoms are bigger than electrons. On the other hand, if theoretical terms are those that can only be used to refer to (sets of [tuples of]) unobservable objects, then supposedly theoretical terms like "electric charge" become observational, because observable objects (e.g. wires) can be electrically charged. (Putnam notes that the presence of an electric charge can be determined through a simple observation in some cases – e.g. by touching a live wire – but this does not seem to be exactly to the point here.) Putnam's claim is that most of the non-logical terms in a real language can be used to refer to both observable and unobservable objects, so there is no reasonable division of non-logical terms in a real language into "observational" and "theoretical".

However, as Suppe (1977) points out, we can contrive an artificial language in which some terms may be used to refer only to observable objects and others to refer only to unobservable objects. For example, where normal English uses the single term "red" the artificial language will use two terms, the first, say red_o, for observable red objects (e.g. red tables) and the second, red_t, for unobservable red objects (e.g. red blood cells). In such a language observational terms are those that refer to observable objects, properties or relations (where observable properties are taken to be sets of observable objects and observable relations are taken to be sets of tuples of observable objects) and theoretical terms are taken to be those that refer to unobservable objects, properties or relations (where unobservable properties are taken to be sets of unobservable objects and unobservable relations are taken to be sets of tuples, whose members include at least one unobservable object [so relations that connect observables to unobservables are assumed to be theoretical]).

That is one possible response. It is not an intuitively appealing one, because the language Suppe describes seems to be *very* different from the language we actually use (for scientific, and other, purposes). However, it is also not the only possible response to Putnam. Putnam assumes that either a predicate is observational if and only if its extension contains only (tuples of) observable objects or that a predicate is theoretical if and only if its

extension contains only (tuples of) unobservable objects. What Putnam's argument shows is (at best) that this assumption is not tenable. But why make this assumption?

The assumption seems to rest on the idea that we can determine whether an object is observable or unobservable, but that we cannot (directly) determine whether a property or relation is observable or unobservable, so whether or not a property or relation is unobservable must be defined in terms of whether or not the objects to which it applies are. But this idea is surely wrong. Surely we know that *red* is an observable property as surely and directly as we know that *Hilary Putnam* is an observable object. And surely we know that *being a collection of superstrings* is an unobservable property as surely and directly as we know that *the nearest black hole* is an unobservable object. It is true that there are some properties and relations that we might hesitate to class either way. But it is equally true that there are some objects that we might hesitate to class either way (small particles of dust, for example). So the idea that we must define the observability or unobservability of properties and relations in terms of the observability or unobservability of the objects to which they apply is at least questionable.

It seems then that there are at least two possible responses to Putnam's argument against the observational term/theoretical term distinction. However, the observational term/theoretical term distinction clearly presupposes a distinction between observable objects, properties and relations and unobservable objects, properties and relations and it has also been argued that this distinction cannot be adequately drawn. One of these arguments is discussed in the next subsection.

1.2. Maxwell's Critique of the Observable/Unobservable Distinction

Maxwell (1962) argues that the observable/unobservable distinction is problematic for essentially two reasons:[1]

[1] Observability is a matter of degree. There is a continuous transition between the observable and the unobservable, so drawing a sharp line between the observable and the unobservable will be arbitrary and an arbitrary distinction is inappropriate for most philosophical uses. In particular, it is inappropriate for the WESRist's purposes: the WESRist takes a fundamentally different attitude to observational and theoretical terms (he thinks that the latter, but not the former, need to be Ramseyfied) and such a fundamental difference in attitude would be unjustified if the distinction between observables and unobservables was arbitrary.
[2] There is nothing that is, in principle, unobservable: over the course of time entities tend to become observable (or rather, more observable), through the advance of technology. For instance, microbes became more observable through the invention of the light microscope, atoms through the invention of the electron microscope and electrons through the invention of the cloud chamber. So, if theoretical terms are those that refer to objects, properties and relations that are in principle unobservable, then there are none, and so WESR collapses into conventional realism (because all terms are observational, so none needs to be Ramseyfied). On the other hand if theoretical terms are those that refer to objects, properties and relations that are currently unobservable then WESR may collapse into conventional realism in the future (if at some future time everything is observable).

In the introduction to this thesis, observables were roughly characterised as those external objects, properties and relations that have a direct counterpart in internal experience. This characterisation falls foul of both of Maxwell's arguments. In the first place there are some cases in which it is hard to judge whether or not we have a certain experience (e.g. where we are not sure

[1] Note that Maxwell himself became an ESRist by 1965. But his brand of ESR was a form of SESR, which does not rely on a distinction between the observable and the unobservable, but on a distinction between the internal and the external.

whether or not we can hear a high-pitched whistle). In such cases it is hard to say whether or not the putatively corresponding external object, property or relation has a direct counterpart in internal experience or not. Furthermore, it is not clear what counts as a direct counterpart in some cases: is the sense-data track in a sense-data cloud chamber a direct counterpart to an electron? In the second case, external objects, properties and relations that previously had no direct counterpart in internal experience (e.g. microbes) can come to have direct counterparts through the advance of technology (e.g. the invention of the microscope). In the light of this it seems doubtful that there is anything that could never, in principle, have a direct counterpart. Of course, that does not mean that we cannot find some form of the distinction that answers Maxwell's objections.

1.3. Van Fraassen on the Observable/Unobservable Distinction

The most notable attempt to draw the observable/unobservable distinction in recent times was made by van Fraassen (1980), who needs the distinction for his constructive empiricism. Van Fraassen concedes that there is a continuum between observables and unobservables, however, he argues that this only shows that "observable" is a vague predicate (like almost all predicates in natural language) and that this doesn't mean it isn't useful. He thus suggests that a fuzzy distinction between observables and unobservables is good enough. Of course, whether or not a fuzzy observable/unobservable distinction is good enough depends on what one wants to do with the distinction. A fuzzy distinction is inappropriate for the WESRist's purposes: the WESRist Ramseyfies theoretical terms but not observational terms and Ramseyfication is an all or nothing matter (a term can't be partially Ramseyfied) so he assumes a sharp distinction between observational terms and theoretical terms.[2]

[2] In fact, it seems that van Fraassen also needs a sharp distinction. Van Fraassen needs the distinction because he suggests that we ought to believe in the existence of observables but be agnostic with regard to the existence of unobservables. On the face of it this suggests that he could work with a fuzzy distinction by matching the fuzziness of the distinction with the fuzziness of belief. Intuitively, belief (and agnosticism) is not an all or nothing matter,

Although he protests that he doesn't need a sharp observable/unobservable distinction, in places van Fraassen can be read as attempting to draw one. He characterises the distinction as follows: "X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it" (van Fraassen, 1980, p. 16) and elaborates by commenting that:

> The human organism is, from the point of view of physics, a certain kind of measuring apparatus. As such it has certain inherent limitations – which will be described in detail in the final physics and biology. It is these limitations to which the "able" in "observable" refers – our limitations, *qua* human beings. (van Fraassen, 1980, p. 17)

This suggests the following characterisation of "observable":

X is observable if and only if there are circumstances (in which we retain our capabilities and limitations as human beings) such that, if X is present to us under those circumstances, then we can observe it without the aid of instruments.[3]

It will be noted that this definition is not at all in line with our normal use of the word observable: we surely do think that we observe objects with microscopes that we do not observe unaided. However, this is no real objection in this context: if the characterisation succeeds in drawing a sharp

---

but comes in degrees. However, if van Fraassen doesn't make a sharp binary distinction between belief and agnosticism it is hard to see how his position differs from realism (even a realist would presumably admit to *some* doubt about the existence of gluons). So, as he attempts to found this divide on the observable/unobservable distinction, he must also draw the latter sharply.

[3] An objection that is sometimes raised against van Fraassen's characterisation is that it implies that dinosaurs (for example) are unobservable, since (being unable to travel back in time) we are unable to observe them. However, this objection has no force against the characterisation attributed to van Fraassen here, since if dinosaurs were present to us, we could observe them.

and principled distinction then it is good enough for the WESRist's purposes.

A more serious objection to this characterisation (in this context) is that, on the face of it, this definition does not evade Maxwell's first objection, because it seems that there are cases in which it is not clear whether or not we can observe something without the aid of instruments (e.g. a high pitched whistle). Furthermore, it makes what is observable relative not only to our status as human beings but also to our status as individuals (some of us are colour blind, some have more acute hearing than others, and so on). Perhaps these objections might be evaded by van Fraassen's reference to the ominous sounding "final biology", which is perhaps meant to tell us exactly what the paradigm human being can observe.[4] It is also not exactly clear what counts as an instrument: is the lensing effect of a temperature gradient in the air an instrument? Perhaps our "final physics" will answer these questions. Or perhaps they are more properly in the domain of our "final engineering". So it is at least doubtful that this characterisation succeeds in drawing a sharp line. On the other hand, the proposal certainly deals with Maxwell's second argument, because in assessing what we can in principle observe, we consider only what we can observe with the unaided senses, which will be unaffected by the advance of technology (let us ignore the fact that it may vary as humans evolve) and relative to which it is clear that not everything is observable.

In fact, even if this characterisation evaded both of Maxwell's arguments Churchland (1982) has argued that it would still fail to draw a distinction between observables and unobservables that justified a fundamental difference in attitude towards the two because, although the distinction that has been drawn is not exactly arbitrary, it does not appear to be epistemically relevant. To make his point Churchland asks us to consider a species of alien who are:

[4] It is certainly possible to doubt the plausibility of this vision of a final biology. However, there is no need to make a fuss about this point because there is a more decisive objection to van Fraassen's proposal, which is discussed in the next paragraph.

> able to observe a domain normally closed to us: the microworld of virus particles, DNA strands, and large protein molecules. Specifically, suppose a race of humanoid creatures each of whom is born with an electron microscope permanently in place over his left "eye". The scope is biologically constituted, let us suppose, and it projects its image onto a human-style retina, with the rest of their neurophysiology paralleling our own. (Churchland, 1982, pp. 43-44)

According to the above definition, virus particles etc. are observable to these creatures. Yet the definition denies that virus particle etc. are observable to us, "*even though we can construct and even if we do construct electron microscopes of identical function, place them over our left eyes, and enjoy exactly the same microexperience as the humanoids*" (Churchland, 1982, p. 44, original emphasis). Churchland's point is that there does not appear to be an epistemically relevant difference between being born with an organ that functions exactly like an electron microscope and being born into a society that has developed the use of electron microscopes. Yet, according to this definition, DNA molecules (etc.) are observable in the former but not the latter case. Hence the observable/unobservable distinction, so characterised, does not mark out a significant difference in the epistemic status of observables and unobservables.

Of course, even if it is true that the proposal discussed here cannot be made to work it does not follow that it is impossible to draw a sharp and principled observable/unobservable distinction, but it does at least raise a doubt as to the possibility, and the WESRist should find this, at the very least, disquieting, because his entire project seems to rest on the assumption that we can make a sharp and principled distinction between observational and theoretical terms, and our ability to do this rests on our ability to draw a sharp and principled distinction between observables and unobservables. In sections 2 and 3, two possible ways in which the WESRist might attempt to evade this problem by replacing the observational term/theoretical term distinction by a somewhat different distinction are considered.

2. The Old Term/New Term Distinction

An option that might tempt the WESRist is to make a distinction not between observational and theoretical terms but between old and new terms (and Ramseyfy the new terms). New terms would be defined as those that have recently been introduced into scientific discourse by a new theory and old terms as any others, including terms that were once new. Lewis (1970) and Hempel (1966) have both advocated making a distinction between old terms and new terms, and the so-called "Munich structuralists" (e.g. Balzer et al, 1987) have advocated a similar distinction.

2.1. Hempel and Lewis

Lewis (1970) proposes that whenever a new (or as he puts it, theoretical) term is introduced by a theory it should be definable using old terms (some of which may once have been new/theoretical terms). Suppose that some theory, T, introduces a number of new terms. According to Lewis these terms "should denote the components of the unique realization of T if there is one, and should not denote anything otherwise" (Lewis, 1970, pp. 437-438). For example, consider a toy theory that consists of the claim:

$\forall x(O_1x \leftrightarrow O_2x \,\&\, Nx)$

where O_1 and O_2 are old terms, and N is a new term introduced by the theory. According to Lewis, N is defined as follows:

$N = X$ if and only if $\forall x(O_1x \leftrightarrow O_2x \,\&\, Xx) \,\&\, (\forall Y \forall x[O_1x \leftrightarrow O_2x \,\&\, Yx] \rightarrow X = Y)$

That is, N denotes the unique property, X, that satisfies the open formula $\forall x(O_1x \leftrightarrow O_2x \,\&\, Xx)$. If there is no such unique property, then N denotes nothing.

Lewis' proposal would not make much sense if one's intention was to go on to Ramseyfy the new terms in the normal way: there would be little point in being able to define a term if one was in any case going to replace it with an existentially bound variable. However, Lewis' proposal does cohere with a non-standard approach to Ramseyfication, whereby one binds the Ramseyfied variables not with standard existential quantifiers but with quantifiers that assert the existence *and uniqueness* of the referents of the variables that they bind. For example, consider the above toy theory. It's standard Ramsey-sentence is:

$\exists X \forall x(O_1x \leftrightarrow O_2x \,\&\, Xx)$

On the non-standard approach this becomes:

$\exists! X \forall x(O_1x \leftrightarrow O_2x \,\&\, Xx)$

which is read "There is exactly one (property) X, such that...".[5] This coheres with Lewis' approach in that this non-standard Ramsey-sentence would be true if and only if the original theory is true, where the term N in the original theory is defined as Lewis proposes.

Lewis' proposal suggests that scientific languages consist of a hierarchy of levels, where each level contains terms that were once newly introduced by some theory (except, perhaps, the lowest level), the newest terms being at the highest level. The terms at the highest level can be replaced without loss (except of economy of expression) by terms from lower levels (being definable using these terms). So terms at the highest level are really redundant. Having got rid of terms at the highest level we

[5] The new quantifier is definable using the standard logical terms (including equality). The formula $\exists! X\alpha$ can be treated as an abbreviation of the formula $\exists X(\alpha \,\&\, \forall Y[\alpha^X_Y \rightarrow X = Y])$, where Y is a variable that does not occur free in α, and α^X_Y is the formula obtained from α by replacing every free X in α by a Y. So the sentence $\exists! X \forall x(O_1x \leftrightarrow O_2x \,\&\, Xx)$ can be treated as an abbreviation of $\exists X \forall x(O_1x \leftrightarrow O_2x \,\&\, Xx \,\&\, \forall Y[\forall x(O_1x \leftrightarrow O_2x \,\&\, Yx) \rightarrow X = Y)$.

can then proceed to get rid of terms at the next highest level. After all, these terms were once new, and so were (and thus, still are) definable using older terms (i.e. terms from lower levels). Clearly, by repeating this process we can get rid of all non-logical terms, except those at the lowest level. Ultimately then *all* terms can be defined using only these lowest level terms.

Lewis' multi-partite distinction between terms at different levels thus reduces to a bipartite distinction between those terms at the lowest level and those terms at other levels (all of which can ultimately be defined using terms from the lowest level and are thus ultimately redundant). Strangely, Lewis does not mention this obvious consequence of his proposal.[6] Consequently, he does not address an obvious question that it suggests: what is special about the lowest level terms (that means that they can be understood directly, whereas everything else must be ultimately defined using them)? A possible answer would be that they are observational terms but if this answer is given then Lewis' account relies on the observable/unobservable distinction after all.[7]

However, this consequence *has* been noted by Cruse and Papineau (2002), who advocate Lewis' approach. They deny that the lowest level terms must be observational terms, arguing as follows:

> The only requirement now being imposed on non-theoretical [i.e. the lowest level] terms is that they are not understood as denoting just those things which satisfy some particular theory. It would require extra empiricist presuppositions to infer from this that the relevant vocabulary must be 'observational' in any substantial

[6] It has been suggested to me that Lewis might not accept this consequence, on the grounds that there needn't be a lowest level of terms. However, as scientific languages are finite there can only be finitely many levels of terms. Hence, if no level is lowest, the levels must form a closed loop: apparently low level terms must be defined in terms of apparently high level terms. This is surely not what Lewis had in mind.

[7] Lewis may have been willing to accept this, but it is obviously not acceptable if one's reason for adopting the old term/new term distinction is to avoid the need for the observational term/theoretical term distinction.

> sense. Without prior empiricist prejudices, why not allow that a term could fail to be defined in a theory, and yet be neither observational nor logical? Antecedently understood [i.e. the lowest level] terms could thus refer to such substantial non-logical relations as causation or correlation, or indeed to many kinds of unobservable things. (Cruse and Papineau, 2002, pp. 181-182)

Of course, it may be possible to give some other answer that explains why these supposedly special terms can be directly understood and others cannot, but if no answer is given (and Cruse and Papineau do not provide one) then there is surely no reason to suppose that they are special, and it might reasonably be concluded that all terms can be directly understood.

Hempel's (1966) view of the relation between old terms and new terms is rather different: he does not claim that new terms must be definable using old terms, rather, he is willing to accept that new terms (i.e. terms introduced by our most recent theories) can, in general, be directly understood. It is not clear that this old term/new term distinction is sufficiently sharp: which theories count as our most recent theories? Moreover, terms are often used with somewhat different meanings in new theories (e.g. "mass" in relativistic mechanics) and in such cases it is not clear whether or not the term should count as old or new. Finally, and most seriously, it is not at all clear what the philosophical significance of the fact that a term was introduced by our most recent theories is even supposed to be. It is true that there might be good reasons to be more sceptical with regard to our most recent theories than with regard to older theories, because they're not (yet) so well tested, but hard to see why this scepticism should manifest itself as a belief in the Ramsey-sentences of our most recent theories but not the unRamseyfied versions of these theories: it would more naturally manifest itself as a reduced degree of belief in both the Ramseyfied and unRamseyfied versions of these theories.[8]

[8] It should be noted that Hempel does not put forward the old term/new term distinction in relation to Ramseyfication and that the problems that have been mooted here may arise only

2.2. The Munich Structuralists

The Munich structuralists have introduced a distinction between non-theoretical terms and theoretical terms that is somewhat similar to the distinction Lewis and Hempel make between old terms and new terms. They define a concept to be theoretical, relative to a given theory, if and only if the determination of the concept presupposes the laws of the theory (cf. Balzer et al, 1987, p. 55). They define a concept to be non-theoretical otherwise. So, for example, force is theoretical relative to classical particle mechanics, because "Every experimental device to measure force presupposes, by its very construction, the validity not only of Newton's [second] fundamental law but also of some more special mechanical law as well (e.g. the law of gravitation or some law of elasticity)" (Balzer et al, 1987, p. 53). Could their non-theoretical/theoretical distinction be used for the WESRist's purposes?

The first thing to note is that, according to this definition, it seems that it is (at least in principle) possible for a term to be theoretical relative to one theory in which it is used but non-theoretical relative to another theory in which it is used. Balzer et al themselves note this (1987, p. 55) and suggest that this possibility is in fact realised: they claim that while mass is theoretical relative to classical mechanics it is non-theoretical relative to stoichiometry (because determination of mass – allegedly – does presuppose the laws of classical mechanics whereas it does not presuppose the laws of stoichiometry) (stoichiometry is the study of the relationships between quantities of reactants and products in chemical reactions).

This suggests that it would not in general be appropriate to construct the Ramsey-sentence of a theory by Ramseyfying away the terms that are theoretical relative to that theory, because, following that procedure, whether or not one had to Ramseyfy away a given term would depend on the theory with which one was dealing, but it is difficult to see why this

---

in this connection: Hempel's distinction may be suitable for the purposes to which he puts it, but it cannot be used by the WESRist to justify Ramseyfication.

should be so. For example, why must "mass" be Ramseyfied in classical particle mechanics, but not stoichiometry? If it is acceptable to take it as a primitive non-logical term in one theory, why not in others?[9]

Another suggestion would be that no terms need to be Ramseyfied except those that are theoretical relative to our most recent theories. This proposal is very similar to the one based on Hempel's old term/new term distinction, which was discussed in the previous subsection. Like that proposal it seems to suffer from lack of motivation: it is just not clear why terms that are theoretical relative to our most recent theories need to be Ramseyfied while terms that are theoretical relative to older theories do not.

3. The Crude Fact/Scientific Fact Distinction

Poincaré (1905, especially part 3, chapter 10) makes a distinction between "crude" and "scientific" facts. (Duhem [1906, especially part 2, chapter 4] makes a similar distinction between "practical" and "theoretical" facts.) Crude facts are essentially theory free, whereas scientific facts are arrived at via inferences from crude facts together with a number of theoretical assumptions. For example, "the needle points to '10'" would qualify as a crude fact (if it was true) whereas "the voltage is 10V" would qualify as a scientific fact (if it was true).

The first obstacle facing the WESRist who wishes to make use of this distinction is that he needs a distinction between two types of term, not a distinction between two types of fact. A natural way to extract a distinction between crude and scientific terms from Poincaré's distinction between crude and scientific facts would be to define a term as scientific if and only if all the facts in which it occurs are scientific facts. At first sight it seems that terms like "voltage" could not be involved in crude facts, because

[9] Again, it should be noted that the Munich structuralists do not put forward their non-theoretical term/theoretical term distinction in relation to Ramseyfication and that the problems that have been mooted here may arise only in this connection: their distinction may be suitable for the purposes to which they put it, but it cannot be used by the WESRist to justify Ramseyfication.

it seems that to judge whether or not a claim about voltage was true would necessarily involve some theoretical assumptions. (Whereas it does seem that terms like "needle" could be involved in scientific facts, e.g. "the needle indicates a voltage of 10V".) However, this is not true. It is possible to judge that the claim "the voltage is 10V or it is not the case that the voltage is 10V" is true without any theoretical assumptions (it is just an instance of the law of excluded middle). This means that, according to the definition, "the voltage is 10V or it is not the case that the voltage is 10V" is a crude fact (since one doesn't need any physics to judge that it is true). Hence, according to the proposed definition, "voltage" is not a scientific term. Indeed, it is clear that the proposed definition would classify no terms as scientific. However, it is easy enough to modify the definition to avoid this objection. A term can be defined as scientific if and only if all the *contingent* facts in which it occurs are scientific facts.

However, the second obstacle facing the WESRist who wishes to make use of this distinction is not so easily overcome. The (modified) proposed definition will only draw a sharp and principled distinction between crude and scientific terms (i.e. a distinction of the sort that the WESRist needs) if the distinction between crude and scientific facts is itself sharp and principled, and it is certainly not obvious that this is so.

Worrall (1991) notes that, in a sense, all facts (about the external world) are scientific:

> Are we not, in making any assertion about the world, no matter how "empirical", really making certain assumptions, so that even the crudest of crude facts are assumption or theory-impregnated?
>
> At any rate, if our factual statements remain objective – about the "external world" rather than our present sensations – then the answer...is obviously "yes": even in reporting that the end of a certain pointer coincided roughly with the mark "10" on some scale we are assuming that the pointer really exists, that we are not constantly

> hallucinating, that a malicious demon is not misleading us, and so on. (Worrall, 1991, p. 334)

If all (contingent) facts (about the external world) are scientific then (according to the proposed definition of scientific terms) all terms (that refer to the external world) are scientific. That implies that there is a sharp crude term/scientific term distinction, but it is not appropriate for the WESRist's purposes (because it makes all terms that refer to the external world scientific and hence collapses the WESR into SESR).

However, Worrall goes on to point out that although there is a sense in which all facts (about the external world) are scientific it is nonetheless the case that in practice there is always a level of facts that can be taken as crude, because nearly everyone agrees that they are true. Taking this approach, the distinction between crude and scientific facts is not a distinction between facts that are theory free and facts that are not theory free but a distinction between facts that are widely agreed on and facts that are not so widely agreed on. However, this distinction is not sharp: there is no obvious degree of agreement that is the smallest degree of "wide" agreement (unless perhaps one takes "wide agreement" to mean "universal agreement", but on that account it seems that there would be no crude facts about the external world after all, since there is always a Cartesian who is prepared to dispute any alleged fact about the external world). This means that the corresponding distinction between crude and scientific terms would also not be sharp, and thus would again not be suitable for the WESRist's purposes.[10]

4. Chapter Summary

[10] Note that, in this form, the crude fact/scientific fact distinction becomes rather like Quine's (1960, especially pp. 40-46) observation sentence/non-observation sentence distinction. Roughly speaking, Quine defines an observation sentence as one whose truth-value most speakers of the language would agree on, if they were exposed to the same stimuli for a given period of time. Quine explicitly notes that his distinction between observation sentences and non-observation sentences is not a sharp one, and indeed he does not require a sharp distinction for his purposes.

It has been argued that neither the traditional observational term/theoretical term distinction (as founded on the observable/unobservable distinction), nor the old term/new term distinction, nor a distinction based on Poincaré's crude fact/scientific fact distinction, is suitable for the WESRists' purposes. The difficulty of drawing a sharp and principled bifurcation of the vocabulary that is suitable for the WESRists' purposes is a serious problem for the WESRist. Indeed, WESR appears to be bankrupt unless some such distinction can be made.

Chapter 3

Strong Epistemic Structural Realism

As explained in the introduction to this thesis, the difference between WESR and SESR is that while the WESRist thinks that theoretical terms (i.e. terms referring to unobservable objects, properties and relations) need to be Ramseyfied (leaving Ramsey-sentences containing only logical and observational terms) the SESRist thinks that external terms (i.e. terms referring to external objects, properties and relations) need to be Ramseyfied (leaving Ramsey-sentences containing only logical and internal terms).

In this chapter the forms of SESR espoused by Russell, Maxwell and Zahar are outlined and criticised. A discussion of arguably the most devastating criticism of SESR, Newman's objection (Newman, 1928), is, however, postponed for separate consideration in the next chapter.

Russell's (1912, 1927) argument for his structuralist thesis rests on his so-called "principle of acquaintance". I call it the "semantic argument" for SESR. Russell also attempts to argue for the structural realist's realist thesis, via the so-called "Helmholtz-Weyl principle" and "mirroring relations principle". I call this the "epistemic argument" for SESR. These arguments are the subject of sections 1 and 2 of this chapter. One could also envisage adapting Russell's semantic argument to support the WESRist's structuralist thesis. This possibility is discussed in section 3.

Maxwell puts forward what I call the "scientific argument" for SESR. This is discussed in section 4.

The most notable contemporary SESRist is Zahar (1996, 2001, 2004). His most significant contribution is his attempt to place SESR in an adverbialist indirect realist framework, as opposed to the sense-data indirect realist framework used by Russell and, less explicitly, Maxwell. One can extrapolate from what Zahar says an argument for his adverbialist version of the SESRist's structuralist thesis that is closely linked to Russell's semantic argument. Zahar's approach to SESR is discussed in section 5.

1. Russell's Semantic Argument for SESR

Russell puts forward his version of SESR most explicitly in *The Problems of Philosophy* (especially chapters 3 and 5) (Russell, 1912) and *The Analysis of Matter* (especially chapter 24) (Russell, 1927).

As noted in the introduction to this thesis, all forms of SESR presuppose indirect realism. Russell's arguments presuppose a form of indirect realism that would nowadays be called sense-data indirect realism. Sense-data indirect realism is to be contrasted with adverbialist indirect realism (as espoused by, for example, Ducasse, 1942 and Chisholm, 1957). The sense-data indirect realist asserts that what we directly perceive are not external objects but internal objects (sense-data). The adverbialist, on the other hand, takes an experience that may be described as "the sensing of a table" (for example) to involve not two objects (a sensor and a sensed object) that stand in a certain binary relation (sensing) but a single object (a sensor) that has a certain monadic property ("sensing tablely"). On both accounts an external object (e.g. a table) might be the cause of the sensing, but it also might not. (However, in so far as both accounts are forms of realism, both are committed to the claim that we have good reason to hold that an external table is at least usually the cause of an apparent sensing of a table.)

Russell's arguments will not be criticised for relying on the sense-data approach to indirect realism. This is because it seems likely that Russell's arguments could be translated without loss into the adverbialist indirect realist framework (this possibility is considered more closely section 5). Although the sense-data indirect realist disagrees with the adverbialist indirect realist on ontological matters (with regard to the existence of sense-data), the two positions are essentially in agreement on epistemological matters and Russell's arguments rely ultimately only on the epistemological claims of the sense-data indirect realist framework.

1.1. Russell's Principle of Acquaintance

Russell's semantic argument for SESR is founded on, (i) his theory of reference and meaning, and, (ii) his sense-data indirect realist epistemology. With regard to the former, Russell's view (although he never quite puts it this way) is that there are two types of term: primitive terms and non-primitive terms. There are correspondingly two routes of reference: acquaintance and description. A speaker can use a primitive term to refer only if he is acquainted with the object, property or relation of reference. Non-primitive terms should be taken as abbreviations for definite descriptions, formed, ultimately, using only primitive (and logical) terms.[1] A speaker can use a non-primitive term to refer only if he can use all the terms involved in the underlying definite description to refer: the term refers to the unique object, property or relation that satisfies the description (if there is one: if there isn't the term does not refer). This can be summarised as follows:

[1] If x can use a primitive term to refer then x is acquainted with the referent.
[2] If x can use a non-primitive term to refer then x can use all the terms involved in the underlying definite description to refer: the term refers to whatever satisfies the description.

Russell takes it that meaning and reference are intimately connected. His view (although, again, he never quite puts it this way) is that:

[1] A definite description (of an individual) behaves syntactically like a name and refers to an individual, but contains a formula. For example, if just one object has both the property P and the property Q, then we can form a definite description referring to this object from the formula Px & Qx, i.e. ix(Px & Qx), which is read, "the (object) x such that x is P and x is Q" (see Fitting and Mendelsohn, 1998, chapter 12, for a formal introduction to definite descriptions of individuals). In second-order logic one can form definite descriptions of properties and relations. For example, if there is just one property that two objects, a and b, share, then we can form a definite description referring to this property from the formula Xa & Xb, i.e. iX(Xa & Xb), which is read "the (property) X such that a is X and b is X".

[1] If x understands a primitive term (i.e. if the term is meaningful to x) then x is acquainted with the referent of the term.
[2] If x understands a non-primitive term then x understands all the terms involved in the underlying definite description.
[3] If x understands a sentence then x understands the non-logical terms in the sentence.

Maxwell (1970a, p. 181 and 1970b, p. 15) calls this Russell's "principle of acquaintance". Russell's own formulation of the principle is as follows: "*Every proposition which we can understand must be composed wholly of constituents with which we are acquainted*" (Russell, 1912, p.32, original emphasis). (This is not meant to imply that we can't refer to things with which we are not acquainted, but only that if we do so, we must do so via descriptions, where the direct referents of the terms in these descriptions are things with which we are acquainted.) The formulation given here is closer to Maxwell's: "All the descriptive (non-logical) terms in any meaningful sentence refer to items with which we are acquainted" (Maxwell, 1970a, p. 181).

Note that parts [1], [2] and [3] of the principle of acquaintance provide necessary but not sufficient conditions for x to understand primitive terms, non-primitive terms and sentences, respectively. (Likewise, Russell's and Maxwell's formulations of the principle also state necessary but not sufficient conditions for a sentence to be meaningful.) The principle is thus a partial, not a complete, theory of meaning. However, this is all Russell needs for his semantic argument for SESR. So it is no objection to the principle of acquaintance (at least, not in this context) that it does not provide both necessary and sufficient conditions for meaningfulness. Plausibly, if we wanted to elaborate the principle into a full theory of meaning, what we would need to add to [1] is the claim that x also knows that the relevant thing with which he is acquainted is the referent of the primitive term in question and what we would need to add to [2] is the claim that x knows that the non-primitive term in question is an abbreviation of the relevant definite description. However, as noted, these additions are

superfluous in this context, and, moreover, Russell does not himself suggest them.

It may be wondered on what basis Russell asserts the principle of acquaintance. In fact, he seems to take it to be almost self-evident. He claims that, "The chief reason for supposing the principle true is that it seems scarcely possible to believe that we can make a judgement or entertain a supposition without knowing what it is that we are judging or supposing about." (Russell, 1911, p. 209). It is not obvious how this is supposed to support the principle of acquaintance. If we read it as "The chief reason for supposing the principle true is that it seems scarcely possible to believe that we can make a judgement or entertain a supposition without *being acquainted with the referents of the primitive terms that we use to make the judgement or supposition*" then it does support the principle of acquaintance. But with this reading it seems more than "scarcely possible" that the claim Russell rejects is true: it seems we can entertain the supposition that "Caesar was bald" without being acquainted with Caesar and Caesar appears to be a primitive term in this supposition. Russell would say that "Caesar" is a non-primitive name, which explains how we can entertain suppositions about Caesar even though we are not acquainted with him, but as Russell's only argument for the principle rests on an appeal to intuitions about what seems possible, it is a serious worry that these intuitions do not seem to support the principle.

Moreover, the claim that a primitive term is meaningful to someone only if he is acquainted with the referent is highly problematic when we consider primitive terms referring to properties and relations (assuming that there are such terms). What can it even mean to be acquainted with a property or relation (rather than just instances of the property or relation)? Russell was prepared to bite the Platonic bullet and thought that we really do become "acquainted" with properties and relations (and not just their instances). He claimed that, "Not only are we aware of particular yellows, but if we have seen a sufficient number of yellows and have sufficient intelligence, we are aware of the universal *yellow*" (Russell, 1911, p. 203, original emphasis) and again that, "In addition to our acquaintance with particular existing things, we also have acquaintance with

what we shall call, *universals*...such as *whiteness, diversity, brotherhood,* and so on" (Russell, 1912, p. 28, original emphasis) but few people would be prepared to accept that these days (and if they were prepared to accept it, they ought to explain how and in what sense we can be acquainted with a universal).

In fact, we can get round this problem by modifying the principle of acquaintance, making the condition for the understanding of a predicate not acquaintance with the universal to which it refers but only acquaintance with an instance of the universal to which it refers. That is, we can make do with the following version of the principle of acquaintance:

[1] If x understands a primitive name then x is acquainted with the object to which the name refers.
[2] If x understands a primitive predicate then x is acquainted with at least one of the objects (pairs, triples, etc.) to which it applies.
[3] If x understands a non-primitive term then x understands all the terms involved in the underlying definite description.
[4] If x understands a sentence then x understands the non-logical terms in the sentence.

This version of the principle of acquaintance does not imply that we must be acquainted with universals in order to understand universal terms, and, as we shall see shortly, it is still strong enough for Russell's argument for SESR to go through. The claim here is only that a necessary (but not sufficient) condition for one to understand a universal term is that one must be acquainted with at least one object (or pair or triple etc. of objects) to which it applies. On the face of it this has some plausiblity: it does seem reasonable to suppose that the word "yellow" (for example) could not be meaningful to someone unless he was acquainted with at least one yellow object (it will be argued that this is in fact false, but at least it seems plausible at first sight). Henceforth in this thesis the phrase "acquaintance with a universal" is used to mean acquaintance with an *instance* of a universal.

However, note that it is not plausible to say that this version of the principle of acquaintance could be modified to provide necessary and sufficient conditions for x to understand all primitive terms. What could be added to [2] to provide a sufficient condition for x to understand a primitive predicate? Intuitively, what we require to make the condition sufficient is not just that the speaker is acquainted with a yellow thing but also that he can pick out the yellowness of the thing as the property to which "yellow" is supposed to refer. However, (part of) the point of Wittgenstein's private language argument (Wittgenstein, 1953, section 243) is that we beg the question if we suppose that the speaker can pick out yellowness in this way. Suppose a speaker sees a yellow object and says to himself "I shall call things like that 'yellow' ". If this is all he says to himself – if he does not further specify the way in which things must be like the object in question in order for them to be called "yellow" – then he will call everything "yellow", because everything is similar to the object in question in some respect. On the other hand, if he can tell himself what the relevant respect is (with a statement of the form "I shall call things like that in respect X 'yellow' ") then (since the relevant respect is yellowness) he can already refer to yellow (with the term X) in which case this explanation of how someone can refer to yellow with the word "yellow" presupposes that he can already refer to yellow with some other word, and is in this sense question begging. However, although these problems are serious for someone who wants to elaborate the principle of acquaintance into a full theory of meaning they do not affect Russell's semantic argument for SESR.

1.2. Russell's Indirect Realism

Since, as noted, Russell presupposes sense-data indirect realism his view is that we are not acquainted with external objects, properties and relations but only sense-data objects, properties and relations:[2]

[2] Actually this isn't quite true: Russell does not think we are *only* acquainted with sense-data objects, properties and relations. He claims that we are also acquainted "by *memory*" (Russell, 1912, p. 26, original emphasis) (presumably with memories) and "by

> we have acquaintance with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths. Thus in the presence of my table I am acquainted with the sense-data that make up the appearance of my table – its colour, shape, hardness, smoothness etc...My knowledge of the table as a physical object, on the contrary, is not direct knowledge. (Russell, 1912, pp. 25-26)

Russell's claim that we are acquainted with particular sense-data seems relatively straightforward, but the claim that we are acquainted with properties and relations of sense-data might mean that either, (i) the properties and relations with which we are acquainted are *only* properties and relations of sense-data, or that, (ii) the properties and relations with which we are acquainted are *at least* properties and relations of sense-data but might *also* be properties and relations of external objects. Russell's view (at least, prior to Newman's objection: see chapter 4) is that there are no "cross category" properties or relations (or at least we have no reason to think that there are any) so the universals with which we are acquainted do not pertain to external objects but are solely properties and relations of sense-data. Russell puts this view forward in a letter to A.J. Ayer (and elsewhere):

> You say that from the fact that the perceived qualities of physical objects are causally dependent upon the state of the percipient, it does not follow that the object does not really have them. This, of course, is true. What does follow is that there is no reason to suppose that it has them. From

---

introspection" (Russell, 1912, p. 26, original emphasis), which gives us awareness of our own awareness (acquaintance with our being acquainted) and that we are probably acquainted with ourselves (Russell, 1912, pp. 27-28). However, these exceptions are not important for present purposes.

> the fact that when I wear blue spectacles, things look blue, it does not follow that they are not blue, but it does follow that I have no reason to suppose they are blue. (Russell, 1969, p. 179)

Russell's point is that even if one was, in fact, acquainted with some property of (or relation between) external objects one would not know this to be the case, because one would not be acquainted with it as a property of (or relation between) external objects, but only as a property of (or relation between) sense-data. Given this, the working assumption should be that one is not acquainted with properties of (or relations between) external objects.[3] In fact, it seems to be a category mistake even to say that an external property or relation *could* be like a sense-datum property or relation. The two are so radically different that the comparison seems to make no more sense than the claim that the number eight is like a spray of salt water on the sea front.

1.3. Russell's Argument for his Structuralist Thesis

Given this background, we can now consider Russell's argument for his structuralist thesis. Consider any statement that we know to be (approximately) true. Presumably, such a statement must be meaningful to

[3] This seems to be Russell's view, most of the time prior to Newman's objection, but he does at one point back away from it to a degree, arguing that although we do not know that the properties and relations in the two realms are similar there could be grounds to suppose that they are: "we know nothing of the intrinsic quality of the physical world, and therefore do not know whether it is, or is not, very different from that of percepts...If there is any intellectual difficulty in supposing that the physical world is intrinsically quite unlike that of percepts, that is a reason for supposing that there is not this complete unlikeness" (Russell, 1927, p. 264). An intellectual difficulty (of the sort intended) in supposing that something is true may provide *some* reason for supposing it is not, but not necessarily a very compelling reason. Most people find that there is intellectual difficulty in supposing that there are more than three spatial dimensions, or that there are the same number of even numbers as natural numbers, but that should not stop us believing these things.

us. So, by the principle of acquaintance, all the non-logical terms in the statement are meaningful to us. There are only two types of non-logical term: primitive terms and non-primitive terms. By the principle of acquaintance, a primitive term is meaningful only if it refers to an item of acquaintance. Consequently, given Russell's indirect realism, a primitive name must name an internal object. Moreover, at least some of the objects (pairs, triple, etc.) to which primitive predicates apply must be internal. But then (given that Russell thinks that there are no cross-category predicates) these predicates must apply only to internal objects (pairs, triples, etc.). Hence they must be internal predicates. Hence, the primitive terms are all internal terms. But a non-primitive term is meaningful only if it is an abbreviation of a description formed using ultimately only primitive terms. Hence, in its unabbreviated form, the statement contains only primitive terms. Hence, in its unabbreviated form, the statement contains only logical and internal terms. Hence any statement we know to be (approximately) true is constructed using only logical and internal terms. This implies that:

The knowledge provided to us by our mature scientific theories consists (in its ultimate form) of statements constructed using only logical and internal terms.

This was referred to as a corollary of the SESRist's structuralist thesis in the introduction to this thesis. But, as noted there, Russell was writing before the Ramsey-sentence approach to ESR was developed, so we may take it that, for Russell at least, this "corollary" *is* the structuralist thesis.

Note that, even if this argument works, it does not establish SESR, because it does nothing to support the structural realist's realist claim: it does not follow from the alleged fact that if we have knowledge of the external world then it consists of claims constructed using only logical and internal terms that we actually do have such knowledge. What the argument does suggest (if it is sound) is that something like SESR is probably the strongest form of realism that could reasonably be upheld.

The principle of acquaintance is clearly a crucial part of Russell's argument (the other crucial part being his indirect realism). In subsections 1.4 and 1.5 objections to the principle of acquaintance are put forward.

1.4. Kripkean Objections to the Principle of Acquaintance

As noted, Russell's principle of acquaintance (i.e. his [partial] theory of meaning) is intimately connected to his (partial) theory of reference. Kripke (1981, especially pp. 80-91) has put forward a number of objections to this "description theory of reference". It may be thought that even if these undermined Russell's theory of reference that would not undermine the principle of acquaintance, because although Russell thought that meaning and reference were intimately connected, there is no reason why one could not combine the partial theory of meaning embodied in the principle of acquaintance with a non-Russellian theory of reference, e.g. the causal theory. However, because Russell's theories of reference and meaning are so intimately connected some of Kripke's arguments against the description theory of reference can easily be turned into arguments against the principle of acquaintance.

The principle of acquaintance implies that if we understand the (apparent) name "Julius Caesar" then we must be taking it as an abbreviation for a definite description. One possible objection is that the apparent name is meaningful to many people (children, for example) who know no more about Julius Caesar than that he was a Roman emperor, or some equally *in*definite description, and – to forestall a possible attempted counter – it is quite possible that some of them have forgotten how they came by the name. This objection, however, is not very compelling, because it does seem fairly plausible to say that in such cases people who use the term only partially understand what it means.

A second possible objection is that the principle of acquaintance seems to imply that a number of statements that appear to be synthetic are actually analytic. For example, if the definite description associated with

"Julius Caesar" is "the Roman Emperor from 49 B.C. to 44 B.C." then it is analytically true that Julius Caesar was the Roman Emperor in 49 B.C.[4] This, however, appears to be a synthetic truth. However, the proponent of the principle of acquaintance can perhaps bite the bullet here and argue that more statements are analytic than we normally think because many analytic truths (like many mathematical truths) are (although a priori true) not obviously true.

A third possible objection in this vein is based on the observation that, even in cases where we think that we can substitute a name for a definite description, we may, in fact, be wrong. For example, we might think that the name "Gödel" can be substituted by the definite description "the man who discovered the incompleteness of arithmetic"[5]. Although this appears to be a definite description it might not be: perhaps the incompleteness of arithmetic was in fact discovered by a number of people simultaneously but for some reason Gödel ended up with all the credit. Or perhaps, as Kripke suggests, "the proof simply materialised by a random scattering of atoms on a piece of paper – the man Gödel being lucky enough to have been present when this improbable event occurred" (Kripke, 1981, p. 86). Then, after all, "the man who discovered the incompleteness of arithmetic" is not a definite description, so the principle of acquaintance implies that the name "Gödel" is not meaningful to us, even though we think we can give a definite description as substitute for it. The problem is that the principle of acquaintance implies that the name "Gödel" is meaningful to us, if, in fact, "the man who discovered the incompleteness of arithmetic" *is* a definite description (and we believe this to be the case) but not meaningful to us if, in fact, it is not. Yet whether or not the name is *meaningful* to us does not seem to depend on this sort of contingency (although whether or not the name *refers* might do). In fact, if it did depend on this sort of

[4] This description itself uses predicates with which we are not acquainted. Consequently, before it could qualify as a legitimate Russellian description these predicates would themselves need replacing.

[5] See footnote 4.

contingency then it is doubtful that we would ever know that any non-logical term was meaningful, a deeply counter-intuitive result.[6]

Although the third of these objections has some force, none of them is wholly compelling, because they all rest on our intuitions about meaningfulness, and intuitions on this matter are easily deformed by a pet theory. An objection that does not have this defect is put forward in the next subsection.

1.5. A Formal Objection to the Principle of Acquaintance

As noted, according to Russell, we lack acquaintance not only with long dead Roman emperors, but with all external objects, properties and relations. As most everyday terms, such as "table" and "red" refer to external objects, properties and relations and we obviously understand these terms Russell must treat them as abbreviations for definite descriptions. Russell explicitly commits himself to this view with regard to "table", but he is less clear that this approach should be adopted for colour words, which he sometimes seems to treat as referring to properties of sense-data. However, if one held that "table" was an abbreviation for a definite description referring to an external object but that "red" referred to a property of sense-data then the

[6] It has been suggested to me that this problem can be evaded if we take the appropriate description to be "the man who is most widely believed to have discovered the incompleteness of arithmetic". But this evasion probably does not work. Suppose we carry out a survey asking everyone which man they think discovered the incompleteness of arithmetic, in order to determine the reference of the expression "the man who is most widely believed to have discovered the incompleteness of arithmetic". Most people who had an opinion would probably say "Gödel". This wouldn't help us pin down the reference of the expression, since the reference of "Gödel" is supposed to be determined by the reference of the expression. Unless most people who have an opinion have an independent means of picking out the man they believe to have discovered the incompleteness of arithmetic (e.g. by pointing to him) then it seems that, on this account, the referent of the expression "the man who is most widely believed to have discovered the incompleteness of arithmetic" will never be fixed, and will thus not be an appropriate definite description to associate with "Gödel". The problem is that it seems unlikely that most people who have an opinion do have an independent means of picking out this man (cf. Kripke, 1981, p. 89).

statement, "the table is red" would always be false (or at least we would never have any reason to think that it was true) as it would attribute a sense-data property to an external object. This seems implausible.

It is, therefore, hard to see how appropriate definite descriptions referring to external objects could be constructed: the only non-logical terms that can be used in the description are those that refer to objects, properties and relations with which we are acquainted, but if we are not acquainted with the referents of "table" and "red", with which words' referents are we acquainted? Actually, there's an easy way out of this problem, by taking (practically all) terms to be ambiguous. "Red" for example, would have two meanings and would ideally be replaced with two words. "Internal red" would refer to a property of sense-data, with which we are acquainted, whereas "external red" would refer to a property of external objects, with which we are not acquainted. In its latter sense it would need to be replaced with a definite description. Similarly for "table".

However, it is still not clear that it is possible to produce an appropriate definite description with which to replace even the simplest term that refers to an external object. Take "(external) table" for example. Russell suggests the appropriate definite description is, "the physical object which causes such-and-such sense-data" (Russell, 1912, p. 26), which is, formally, ix(Px & Cxa), where "Px" means x is physical, "Cxy" means "x causes y" and "a" names the sense-datum in question (presumably a sense-data table). This will obviously not do: it seems that neither "physical" nor "cause" should be meaningful according to the principle of acquaintance (unless they are themselves to be taken as abbreviations for appropriate definite descriptions: this approach will be discussed shortly). The word physical could be omitted from the above altogether (leaving us with the definite description "the object that causes such-and-such sense-data", i.e. ix[Cxa]) but the word "cause" seems indispensable. In this case "cause" is taken to be a relation between an external object on the one hand and a sense-datum on the other. Presumably we cannot be acquainted with this causal relation when we are only ever acquainted with one of its two terms anymore than we could grasp the "bigger than" relation by being told (truly) of a variety of objects that, "this is bigger than something".

If we are not acquainted with the "cause" relation then the definite description of a table as "the object that causes such-and-such sense-data", i.e. ix(Cxa), is at best a half-way house, but it might still be legitimate if "cause" could be replaced by an appropriate definite description.[7] One possible attempt would be to define cause as, "the relation holding between two objects such that the first is not a sense-datum and the second is", i.e. $iX(\forall x \forall y[Xxy \rightarrow (\neg Sx \ \& \ Sy)])$ where "Sx" means "x is a sense datum" (it seems plausible to suppose that we might be acquainted with such a property) but this would not do, because it is not a *definite* description of "cause" (assuming that there is not exactly one sense-datum and exactly one object that is not a sense-datum): if we have n non-sense-data objects and m sense-data objects then there are n.m pairs of the form <non-sense-data object, sense datum> and so $2^{n.m}$ subsets of the set of these pairs and so $2^{n.m} - 1$ relations between non-sense-data objects and sense-data objects (the "-1" here is because one of the subsets is the empty set, which is not a relation between non-sense-data objects and sense-data objects).

Perhaps it might be argued that "cause" refers to an unusual relation, one that sometimes holds between two sense-data and sometimes between an external object and a sense-datum. We could then be acquainted with the relation because it does sometimes hold between two sense-data (and being acquainted with a relation does not involve being acquainted with all its instances). We would then have no difficulty in understanding the term "cause" even when the particular instance of the relation we are talking about involves an external object and a sense-datum (rather as the conventional scientific realist might claim that having grasped the relation

[7] It has been suggested to me that there is a third option: Russell could argue that the term "cause" is meaningful, not because "cause" is an abbreviation for an appropriate definite description, nor because we are acquainted with causes, but because we are acquainted with our belief that "A causes B". However, I don't think this works. I don't think we can really believe that "A causes B" unless we know what the word cause means: it would be like believing that "A flubbles B". And, according to the principle of acquaintance, we can't know what "cause" means, unless we are acquainted with causes, or "cause" is an abbreviation for an appropriate definite description.

"bigger than" there is then no difficulty in understanding how it can hold between two objects that are both too small to be observed).

However, this seems a rather dubious approach: sense-data and external objects are supposedly radically different things (in a way that observable and unobservable objects are not, according to the conventional scientific realist). It therefore seems to be something like a category mistake to suppose that the very same relation could be instantiated either by two sense-data or by one external object and one sense-datum. It is hard to believe that a sense-data flame stands in *the same* causal relation to a external flame as it stands in to a sense-data striking of a match. Even if this was possible, it is doubtful that we could ever have reason to suppose that it was true, any more than we could have reason to suppose that external red was like sense-data red.[8]

In fact, assuming the Russellian does not adopt this dubious approach, then it can be formally proved that it is in general impossible for him to create a definite description of an external object. It can be shown that:

It is impossible to create a definite description of an external object using a language of second-order logic whose non-logical terms refer exclusively to internal objects (or properties exclusively thereof or relations exclusively amongst internal objects), if there is more than one external object.

This is a straightforward corollary of the following theorem:

Given a domain with two classes of objects, A and B, and a language of second-order logic whose non-logical terms refer exclusively to objects in class A (or sets of objects in class A, or sets of tuples of objects in class A), it is impossible to create a definite description of an object in class B, if B contains more than one object.

[8] However, Russell does adopt this line in response to Newman's objection: he claims that there is at least one predicate ("co-punctuality") that can apply to both internal and external objects. The difficulties with this proposal will be further elaborated in chapter 4.

which is proved in appendix 2. This result commits an advocate of the principle of acquaintance to the view that it is impossible to refer to a unique external object (as he is already committed to the view that we cannot directly refer to external objects). This appears to be a reductio of the principle, because phrases like "the nearest supernova" do appear to refer to unique external objects (at least potentially). As Russell's semantic argument for the SESRist's structuralist claim crucially depends on the principle of acquaintance, this result also undermines this argument. In section 2 we will turn to consider Russell's argument for the SESRist's realist claim.

2. Russell's Epistemic Argument for SESR

It has been argued that Russell's semantic argument for SESR is far from compelling, because it is based on the principle of acquaintance, a principle that seems to have absurd consequences. Votsis (2004, 2005) identifies two further principles that Russell deploys to further argue for SESR, principles that Votsis calls the "Helmholtz-Weyl principle" (following Psillos, 2001a) and the "mirroring relations principle". These principles form the basis of Russell's epistemic argument for SESR. While Russell's semantic argument is for his structuralist thesis (the thesis that our knowledge of the external world consists of claims constructed using only logical and internal terms) his epistemic argument is for structural realist's realist thesis (the claim that mature scientific theories provide us with substantial knowledge of the external world). (Cf. The way Worrall uses the pessimistic induction to argue for the WESRist's structuralist thesis and the no miracles argument to argue for the structural realist's realist thesis: see chapter 1.) There are two questions to ask about each of these principles: (i) Is it plausible? (ii) If it were true, would it secure substantial (conjectural) knowledge about the external world of the sort that, according to the SESRist, science provides?

2.1. The Helmholtz-Weyl Principle

Russell states (what has since been called) the Helmholtz-Weyl principle as follows: "differing percepts have differing stimuli" (Russell, 1927, p. 255).[9] The term stimuli might refer to a number of things. For example, the stimulus that causes a (visual) sense-data table might be taken to be an external table, a collection of photons, a brain state, or some combination of these. Russell uses the word "stimuli" to refer to, "events just outside the sense-organ" (Russell, 1927, p. 227). The term "differing" also might mean two separate things: do two objects differ simply by being two and not one, or is it also required that they are qualitatively different (i.e. have different properties)? The most coherent reading of Russell suggests that he intended differing in the latter sense. The Helmholtz-Weyl principle thus amounts to the following:

Qualitatively different percepts are caused by qualitatively different events just outside the sense-organ.

However, as stated, the principle is false. It is certainly possible that someone might be exposed to exactly similar events just outside the sense organ on two occasions but have completely different percepts on the two occasions – if, for example, he was blinded in between the two occasions. This suggests that the principle is false because of Russell's intended meaning of "stimuli" and that it might be true if we took the stimuli to be the total external state of affairs (including the state of the sense organs involved). This would yield the following modified form of the Helmholtz-Weyl principle:

Qualitatively different percepts are caused by qualitatively different external states.

[9] Cf. Votsis' statement of the principle: "Different effects (i.e. percepts) imply different causes (i.e. stimuli/physical objects [footnote omitted])" (Votsis, 2005, p. 1362).

Although not incontrovertible, this has at least some plausibility (except, perhaps, for the mention of "causation" it sounds like the familiar claim that the mental supervenes on the physical).

Does this modified form of the Helmholtz-Weyl principle enable one to infer substantial knowledge about the external world? The answer is no. Suppose one has two differing percepts, one of a sense-data-blue sense-data-table and one of a sense-data-red sense-data-table. The modified form of the Helmholtz-Weyl principle allows one to infer that they are caused by different external states. But it does not licence the inference to anything about these states, *except* that they are different. For all it says the first might be caused by a single external object having a single monadic property, whereas the second might be caused by an uncountably infinite collection of external objects standing in a complex web of relations. The modified form of the Helmholtz-Weyl principle maybe plausible, but it is nowhere near strong enough to give Russell what he wants. It does give us some knowledge of the external world – that there is at least one distinct external state of affairs for each distinct percept – but no indication of how (or if) these states of affairs are related, which is far less knowledge than the structural realist claims that we have.

2.2. The Mirroring Relations Principle

Russell states (what has since been called) the mirroring relations principle (MR) as follows: "the objective counterparts [of phenomena]...form a world having the same structure as the phenomenal world" (Russell, 1919, p. 61).[10, 11] For present purposes the following formulation of MR will be used:

[10] Cf. Votsis' statement of the principle: "Relations between percepts mirror (i.e. have the same logico-mathematical properties as) relations between their non-perceptual causes" (Votsis, 2005, p. 1362).

[11] This principle is substantially stronger than the Helmholtz-Weyl principle. Indeed, as we shall see, it makes the Helmholtz-Weyl principle altogether redundant, which makes one

The structure exemplified by the external world is isomorphic to the structure exemplified by the internal world.

Is MR plausible? Russell says little in its favour. Votsis claims that, "it is hard to imagine how we can have knowledge of the external world without accepting something like MR" (Votsis, 2005, p. 1366). If the only argument for MR is that it must be accepted if one is to have knowledge of the external world, then, if MR is used to support the claim that we do have substantial knowledge of the sort the structural realist claims that we have, then the argument, although not quite circular, is certainly elliptical enough not to convince the antirealist.[12] However, it is easy to see that even if MR were plausible it would not justify the claim that we do have substantial knowledge of the sort the structural realist claims.

What the Russellian ESRist would like to show using MR is that something like the following version of the structural realist's realist thesis is true:

The structure that our mature scientific theories ascribe to the external world is isomorphic to the structure of the external world.

Let's call this R. An initial worry with R is that it just doesn't make sense to talk about *the* structure that our mature scientific theories ascribe to the external world: different theories describe the same system in different ways. For example, a hydrogen molecule may be treated as a single object (a molecule) by one theory, as a pair of objects (two hydrogen atoms) by another, as a collection of four objects (two protons and two electrons) by a

---

wonder why, given that Russell held MR, he bothered to formulate the Helmholtz-Weyl principle.

[12] Perhaps though, converting the antirealist is not really Russell's intention. Perhaps he is just trying to be explicit about the principles he believes that we (in our commonsensical realist outlooks) actually do presuppose and trying to demonstrate how (if true) they would justify SESR.

third, as a collection of eight objects (six quarks and two electrons) by a fourth... These theories will be satisfied by different structures. So what is *the* structure that our mature scientific theories ascribe to the hydrogen molecule? Moreover, even the description provided by a single theory might be satisfied by more than one (non-isomorphic) structure. There is nothing profound about this. It's just the case that a set of sentences can (in general) be satisfied by more than one structure. Indeed, by the Löwenheim-Skolem theorems, if a set of first-order sentences has one infinite model (i.e. a model with an infinite domain) then it has models of every infinite cardinality. That is profound, but not so relevant. However, let's assume for the sake of argument (since there is worse to come) that these difficulties can be resolved: perhaps the structure that our mature scientific theories ascribe to the external world is the simplest structure (a tricky notion) that satisfies our most fundamental theory or the simplest structure that satisfies all our theories (in the unlikely event that there is one).

MR would justify R if and only if we make the following assumption (let's call it A):

The structure that our mature scientific theories ascribe to the external world is isomorphic to the structure of the internal world.

in the sense that:

$(MR \,\&\, A) \rightarrow R$ and $(MR \,\&\, \neg A) \rightarrow \neg R$

but (even though it is not really clear what *the* structure that our mature scientific theories ascribe to the external world is supposed to be) A is blatantly false: our theories describe the world to consist not only of tables and other macroscopic objects that have a counterpart in the internal world but also of many other objects (our own brains, photons, molecules, atoms...) that do not have internal counterparts. The point is that the structure that we (as adults in a scientifically sophisticated society) actually infer that the external world exemplifies is generally much more complex

than the structure that the internal world exemplifies. For example, faced with a sense-data human one can infer that there is not just one external object (corresponding to the single sense-data human) but many: numerous external cells, an even greater number of molecules.... Given that (MR & $\neg$A) $\rightarrow$ $\neg$R and A is false then:

MR $\rightarrow$ $\neg$R

so MR, far from justifying R, in fact implies that it is false. This is because it implies that that the most naive common-sense view possible is right.[13]

What this problem seems to suggest is that there is no simple mapping linking the structures exemplified by the internal and external worlds, if our best theories are anything like right. Scientific practice presumably does presuppose that there is some relation between the two, but not, in general, a simple one, and not one that can be specified in advance to cover all occasions. This really should not be surprising: if MR was true physics would be a very much easier subject than it in fact is.[14]

---

[13] Psillos (2001a) also objects to the Russellian contention that the structure of the internal world is isomorphic to the structure of the external world: "From 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." (Psillos, 2001a, p. S15). While Psillos claims that the realist requires that this possibility be open what is claimed here is something stronger: that the (scientific) realist believes that this possibility is in fact realised (since he believes that mature scientific theories are approximately true).

[14] It should also be noted that the Ramsey-sentence approach to ESR seems to be incompatible with MR, because, as Votsis notes, "the Ramsey-sentence of a theory preserves the *logical structure* of the whole theory, something directly at odds with Russell's insistence that we infer the structure of the world from the structure of our perceptions" (Votsis, 2004, pp. 55-56, original emphasis). The two are at odds because the structure of a theory is typically much richer than the structure of perceptions, so MR could not justify (and would in fact undermine) belief in a theory's Ramsey-sentence. This will not worry the advocate of the Ramsey-sentence approach, because, as noted, there appears to be a tension between MR and any form of scientific realism. Perhaps unsurprisingly, most contemporary structural realists are not explicitly committed to MR (Votsis [2003, 2004, 2005] appears to be the only exception).

3. Russell's Semantic Argument Adapted for WESR

In principle, the WESRist could adapt Russell's semantic argument and use it to argue for his own structuralist thesis, i.e. he could argue that the only non-logical terms that are directly *meaningful* are those that refer to observables.[15] To this end, he would propose a variant of the principle of acquaintance, the conjunction of the following claims (henceforth, this will be called the principle of observation):

[1] If x understands a primitive name then x has observed the object to which the name refers.
[2] If x understands a primitive predicate then x has observed an instance of the property or relation to which it refers.
[3] If x understands a non-primitive term then x understands all the terms involved in the underlying definite description.
[4] If x understands a sentence then x understands the non-logical terms in the sentence.

The principle of observation is not equivalent to the principle of acquaintance as observing an object is not the same as being acquainted with it. Recall that observables are (roughly speaking) those external objects, properties and relations that have direct counterparts in internal experience (e.g. tables and blueness, but not quarks or strangeness). Observing such an object, property or relation would not involve being acquainted with it in Russell's sense (which, for the indirect realist, would be impossible) but only being acquainted with its internal counterpart.

Consider any statement that we know to be (approximately) true. Presumably, such a statement must be meaningful to us. So, by the principle

[15] In fact, although WESRists have not explicitly adopted this argument the logical positivists did put forward something like this view, without taking it to imply WESR (see, for example, Carnap, 1956).

of observation, all the non-logical terms in the statement are meaningful to us. There are only two types of non-logical term: primitive terms and non-primitive terms. By the principle of observation, a primitive term is meaningful only if it refers to an item that has been observed. Consequently a primitive term must refer to an observable object, property or relation. Hence, the primitive terms are all observational terms. But a non-primitive term is meaningful only if it is an abbreviation of a description formed using (ultimately) only primitive terms. Hence, in its unabbreviated form, the statement contains only primitive terms. Hence, in its unabbreviated form, the statement contains only logical and observational terms. Hence any statement we know to be (approximately) true is constructed using only logical and observational terms. This implies that:

The knowledge provided by our mature scientific theories consists (in its ultimate form) of statements constructed using only logical and observational terms.

This was referred to as a corollary of the WESRist's structuralist thesis in the introduction to this thesis, but if the argument established the corollary, it would make the thesis highly plausible: it is hard to see how our knowledge could go beyond the Ramsey-sentences of our best theories, if our knowledge must consist of claims constructed using only logical and observational terms.

Again, however, note that even if this argument works, it does not establish WESR, because it does nothing to support the structural realist's realist thesis: it does not follow from the alleged fact that if we have knowledge of the (observable and unobservable) external world then it consists of claims constructed using only logical and observational terms that we actually do have such knowledge. What the argument does suggest (if it is sound) is that something like WESR is probably the strongest form of realism that could reasonably be upheld.

The principle of observation is somewhat more tenable than the principle of acquaintance. It is true that the arguments based on intuitions about meaningfulness directed against the principle of acquaintance seem to

have equal force in the present case (these arguments will not be repeated here). However, the final objection to the principle of acquaintance, which seemed to be the most powerful, is perhaps not as compelling when applied to the principle of observation. This objection is that it is in general impossible to create a definite description referring to an external object using only logical terms and non-logical terms that refer to sense-data (or properties thereof or relations exclusively amongst sense-data) so the principle of acquaintance implies that it is impossible to refer to a unique external object. This is a corollary of the theorem proved in appendix 2. Another corollary of this theorem is the following:

It is impossible to create a definite description of an unobservable object using a language of second-order logic whose non-logical terms refer exclusively to observable objects (or properties exclusively thereof or relations exclusively amongst observable objects), if there is more than one unobservable object.

However, the significance of this theorem for the WESRist is debatable. If he takes an observational predicate to be one that refers exclusively to sets of (tuples of) observable objects then this result will commit the WESRist to the view that it is impossible to refer to a unique unobservable object. That would appear to be a reductio of WESR, as phrases like, "the nearest black hole" do appear to successfully refer to unique unobservable objects (at least potentially). However, if the WESRist believes that observational predicates can apply to unobservable objects (and this seems intuitively reasonable: it does seem that observational predicates such as "larger than" can be used to describe a relation between atoms and electrons) then the result is irrelevant, because the WESRist does not believe that we have a language whose non-logical (observational) terms refer exclusively to observable objects (or

properties exclusively thereof or relations exclusively amongst observable objects).[16]

This suggests that the semantic argument for ESR, even if not entirely unproblematic for the WESRist, works better as an argument for WESR than as an argument for SESR. On the other hand, it has already been argued (in chapter 2) that drawing the observational term/theoretical term distinction in a manner suitable for the WESRist's purposes is problematic.

Contrast this semantic argument for WESR with the historical argument (based on the pessimistic induction) discussed in chapter 1: note that these two arguments lead one to different opinions with regard to theoretical terms (i.e. terms referring to unobservable objects, properties and relations) and with respect to the relationship between the sentence, A, that expresses the theory and the Ramsey-sentence, A*, obtained from A. If one adopts the historical argument for WESR then one can accept that theoretical terms have the logical form of names and predicates (i.e. one can accept that theoretical terms are primitive terms) and so one can regard A and A* as different sentences, where only the latter is known to be true. On the other hand, if one adopts the semantic argument for WESR then theoretical terms cannot have the logical form of names and predicates and A will be regarded as merely an abbreviated way of writing A* (i.e. a theory's Ramsey-sentence will be regarded as the true logical form of the theory).

4. Maxwell's Scientific Argument for SESR

In a paper presented at a conference in 1965 Maxwell claims that, "the only aspects of the nonmental world of which we can have any knowledge or conception are purely structural (or, in other words, purely formal)" (Maxwell, 1968, p. 153). A purely formal claim is, presumably, one

[16] The SESRist cannot plausibly evade the analogous corollary of the theorem in analogous manner, because the internal and external realms are completely unalike and so it seems

constructed using only logical terms. Consequently Maxwell's claim is in this respect stronger than the structuralist thesis attributed to SESRists in this thesis (as noted, their position implies that our knowledge of the external world consists, in its ultimate form, of statements constructed using only logical *and internal* terms). However, Maxwell concedes (in a footnote presumably retrospectively added for the published version, which came out in 1968) that, "the account in the present essay is not only incomplete but contains a serious error: structure should not be identified with form; rather it is form plus causal connections with experience" (Maxwell, 1968, p. 154). He elaborates this point in later work, maintaining that our knowledge of external objects is at most knowledge of their "structural properties" but defining structural properties so that "not all structural properties are also purely formal" (Maxwell, 1970a, p. 188) thus allowing that we are not limited to purely formal knowledge of external objects.

Maxwell's proposal is that a "structural" description of an object, property or relation can make use of non-logical terms, as long as these refer to items of internal experience. In other words, Maxwell (after 1965) defines a "structural" claim to be one constructed using only logical and internal terms. This seems to be a confusing choice of terminology, but once we understand that this is what Maxwell means by "structural" (after 1965) his claim that knowledge of the world is "structural" is (after 1965) in line with the structuralist thesis attributed to SESRists here.

Maxwell's argument for this claim runs as follows:[17]

> If our current theories in physics, neurophysiology, and psychophysiology are at all close to the truth or even if they are at all headed in the right direction, then a complete description, including a complete causal account, of everything that is involved in perception except the private experience itself would mention only such entities and events as submicroscopic particles, electromagnetic quanta,

---

that there cannot be any predicates that apply to objects in both domains.

> etc., and their relations and interactions with one another and with, for example, neural termini in the retina, afferent neural impulses, and patterns of neuronal activity in the brain. At no point in the entire complete description and causal explanation is there any mention of any first order property such as colors *until* we come to the private experience that results from the pattern of neuron firings in the brain. [footnote omitted] It seems to me that we must conclude that colors are exemplified *only* in our private experiences and that there is no reason to believe that they are ever properties of the material objects of the external environment. *What holds for colors must be true for all of the first order properties that we perceive directly.* We do not know what any of the first order properties of material objects are, although our (Ramseyfied) theories tell us that they exist and what some of *their* (second and higher order) properties are. This ends the summary of the scientific argument for Russell's contention that our knowledge of the external (or *physical*, or nonmental) realm is limited to its structural aspects. (Maxwell, 1970b, p. 19, original emphasis)

Maxwell argues here that a complete description of the physical process that terminates in the perception of a colour would not make any reference to colours, but only to such entities and events as "submicroscopic particles" etc. and that we should infer from this that colours do not exist. However, there is something wrong with this reasoning, as is perhaps more apparent when we note that it could just as well (or rather badly) be argued that a complete description of the physical process that terminates in the perception of a cat would not make any reference to cats, but only such entities and events as "submicroscopic particles" etc. and that we should

[17] Maxwell offers no argument for the structural realist's realist thesis.

infer from this that cats do not exist. This is an affront not only to common sense (and cats) but also to a number of scientific (biological) theories that purport to deal with the behaviour, physiology, etc. of cats.

This absurdity arises because Maxwell (in effect) assumes that there is a single uniquely accurate description of the external world (given by our most fundamental theories). It is true that if we assume a very strong reductionist thesis then only the most fundamental theory is needed (in principle) in the sense that it should be able to account for all phenomena without recourse to any less fundamental theories. Such a thesis is of course not uncontroversial, but even if it is maintained it doesn't seem to make any difference: even if it is possible to account for all phenomena with a theory of the world that doesn't mention cats that surely would not imply that cats do not exist in the world, only that they are supervenient on other entities. The same applies to Maxwell's own example of colours. Colours do not appear in our most fundamental theories of the world. It would be fair to infer from this that objects' colours supervene on their other properties but it would be wrong to infer from this that objects are not coloured. As Quine notes in his comment on Maxwell's paper:

> Water remains water gallon by gallon, I say, even though its submicroscopic bits are rather oxygen and hydrogen; there is no paradox in this, and there is none in saying that a table remains smooth and brown, square inch by square inch, even though its submicroscopic bits are discrete, vibrant, and colorless. The quality of being aqueous, also of being smooth and brown, are like swarming, or waging war: they are traits only of a congeries. This does not make them unreal or subjective. There is no call for a predicate to hold of each part of the things it holds of. Even a predicate of shape, after all, would fail that test. It is a modern discovery in particular that aqueousness, smoothness, and brownness resemble squareness and swarming on this score; but it is not a contradiction. (Quine, 1968, p. 162)

However, although Maxwell's argument is flawed, his conclusion, that the objects, properties and relations that we perceive are not external objects, properties and relations follows immediately from indirect realism. For the indirect realist the colours that we experience in percepts are not the same as the colours of external objects. More generally, we must distinguish between external cats, colours, electric charges, spatial extensions and atoms and their internal counterparts (if they have any). According to the indirect realist we directly perceive the latter but not the former. Maxwell's argument from science for this conclusion is completely redundant if one accepts indirect realism, and, for the reasons discussed, cuts no ice if one does not. However, whilst (if we are indirect realists) it is reasonable to give Maxwell the lemma that we do not directly experience the properties or relations of external objects (or the objects themselves), his conclusion, that our knowledge of external objects is "structural" (in Maxwell's sense: i.e. consists of claims constructed using logical and internal terms) only follows if we further assume something like the principle of acquaintance: that any meaningful primitive non-logical term must refer to something that we directly experience. Maxwell does endorse this principle, but it has already been argued (in section 1) that it is false.

5. Zahar's Adverbialist Approach to SESR

5.1. Zahar's Version of SESR

Zahar (1996, 2001, 2004) advocates a version of SESR based on the Ramsey-sentence approach. However, certain comments he makes suggest that his view may be substantially different from those of other SESRists, such as Russell and Maxwell. The latter make a distinction between internal objects, properties and relations that can be referred to directly by primitive non-logical terms, and external objects, properties and relations that can only be referred to by descriptions (formed using logical and internal terms). Zahar, however, sometimes appears to draw a quite different distinction *in addition* to the distinction drawn by Russell and Maxwell, between our

knowledge of external objects and our knowledge of external relations (whether or not he intends this to include monadic relations, i.e. properties, is not clear). For example, he claims that, "in a unified and highly confirmed theory, the basic relations mirror some objective reality while the relata remain inaccessible to human knowledge" (Zahar, 2001, p. 38).[18]

Notice how Zahar here juxtaposes an ontological claim about relations (that they mirror some objective reality) with an epistemic claim about the relata (that they are inaccessible to human knowledge). If this is taken to imply that the relata do not mirror some objective reality (or that we cannot know this to be the case) then it is incompatible with the version of ESR that has been ascribed to Russell and Maxwell in the previous sections. According to Russell and Maxwell, each relatum in the theory mirrors a relatum in the external world every bit as much as the relations in the theory mirror the relations in the external world: the "mirroring" (in its ideal form, which occurs when the theory is exactly true) is a one-one mapping from the objects in the domain of the structure of the theory to the objects in the domain of the structure of the world (that preserves the relations), so each relatum in the theory has a counterpart in the external world, as well as each relation. Of course, according to the indirect realist, we don't have direct access to the external world, so we could never be *certain* that there was such a mapping (i.e. we could never be certain that the theory in question was true). Moreover, the structural realist does not maintain that our theories are ever exactly true, but only that they are approximately true, so he would not claim that the mapping would take exactly this form, only something that approached this form. However, neither of these considerations imply the sort of difference between relata and relations that Zahar seems to suggest.

In fact, it is hard to see in what sense the relations in a theory could mirror the external world without the relata doing likewise, because the relations just are sets of n-tuples of relata. As Zahar puts it, "classical semantics seems unable to interpret relations except through their relata"

[18] In actual fact, Zahar puts this forward as Poincaré's view, but it is fairly clear that he

(Zahar, 2001, p. 38). This leads Zahar to suggest that we need a new "structural semantics" to replace the normal semantics of predicate logic. However, if Zahar's position really requires a radical new semantics for predicate logic to become viable then, given that he doesn't even begin to suggest what such a new semantics might look like, that is surely so much the worse for his position. With regard to the second part of the above quotation it is true, according to Russell and Maxwell, that external objects remain, "inaccessible to human knowledge" in so far as we are not acquainted with them, but the same also holds for external relations (and neither is completely inaccessible: we have knowledge by description – knowledge consisting of claims formed using logical and internal terms – of both).

In attempting to further elucidate why it is that SESR implies relations are more important than relata Zahar makes the similar claim that a relation in a theory may reflect, "a real connection between elements about whose intrinsic nature we know next to nothing" (Zahar, 2001, p. 38). Notice how Zahar (again) juxtaposes an ontological claim about relations (that they are real) with an epistemic claim about the relata (that we don't know anything about their intrinsic natures). Both may be true, but the comparison does not seem to support the claim that the relations are (either ontologically or epistemically) more important than the relata. On the one hand, if a relation in a theory reflects "a real connection" why don't the relata reflect real objects? On the other hand, if we know nothing about the "intrinsic nature" of objects isn't the same true of relations? The second case is not so trivial: it depends what is meant by the intrinsic nature of an object. Such talk suggests that Zahar is chasing – or perhaps running away from – Lockean substances, but isn't the nature of an object exhausted by its properties and relations? If so then it is true – according to SESR – that our knowledge of the nature of external objects is limited (to knowledge consisting of claims formed using logical and internal terms) but again, the same holds with regard to the nature of external properties and relations.

---

endorses it.

It is true that, on any approach to SESR, there is a sense in which we have less knowledge of external objects than of external relations. According to SESR when we consider an external object or external relation *in itself* (i.e. when we ignore how it is related to other objects and relations, in particular internal objects and relations) we are restricted to knowledge of its formal properties. An object, considered in itself (i.e. ignoring its properties and relations) doesn't have any interesting formal properties, whereas a relation, considered in itself, can display a variety of formal properties: it may be transitive or symmetric or reflexive or Euclidean etc. Properties in this respect are more like objects than relations, there seems to be nothing (formal) to say about them, in themselves, except that they contain a certain number of elements. However, this can't be the point that Zahar was trying to make, for clearly it does not suggest the need for a new semantics for predicate logic. On the contrary, it is completely trivial: there is obviously nothing interesting (formal or otherwise) to say about an object, *if we ignore its properties and relations* (even if there is something more to its "nature").

5.2. Zahar's Argument for SESR

Like Worrall, Zahar employs the no miracles argument to argue for the structural realist's realist thesis. One can also reconstruct from his writings an argument for the SESRist's structuralist thesis that is closely linked to Russell's semantic argument, but has an important difference: Zahar rejects Russell's sense-data version of indirect realism in favour of an adverbialist indirect realism. Zahar's reasons for rejecting the former are traditional worries about the ontological status of sense-data. Sense-data are certainly supposed to have some very unusual properties. Firstly, they have only a minimal independence of the individuals that sense them: they presumably only exist when they are sensed, but what sort of object can be destroyed by closing one's eyes or (re?-)created by opening them? For that matter is it the same sense-data table with which I am acquainted before and after I blink? Furthermore, as Zahar (2001) has noted, sense-data would often be incomplete objects:

> A speckled hen is apprehended as having a finite number of speckles but neither as having 100 speckles nor as not having 100 speckles...the phenomenal hen will have a finite number of speckles but it is not the case that this number will be either equal to or different from 100. (Zahar, 2001, p. 28)

Adverbialists, such as Zahar, do not postulate sense-data at all. The original motivation for postulating sense-data is that apparently seeing a speckled hen (for example) does not (necessarily) involve seeing a speckled hen, and this can be explained by supposing that what we (directly) see is something else – a sense-data speckled hen – that may or may not be caused by a real speckled hen. But in fact there is no need to suppose that (apparently) seeing a speckled hen (ever) involves seeing a sense-data hen. Instead of construing the experience of apparently seeing a speckled hen as involving a two-place relation of "seeing" in which the viewer stands to another object (a sense-data speckled hen) it would be better (in so far as it would avoid the need for such troublesome objects as sense-data) to construe "apparently seeing a speckled hen" as a one-place relation in which the "viewer" alone stands, i.e. as a mode of being that an individual undergoes. (Of course, the common-sense view is that it is a two place relation in which the viewer stands to a real hen, but this is implausible because, as noted, there need be no real hen for the viewer to undergo an experience of apparently seeing a speckled hen.)

However, there is still a significant distinction to be drawn between sentences that make claims about the external world (such as "There is a table in Elie's room" or "I saw a table in Elie's room") and sentences that make claims only about one's own experiences (such as "I believe that there is a table in Elie's room" or "I had an experience like seeing a table in Elie's room"). Zahar (2001) calls the latter autopsychological sentences. It seems that making autopsychological claims is a relatively safe business: for any non-autopsychological claim, p, that I take to be true I am always less likely to be wrong if I assert (the

autopsychological) "I believe that p" rather than just (the non-autopsychological) "p". Similarly, whenever we have a non-autopsychological sentence of the form "I observed q" the autopsychological statement "I had an experience as of observing q" is less likely to be false (in fact in this case – but not in the earlier case – it seems that the former implies the latter, but not vice-versa).

However, this difference, in itself, does not argue for the SESRist's structuralist thesis, nor even for the corollary of this thesis: the fact that claims about the external world are less certain than claims about one's own experiences does not imply that our knowledge of the external world consists of claims constructed out of logical and internal terms. To argue for the latter Zahar would need, although does not explicitly put forward, something like the following variant of the principle of acquaintance:

[1] If x understands a primitive name then it names x himself (the only object with which x is acquainted, according to the adverbialist).
[2] If x understands a primitive predicate then it ascribes to its subject a psychological state that x has experienced (the only properties with which x is acquainted, according to the adverbialist).
[3] If x understands a non-primitive term then x understands all the terms involved in the underlying definite description.
[4] If x understands a sentence then x understands the non-logical terms in the sentence.

Given this version of the principle of acquaintance Zahar could argue for the corollary of the SESRist's structuralist thesis in essentially the same way as Russell but it seems that this version of the principle of acquaintance faces essentially the same problems as the original principle of acquaintance. In particular, the formal objection to Russell's principle of acquaintance is equally compelling when applied to this version of the principle of acquaintance. The objection was based on the following corollary of the theorem proved in appendix 2:

It is impossible to create a definite description of an external object using a language of second-order logic whose non-logical terms refer exclusively to internal objects (or properties exclusively thereof or relations exclusively amongst internal objects), if there is more than one external object.

This is relevant to the modified version of the principle of acquaintance because, according to this principle, we *are* restricted to a language whose non-logical terms refer exclusively to internal objects (or properties exclusively thereof or relations exclusively amongst internal objects). In fact, we are restricted to a language with only one name (denoting ourselves), a number of extensionally identical one-place predicates (the extension of each being the set containing ourselves) denoting the various psychological states that we have experienced and no other non-logical terms (so, in particular, no non-logical predicates denoting relations). So the modified version of the principle of acquaintance implies that it is impossible to create a definite description of a unique external object. This implies that it is impossible to refer to a unique external object (assuming that there is not only one external object). This appears to be a reductio of Zahar's approach to SESR, because phrases like "the nearest supernova" do appear to successfully refer to unique external objects (at least potentially).

6. Chapter Summary

The main claim of this chapter is that the principle of acquaintance is not tenable and that, consequently, the semantic arguments for the SESRist's structuralist thesis have no force. It has also been argued that Maxwell's "scientific" argument for the thesis is unconvinvcing. This does not mean that the thesis is not true, but it does mean that we have been given no good reason to think that it is true. In the next chapter we will look at Newman's objection, which (it will be argued) shows that both the SESRist's and the WESRist's structuralist claims are much stronger than they might initially appear and that they make SESR and WESR only trivially distinct from antirealism.

Chapter 4

Newman's Objection

Max Newman put forward his now famous objection to Russell's ESR in his review (Newman, 1928) of *The Analysis of Matter* (Russell, 1927). More recently, Demopoulos and Friedman (1985) and Ketland (2004) have put forward variants of the objection aimed at the Ramsey-sentence approach to ESR. Following the recent revival of interest in ESR numerous authors (Cruse, 2005, French and Ladyman, 2003a, Melia and Saatsi, 2006, Psillos, 1999, Redhead, 2001b, Votsis, 2003, 2004 and Zahar, 2001, 2004) have suggested a variety of ways in which one might attempt to evade the objection.

This chapter consists of four main sections. In section 1, the objection and its recent variants are outlined. In section 2, two responses that argue that the objection can be evaded by abandoning the Ramsey-sentence approach to ESR are considered. In section 3, three responses that have been put forward specifically to rescue the Ramsey-sentence approach to ESR from the modern versions of the objection are discussed. Finally, in section 4, three responses are considered that are neutral with respect to one's approach to ESR and all argue (in different ways) that the objection can be evaded by introducing the notion that some relations/structures are privileged over others. It is concluded that none of these suggestions is an adequate response to Newman's objection.

1. The Objection

1.1. Newman's Version

Newman (1928) takes Russell's ESR to imply that the most that we can know about the external world is its structure. He ascribes this view to Russell on the basis of passages like the following (which Newman quotes

[Newman, 1928, p. 144]): "Thus it would seem that wherever we infer from perceptions it is only structure that we can validly infer; and structure is what can be expressed by mathematical logic" (Russell, 1927, p. 254), "The only legitimate attitude about the physical world seems to be one of complete agnosticism as regards all but its mathematical properties" (Russell, 1927, p. 270). Newman then launches the following objection to this view:

> Any collection of things can be organised so as to have the structure W [where W is an arbitrary structure], provided there are the right number of them. Hence the doctrine that *only* structure is known involves the doctrine that *nothing* can be known that is not logically deducible from the mere fact of existence, except ("theoretically") the number of constituting objects. (Newman, 1928, p. 144, original emphasis)

For example, being told that a system has domain D = {a, b, c} (where a, b and c are arbitrary names for three distinct but unspecified objects) and instantiates a relation R= {<a, b>, <a, c>, <b, c>} tells us no more than that the system consists of three objects, because some elementary set-theory reveals that any three objects instantiate seven non-empty one-place relations, 511 non-empty two-place relations (of which R is one) and 134,217,727 non-empty three-place relations.[1] Being told that they instantiate R is both trivial (in so far as it follows from some elementary set-theory) and perversely specific (in so far as R is just one of the 134,218,245 non-empty relations they instantiate). Thus being told that the system has structure <D, R> is being told no more than that it contains three objects, because any system containing three objects can be taken to have this structure, along with a vast number of other structures (any tuple whose first

[1] A set of n objects has $2^n - 1$ non-empty subsets. n objects can be arranged into a set of n.n distinct pairs, which has $2^{n.n} - 1$ non-empty subsets. n objects can be arranged into a set of n.n.n triples, which has $2^{n.n.n} - 1$ non-empty subsets.

member is D and whose other members are amongst the 134,218,245 relations instantiated by the members of D is a structure that can be taken to be possessed by any system containing three objects).

The objection arises because our purely structural knowledge gives us only extensional information about the structure of the system: if we had an intensional interpretation of R, we would not have this problem. For instance, if we knew that R was the "heavier than" relation (restricted to the system) then we would have some more useful information: we would know that the three objects in the system were of unequal weights. However, the claim that we have any such intensional information about the external world is exactly what Newman thinks that Russell denies.

Newman considers two possible responses to this objection. The first is to distinguish "real" relations from "fictional" relations and assume that when the Russellian tells us what relations hold in a system he is talking only about real relations. A fictional relation is defined as "one whose only property is that it holds between the objects that it does hold between" (Newman, 1927, p. 145). By this Newman does not mean to call real only those relations that have interesting formal properties (like reflexivity, transitivity etc.) because the Russellian would presumably not wish to ignore all relations that lack interesting formal properties (for example, all one-place relations lack interesting formal properties, but the Russellian would presumably not wish to ignore all one-place relations). Rather he means that real relations are those that are the extensions of intensionally interpreted predicates. At first it seems that the problem here is that all relations are fictional for Russell, because, according to Newman, he denies that our knowledge claims about the external world involve any intensionally interpreted predicates (except perhaps equality). However, Newman claims that the problem is just the opposite, i.e. that all relations are real, because, having named the objects in the domain, each relation will be the extension of some intensionally interpreted predicate. In the above case, for example, R is the extension of the relation that holds between x and y just in case (x = a and y = b) or (x = a and y = c) or (x = b and y = c).

This leads Newman to consider a second possible response, which is to distinguish between "important" and "trivial" relations and assume that

when the Russellian tells us what relations hold in a system he is talking only about important relations. Newman dismisses this response as follows:

> we should have to compare the importance of relations of which nothing is known save their incidence (the same for all of them) in a certain aggregate. For this comparison there is no possible criterion, so that "importance" would have to be reckoned among the prime unanalysable qualities of the constituents of the world, which is, I think, absurd. (Newman, 1928, p. 147)

Although Newman thought that this response was absurd, a number of philosophers have put forward variants of it. These will be discussed in section 4.

However, isn't there a much more obvious response to Newman: hasn't he misunderstood the position he attacks? Newman imputes to Russell something like the following claim:

Our knowledge of the world is purely structural (i.e. it consists of claims constructed using only logical terms).

However it has been suggested (in chapter 3) that Russell actually held the following (weaker) view:

Our knowledge of the world consists of claims constructed using only logical and internal terms.

It is true that the passages of Russell that Newman quotes seem to suggest he would go along with the former (stronger) of these claims, but elsewhere Russell is quite explicitly committed to the view that we know how external objects are connected to internal experience, which he allows we can legitimately describe with non-logical terms. For example, Russell says that, "My knowledge of the table is of the kind which we shall call 'knowledge by description'. The table is 'the physical object which causes

such-and-such sense-data.' (Russell, 1912, p. 26). This seems to imply that he held only the latter (weaker) claim. If we understand Russell this way then it seems that Newman has misunderstood his position.

However, Russell did not respond to Newman this way (Russell's response is discussed in passing in subsection 3.2, as it relates to Cruse's [2005] response). It is difficult to believe that Russell just missed such an obvious rejoinder. Perhaps the reason he did not offer it is that he could see that although Newman's objection, as Newman states it, is not strictly speaking to the point, nonetheless Newman's line of thinking does lead one to the conclusion that ESR is not significantly distinct from standard antirealism. This point has been made by Demopoulos and Friedman (1985) and Ketland (2004) with respect to the Ramsey-sentence approach to ESR and is discussed in the next subsection.

1.2. Demopoulos and Friedman's and Ketland's Versions

Demopoulos and Friedman (quite anachronistically) impute to Russell a form of the Ramsey-sentence approach to ESR (Demopoulos and Friedman, 1985, p. 622) whereby the knowledge a scientific theory provides is expressed by the Ramsey-sentence of that theory, which will contain non-logical terms that refer to either internal objects, properties and relations (for the SESRist) or observable objects, properties and relations (for the WESRist) but will in either case not be purely structural. They claim (although without substantial argument) that if "our theory is consistent, and if all its purely observational consequences are true, then the truth of the Ramsey-sentence *follows* as a theorem of set theory or second-order logic, provided our initial domain has the right cardinality" (Demopoulos and Friedman, 1985, p. 635, original emphasis). Although Demopoulos and Friedman do not really back up this claim, Ketland (2004) does provide a strong argument for (a slight variant of) the claim.

To understand Ketland's argument we must distinguish between intended and arbitrary interpretations of a language. An intended interpretation is a function from the non-logical terms of the language to the objects, properties and relations in a structure that respects the intended

meanings of the non-logical terms. For example, under its intended interpretation, the name "Julius Caesar" is assigned the person Julius Caesar. Similarly, under its intended interpretation, the predicate "larger than" is assigned the set of pairs <x, y> such that x is larger than y. An arbitrary interpretation does not respect intended meanings in this way. In an arbitrary interpretation the name "Julius Caesar" might be assigned any arbitrary object and the predicate "larger than" might be assigned any arbitrary set of pairs.

Ketland assumes that we have a two-sorted second-order language. A two-sorted language (see, for example, Enderton, 2001, pp. 295-296) has two types of individual variables that range over two different domains. Many-sorted languages are harmless in the sense that they can be reduced to standard one-sorted languages without loss (except of convenience) (see Enderton, 2001, pp. 296-299). In this case the two domains (in the intended interpretation of the language) are observable objects and unobservable objects.

The language is also assumed to have three types of predicates: observational predicates, which (in the intended interpretation of the language) refer to observable properties and relations (which Ketland takes to be sets [of tuples] of observable objects), theoretical predicates, which (in the intended interpretation of the language) refer to unobservable properties and relations (which Ketland takes to be sets [of tuples] of unobservable objects) and mixed predicates, which (in the intended interpretation of the language) refer to mixed relations (which Ketland defines as sets of tuples such that each tuple contains at least one observable object and at least one unobservable object).[2]

---

[2] Ketland himself notes that the characterisation of observable (etc.) properties and relations as sets (of tuples) of observable (etc.) objects has some counterintuitive consequences. In particular he notes that, "many scientifically significant relations and quantities (e.g., various space-time relations and quantities, various scientific quantities such as mass, length, duration, location, etc.) will 'decompose' into three strangely distinct relations, depending upon the observational status of their relata" (Ketland, 2004, p. 289, footnote 5). Cruse (2005) argues that the ESRist (or at least the WESRist) can respond to Ketland's version of Newman's objection by denying that this is an accurate characterisation of the

Let $<D_O, O_1, O_2...>$ be the structure associated with the intended interpretation of the observational part of the language, i.e. let D_O be the set of observable objects in the world and let O_1, O_2 etc. be (the sets corresponding to) the observable properties and relations referred to by the observational predicates of the language.

We can now define what it means for an arbitrary structure for the language, $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ to be empirically correct ($<D_1, D_2>$ is an arbitrary two-sorted domain, each $R_{1.i}$ is an arbitrary interpretation of an observational predicate over D_1 [i.e. a subset of D_1 or a subset of $D_1 \times D_1$ etc.], each $R_{3.i}$ is an arbitrary interpretation a theoretical predicate over D_2 [i.e. a subset of D_2 or a subset of $D_2 \times D_2$ etc.] and each $R_{2.i}$ is an arbitrary interpretation of a mixed predicate over $D_1 \cup D_2$ [i.e. a subset of $D_1 \times D_2$ or a subset of $D_2 \times D_1$ etc.]). We do this as follows:

Definition 1: $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ is *empirically correct* if and only if its reduct $<D_1, R_{1.1}, R_{1.2}...>$ is isomorphic to $<D_O, O_1, O_2...>$. (cf. Ketland's "Definition E" [2004, p. 296])

In other words, a structure is empirically correct if the appropriate reduct of the structure is isomorphic to the structure of the observable world (relative to some choice of predicates). This definition of empirical correctness is in line with van Fraassen's (1980) notion of empirical adequacy: van Fraassen says that, "a theory is empirically adequate exactly if...[it] has at least one model that all the phenomena fit inside" (van Fraassen, 1980, p. 12). We will return to the issue of how this definition of empirical correctness compares with other notions of empirical correctness (or adequacy) later. Let's assume that the Ramsey-sentence of a theory in this language is obtained by Ramseyfying away the mixed and theoretical predicates that appear in the theory (cf. Ketland, 2004, p. 292). It follows that:

---

observable/unobservable distinction that he wishes to draw. This response is discussed in subsection 3.2.

Theorem 1: The Ramsey-sentence of a theory A is true if and only if there is some sequence of relations $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$... such that $<< D_O, D_T >, O_1, O_2..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...> \models A$ (cf. Ketland's "Theorem 4" [2004, p. 293])

where D_T is the set of unobservable objects in the world. (Ketland gives a proof of this result, this has been omitted, as the result itself seems intuitively obvious.) We need one more definition:

Definition 2: $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ is *T-cardinality correct* if and only if $|D_2| = |D_T|$. (cf. Ketland's "Definition G" [2004, p. 298])

and we can then prove:

Theorem 2: The Ramsey-sentence of a theory A is true if and only if A has a model that is empirically correct and T-cardinality correct. (cf. Ketland's "Theorem 6" [2004, p. 298])

The proof is in two steps:

[1] Left-to-right: Suppose the Ramsey-sentence of A is true. Then, by Theorem 1, there is some sequence of relations $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$... such that $<< D_O, D_T >, O_1, O_2..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...> \models A$, i.e. $<< D_O, D_T >, O_1, O_2..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ is a model of A. Clearly, $<D_O, O_1, O_2...>$ is isomorphic to $<D_O, O_1, O_2...>$, so $<< D_O, D_T >, O_1, O_2..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ is empirically correct (by definition 1). Equally clearly, $|D_T| = |D_T|$, so $<< D_O, D_T >, O_1, O_2..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ is T-cardinality correct (by definition 2).

[2] Right-to-left: Suppose A has a model, $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$, that is empirically correct and T-cardinality correct. As $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$ is empirically correct, $<D_1, R_{1.1}, R_{1.2}...>$ is isomorphic to $<D_O, O_1, O_2...>$, i.e. there is a bijection, f (D_1

$\rightarrow D_O$), such that, for every $R_{1.i}$ and every n-tuple, $<x_1, x_2,\ldots x_n>$, of elements of D_1:

$<x_1, x_2,\ldots x_n> \in R_{1.i}$ if and only if $<f(x_1), f(x_2),\ldots f(x_n)> \in O_i$

As $<<D_1, D_2>, R_{1.1}, R_{1.2}\ldots, R_{2.1}, R_{2.2}\ldots, R_{3.1}, R_{3.2}\ldots>$ is T-cardinality correct, $|D_2| = |D_T|$, i.e. there is a bijection, g ($D_2 \rightarrow D_T$). We can use f and g to define a new function, f*g ($<D_1, D_2> \rightarrow <D_O, D_T>$) such that f*g(x) = f(x) if $x \in D_1$ and f*g(x) = g(x) if $x \in D_2$ and we can use f*g to define new relations such that, for every $R_{2.i}$ and every $R_{3.i}$:

(i) $R_{2.i}' =_{df} \{<f*g(x_1), f*g(x_2),\ldots f*g(x_n)> : <x_1, x_2,\ldots x_n> \in R_{2.i}\}$
(ii) $R_{3.i}' =_{df} \{<f*g(x_1), f*g(x_2),\ldots f*g(x_n)> : <x_1, x_2,\ldots x_n> \in R_{3.i}\}$

By the construction of the $R_{2.i}$'s and $R_{3.i}$'s, f*g is an isomorphism between $<<D_1, D_2>, R_{1.1}, R_{1.2}\ldots, R_{2.1}, R_{2.2}\ldots, R_{3.1}, R_{3.2}\ldots>$ and $<<D_O, D_T>, O_1, O_2\ldots, R_{2.1}', R_{2.2}'\ldots, R_{3.1}', R_{3.2}'\ldots>$. We know that $<<D_1, D_2>, R_{1.1}, R_{1.2}\ldots, R_{2.1}, R_{2.2}\ldots, R_{3.1}, R_{3.2}\ldots> \models A$, so $<<D_O, D_T>, O_1, O_2\ldots, R_{2.1}', R_{2.2}'\ldots, R_{3.1}', R_{3.2}'\ldots> \models A$. So by theorem 1, the Ramsey-sentence of A is true.

QED.

How does Ketland's result compare to Demopoulos and Friedman's claim that if "our theory is consistent, and if all its purely observational consequences are true, then the truth of the Ramsey-sentence *follows* as a theorem of set theory or second-order logic, provided our initial domain has the right cardinality" (Demopoulos and Friedman, 1985, p. 635, original emphasis)? Ketland notes that his result is, strictly speaking, weaker, because it is in principle possible that all a theory's purely observational consequences could be true whilst it might not have an empirically correct model but not vice-versa. (An observational consequence here is assumed to be any statement formed using only observational predicates and logical terms, excluding predicate variables.

Observational consequences in this sense are thus assumed to include empirical generalisations. Zahar's reply to the modern version of the objection is based on the claim that this is inappropriate; this is discussed in subsection 3.1.)

It is not easy to give an example that demonstrates how all a theory's purely observational consequences could be true whilst it might not have an empirically correct model, but Ketland gives an example that demonstrates how something analogous can occur in number theory. Say that a model is "arithmetically correct" if and only if it has a reduct isomorphic to the standard natural number structure. Say that a theory has true "arithmetical consequences" if and only if all the consequences of the theory that are stated in the language of arithmetic are satisfied in the standard natural number structure. Let L be the language of arithmetic, and let L_T be the language of arithmetic extended by a monadic predicate symbol, T (which is intended to behave like a truth predicate). Now consider the theory FS*: this contains the axioms of Peano arithmetic and certain axioms concerning the predicate T (these are axioms that, intuitively speaking, a truth predicate should satisfy) (see Halbach, 1999, pp. 368-369). It can be shown that, (i) if M is a model of FS* then M does not have a reduct isomorphic to the standard natural number structure, so FS* does not have a model that is arithmetically correct, but nonetheless, (ii) every consequence of FS* that does not involve T is true in the standard natural number structure, so FS* has only true arithmetical consequences. For more on this see Ketland (2004, pp. 295-298).

However, despite the fact that Ketland's result is strictly speaking weaker than the one Demopoulos and Friedman claim, the difference appears to be immaterial: Ketland's result still suggests that ESR is not significantly distinct from antirealism. As noted, Ketland's notion of empirical correctness formalises van Fraassen's notion of empirical adequacy. Since van Fraassen argues that it is rational to believe that our mature theories are empirically adequate (but not necessarily true) Ketland's result shows that what the knowledge that a theory's Ramsey-sentence is (approximately) true would amount to, beyond what van Fraassen's antirealism allows, would be (at most) only knowledge of the cardinality of

the unobservable world. Thus ESR (in its Ramsey-sentence form) is just van Fraassen's antirealism, augmented by the peculiar claim that we can (perhaps) know the cardinality of the unobservable world.[3]

2. Replies that Abandon the Ramsey-Sentence Approach to ESR

In this section two arguments that claim that Newman's objection can be evaded if one abandons the Ramsey-sentence approach to ESR are discussed. On the face of it, this appears to be a strange line to take, because although Demopoulos and Friedman's and Ketland's versions of the objection are directed at the Ramsey-sentence approach to ESR Newman's original version of the objection is not. This first impression remains on closer inspection of these arguments.

2.1. Redhead's Reply

Redhead (2001b) argues that the Ramsey-sentence approach to ESR is indeed undermined by Newman's objection, commenting that, "the Ramsey sentence $\exists R(S[R])$, asserting the existence of a relation R which has structure S, is in fact a logical truth, modulo the specification of the cardinality of the domain over which the relation is defined" (Redhead, 2001b, pp. 345-346). This is false: the Ramsey-sentence of a theory is not satisfied by every model of the right cardinality, so the Ramsey-sentence is not "a logical truth, modulo the specification of the cardinality of the domain" (on the most obvious reading of this expression). For example consider the "theory" expressed by the sentence $\exists w \exists x ([w \neq x] \ \& \ \forall y [(y = w) \ v \ (y = x)]) \ \& \ \forall z(Pz)$ (which says that there are two things, and that

[3] The qualifications "at most" and "perhaps" appear here because knowing that a theory's Ramsey-sentence is true doesn't guarantee that we know the cardinality of the unobservable world, it only guarantees that we know that the theory has *some* model with the right cardinality (and the theory might have models of different cardinalities).

everything is P). It's Ramsey-sentence is $\exists X[\exists w \exists x ([w \neq x] \ \& \ \forall y [(y = w) \ v \ (y = x)]) \ \& \ \forall z(Xz)]$ (assuming P needs to be Ramseyfied). This is satisfied by the structure $<\{1,2\}; \{1,2\}>$ but not by every structure with two elements in its domain. For example, it is not satisfied by the structure $<\{1,2\}; \{1\}>$. Hence it is not "a logical truth, modulo the specification of the cardinality of the domain". What (Ketland's version of) Newman's objection actually shows is that *if the theory has an empirically correct model* then the theory's Ramsey-sentence is bound to be satisfied by *some* structure (instantiated by the world) as long as the world contains an appropriate number of (unobservable) things.

Redhead puts forward an alternative approach to ESR, which, he claims, avoids Newman's objection. He describes this alternative approach as follows:

> We need not deny that there are real physical relations posited by physical theories...Thus S(R), where R refers to a specific relation having the structure S, is of course logically stronger than the Ramsey sentence, and is by no means a logical truth. But this means...that the reference of R must be picked out in non-structural terms. But this is not denied in the above account. Our claim is merely that R is hypothesised in some explanatory theoretical context so it exists as an ontological posit, but all that we have epistemic warrant for is the second-order structure S. (Redhead, 2001b, p. 346).

Redhead appears to suggest that the Ramsey-sentence approach to ESR denies that there are real physical relations posited by physical theories. This is also false. The Ramsey-sentence approach does claim (roughly speaking) that all we know about (some of) these relations is structural, and this is (roughly speaking) why Newman's objection operates against it. But this is a claim that Redhead apparently endorses. The point of Newman's objection is (roughly speaking) that if all we know is that there is *some* (real) relation

R (of which we have only structural knowledge) then we know nothing more than a cardinality constraint on the domain over which the relation is defined.

On the other hand, in parts of the above quote Redhead seems to suggest that we can specify the relation R intensionally. If he does think this then he appears to have abandoned ESR, and it is hard to see how he can maintain the claim that we have only structural knowledge of R.

Perhaps the most charitable reading of Redhead's position would be to substitute the word "important" or "natural" for the words "real" and "specific" in the above quote. This would lead us back to Newman's own "absurd" response to his objection (variants of which are discussed in section 4). However, it is doubtful that this is really what Redhead intended. Neither "natural" nor "important" means the same as "real", much less is either synonymous with "specific". Moreover, there seems to be no reason why taking this line forces one to abandon the Ramsey-sentence approach (as will be seen when this approach is discussed in section 4).

2.2. French and Ladyman's Reply

French and Ladyman (2003a) appear to suggest that Newman's objection does not arise if one adopts the semantic view of theories (whereby a theory is taken to be a collection of structures) as opposed to the syntactic view (whereby a theory is taken to be a collection of sentences):

> 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. [Footnote omitted] 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 pseudo-problems 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. One of us (Ladyman 1998) has suggested an alternative descriptive framework for SR [structural realism], namely the "semantic" or model theoretic approach to theories. (French and Ladyman, 2003a, p. 33)

On the face of it, it seems highly unlikely that moving to the semantic view would really allow the ESRist to evade Newman's objection. In fact, Newman's original version of the objection is posed against the view that scientific theories directly specify a structure that represents the world. It is true that Demopoulos and Friedman (1985) and Ketland (2004) aim their objections at the Ramsey-sentence approach, which does assume that science presents us with a linguistic representation of the world. However, it is easy to show that an analogue of their objections applies to a version of ESR framed using the semantic view.

Framed in terms of the semantic view, ESR would imply something like the following limitation on our knowledge:

The most that we can know about the world is that some structure (provided by our scientific theories) is empirically correct and isomorphic to a structure instantiated by the world.

Let $<D_O, O_1, O_2...>$ be the structure instantiated by the observable world (relative to some chosen observational predicates): i.e. let D_O be the set of observable objects in the world and let O_1, O_2 etc. be the intended extensions of observational predicates. Let a theory present us with a structure, $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$, where D_1 is a domain that is to represent the set of observable objects, D_2 is a domain that is to represent the set of unobservable objects, each $R_{1.i}$ is to represent an observable relation (where an observable relation is taken to be a set of [tuples of] observable objects), each $R_{2.i}$ is to represent a mixed relation (where a mixed relation is taken to be a set of tuples that each contain at least one observable object and at least one unobservable object) and each

$R_{3.i}$ is to represent an unobservable relation (where an unobservable relation is taken to be a set of [tuples of] unobservable objects). Define such a structure to be empirically correct as before:

Definition 1: $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is *empirically correct* if and only if its reduct $\langle D_1, R_{1.1}, R_{1.2}...\rangle$ is isomorphic to $\langle D_O, O_1, O_2...\rangle$. (cf. Ketland's "Definition E" [2004, p. 296])

Let D_T be the set of unobservable objects in the world. Define $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ to be T-cardinality correct as before:

Definition 2: $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is *T-cardinality correct* if and only if $|D_2| = |D_T|$. (cf. Ketland's "Definition G" [2004, p. 298])

We can now prove that:

Theorem 3: $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is empirically correct and isomorphic to a structure instantiated by the world if and only if it is empirically correct and T-cardinality correct.

The proof comes in two stages:

[1] Left to right: Suppose $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is empirically correct and isomorphic to a structure instantiated by the world. I.e. $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is isomorphic to a structure of the form, $\langle\langle D_O, D_T\rangle, O_1, O_2..., R_{2.1}', R_{2.2}'..., R_{3.1}', R_{3.2}'...\rangle$, where O_1, O_2 etc. are the intended extensions of observational predicates. By stipulation it is empirically correct. And clearly, $|D_2| = |D_T|$, i.e. $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is T-cardinality correct.

[2] Right to left: Suppose $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is empirically correct and T-cardinality correct. By stipulation it is empirically

correct. As $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is empirically correct there is some isomorphism between $\langle D_1, R_{1.1}, R_{1.2}...\rangle$ and $\langle D_O, O_1, O_2...\rangle$. I.e. there is a bijection f ($D_1 \rightarrow D_O$), such that, for every $R_{1.i}$ and every n-tuple, $\langle x_1, x_2..., x_n\rangle$, of elements of D_1:

$\langle x_1, x_2,...x_n\rangle \in R_{1.i}$ if and only if $\langle f(x_1), f(x_2),...f(x_n)\rangle \in O_i$

As $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ is T-cardinality correct there is some bijection, g ($D_2 \rightarrow D_T$). We can use these functions to define a new function, f*g ($\langle D_1, D_2\rangle \rightarrow \langle D_O, D_T\rangle$) such that f*g(x) = f(x) if $x \in D_1$ and f*g(x) = g(x) if $x \in D_2$ and we can use f*g to define new relations such that, for every $R_{2.i}$ and every $R_{3.i}$:

(i) $R_{2.i}' =_{df} \{\langle f^*g(x_1), f^*g(x_2),... f^*g(x_n)\rangle : \langle x_1, x_2,... x_n\rangle \in R_{2.i}\}$
(ii) $R_{3.i}' =_{df} \{\langle f^*g(x_1), f^*g(x_2),... f^*g(x_n)\rangle : \langle x_1, x_2,... x_n\rangle \in R_{3.i}\}$

By the construction of the $R_{2.i}$'s and $R_{3.i}$'s, f*g is an isomorphism between $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle$ and $\langle\langle D_O, D_T\rangle, O_1, O_2..., R_{2.1}', R_{2.2}'..., R_{3.1}', R_{3.2}'...\rangle$. $\langle\langle D_O, D_T\rangle, O_1, O_2..., R_{2.1}', R_{2.2}'..., R_{3.1}', R_{3.2}'...\rangle$ is a structure instantiated by the world: we know that O_1, O_2 etc. are the intended extensions of observational predicates and each $R_{2.i}'$ is an arbitrary mixed relation that is obviously instantiated by the world (all its tuples are built from objects in the set $D_O \cup D_T$, i.e. the set of objects in the world) and similarly, each $R_{3.i}'$ is an arbitrary unobservable relation that is obviously instantiated by the world (all its tuples are built from objects in the set D_T, i.e. the set of unobservable objects in the world).[4]

QED.

[4] It may be objected that the $R_{2.i}$'s and the $R_{3.i}$'s, presumably unlike the O_is, may not correspond to any natural relations; this again leads to essentially Newman's own "absurd" response to his objection (variants of which are discussed in section 4).

Theorem 3 states that the semantic view formulation of ESR (i.e. the claim that the most that we can know about the world is that some structure – provided by our scientific theories – is empirically adequate and isomorphic to a structure instantiated by the world) is equivalent to the claim that the most that we can know about the world is that some structure (provided by our scientific theories), $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...>$, is empirically correct and T-cardinality correct, so Theorem 3 provides a Newman-esque reductio of the semantic view formulation of ESR. It is true that there is a slight difference between the ESRist who works with the semantic view and the ESRist who works with the syntactic view/Ramsey-sentence approach (because the latter's claim is equivalent to the even less optimistic view that the most we can know about the world is that one of the – possibly many – structures that satisfy a given theory is empirically correct and T-cardinality correct) but this difference does not appear to be very significant. (The difference arises because the proponent of the semantic view – at least as he has been characterised here – thinks that the theory specifies a unique structure, whereas the proponent of the syntactic view thinks that a theory is a sentence that picks out only a family of structures, all of which satisfy the sentence.)

3. Replies Designed to Rescue the Ramsey-Sentence Approach

In this section three replies that are designed to save the Ramsey-sentence approach to ESR from the modern versions of Newman's objection are considered.

3.1. Zahar's Reply

Zahar (2001 [in appendix 4, co-written with John Worrall], 2004) takes issue with Demopoulos and Friedman's version of Newman's objection, which he characterises as the claim that it is "only what the Ramsey-sentence asserts *over and above its observational content* [that] is reducible

to...a cardinality constraint" (Zahar, 2004, p. 10, original emphasis). Zahar goes on to say that:

> This "over and above" however proves to be essentially indefinable; for on the one hand, the Ramsey-sentence does not normally follow from its empirical basis, i.e. from the set of true and empirically decidable, hence *singular* sentences. If, on the other hand, all the – generally undecidable – "empirical generalisations" were included in the observational content of a theory, then the Ramsey-sentence might well turn out to be one of them; in which case Demopoulos's and Friedman's thesis collapses into the trivial claim that the Ramsey-sentence follows from itself (Zahar, 2004, p. 10, original emphasis).

Zahar's first "hand" holds the proposition that only singular sentences form the observational content of a theory. He then demonstrates that Ramsey-sentences can go beyond such observational content. He asks us (Zahar, 2004, p. 11) to consider a theory expressed by the sentence:

$A: \forall x(Fx \rightarrow Tx) \ \& \ \forall y(Ty \rightarrow Ky)$

where F and K are observational predicates and T is a theoretical predicate. The Ramsey-sentence of this theory is:

$A^*: \exists X(\forall x[Fx \rightarrow Xx] \ \& \ \forall y[Xy \rightarrow Ky])$

which is equivalent to:

$\forall x(Fx \rightarrow Kx)$

and this last sentence (and hence the equivalent A*) goes beyond any number of singular statements of the form:

$Fa_i \rightarrow Ka_i$

in the sense that no matter how many statements of this form we have there is always a model in which they are all true but in which the generalisation $\forall x(Fx \rightarrow Kx)$ (and hence the equivalent A*) is false. So there is a clear sense in which the Ramsey-sentence of the theory goes beyond (Zahar's understanding of) the observational content of the theory.

However, given Zahar's understanding of the observational content of a theory as consisting of the singular sentences (containing only observational terms) entailed by the theory then not only do Ramsey-sentences typically go beyond observational content but, as Zahar's example clearly illustrates, universal generalisations that involve only observational predicates (i.e. "empirical generalisations") also go beyond observational content. Even the antirealist would typically agree that we can know such generalisations to be true, so even the antirealist would agree that we can know more than the observational content of a theory, in Zahar's sense of observational content.[5] Unless the ESRist can demonstrate that Ramsey-sentences also go beyond empirical generalisations he has failed to distinguish his position from antirealism. In fact, Zahar clearly states (in the above quote) that the Ramsey-sentence of a theory might often be (equivalent to) an empirical generalisation. At this point he seems to concede to Demopoulos and Friedman even more than they ask for: they claim that ESR is antirealism plus a cardinality constraint, while Zahar seems to concede that ESR is (often) plain antirealism (because he claims

[5] Typically, the antirealist would say that we can have knowledge of singular empirical statements and empirical generalisations (i.e. generalisations that do not involve theoretical terms or second-order variables) but would deny that we generally have knowledge of a theory's Ramsey-sentence (since, notwithstanding Zahar's example, a theory's Ramsey-sentence is not typically equivalent to an empirical generalisation). (Although what Newman's objection shows is – roughly speaking – that knowing a theory's Ramsey-sentence to be true is knowing very little beyond knowing that the singular empirical statements and empirical generalisations that follow from the theory are true, so ESR collapses into a position not significantly distinct from antirealism.)

that Ramsey-sentences are [often] equivalent to empirical generalisations, so the ESRists' claim that we can have knowledge of at most Ramsey-sentences is [often] equivalent to the antirealists claim that we can have knowledge of at most empirical generalisations).

Nonetheless, Zahar denies that ESR is just antirealism. This denial seems to rest on an equivocation over the meaning of "observational content". On the one hand he suggests that the difference between realists and antirealists is that the latter deny that we can have knowledge that goes beyond the observational content of a theory, which is true, but only if the observational content of a theory is taken to *include* empirical generalisations. He then demonstrates that the Ramsey-sentence of a theory goes beyond the observational content of a theory, where this is now taken to *exclude* empirical generalisations. Thus his conclusion that knowing the Ramsey-sentence of a theory is knowing more than the antirealist would allow does not follow.

3.2. Cruse's Reply

Cruse's (2005) reply is directed at Ketland's version of the objection. In particular, Cruse objects that not all ESRists need draw the observational term/theoretical term distinction in the way that Ketland suggests (and given a suitably different construal of the distinction, Ketland's proof of theorem 2 would not go through). Recall that Ketland takes observational predicates to refer exclusively to sets of (tuples of) observable objects, theoretical predicates to refer exclusively to sets of (tuples of) unobservables objects and mixed predicates to refer to sets of tuples that each contain at least one observable and one unobservable object. As noted, Ketland acknowledges that this has some rather counter-intuitive consequences. Cruse emphasises this point:

> consider the relation denoted by the predicate "larger than". On Ketland's taxonomy, there is no such single relation; there are three. First, there is the relation we might call *observably larger than*, which ranges entirely over

> observable objects. Second[,] there is the relation we might call *unobservably larger than*, which ranges entirely over unobservable objects. Third, there is the relation we might call *miscellaneously larger than*, which applies to all and only pairs of objects such that the first is observable, the second unobservable, and the first larger than the second.[6] On Ketland's terminology, only the first class of relations – those which range entirely over observable objects – count as observable...I will call this the *strong* version of the observational-theoretical (O/T) distinction. (Cruse, 2005, p. 561, original emphasis, footnote added)

Cruse's reply is based on rejecting the strong observational term/theoretical term distinction.

Cruse notes that *some* ESRists do appear to be committed to this form of the observational term/theoretical term distinction or rather (as he puts it) "something isomorphic to it" (Cruse, 2005, p. 563). Translated into the terminology of this thesis, he suggests that the internal term/external term distinction employed by SESRists (such as Russell) must take this form because, "the mental and physical domains are entirely disjoint, so no (non-mathematical) property which applies to a mental event applies to a physical event or vice versa. Similarly, no (non-mathematical) predicate which applies to a mental event will apply to a physical event" (Cruse, 2005, p. 563).

In fact, in the face of Newman's objection, Russell decided that he was not committed to this form of the distinction. In a letter to Newman (reprinted in his autobiography [1968] and by Demopoulos and Friedman [1985]) he wrote:

[6] This is splitting hairs, but it seems that we could also have an unobservable object that was larger than an observable object if, for example, "the universe" or "the nearest black hole" qualify as unobservable objects.

> It was quite clear to me, as I read your article, that I had not really intended to say what in fact I did say, that nothing is known about the physical world except its structure. I had always assumed spacio-temporal continuity with the world of percepts, that is to say, I had assumed that there might be co-punctuality between percepts and non-percepts...And co-punctuality I regarded as a relation which might exist among percepts and is itself perceptible. (Russell, 1968, p. 176)

Russell is here putting forward the view that there is at least one non-logical predicate ("co-punctuality") that refers to a relation that can hold between (i) pairs of external objects, (ii) pairs consisting of one external object and one internal object (in either order) and (iii) pairs of internal objects. Russell's own reply to Newman is actually essentially the same as Cruse's reply to Ketland, which is perhaps surprising, as Cruse cites Russell as the sort of ESRist for whom this reply is not available.

There are good reasons to think that this sort of reply isn't available to Russell. As Demopoulos and Friedman (1985) point out, Russell's move is completely ad hoc:

> in the earlier theory [i.e. Russell's (1927) theory] we could not assume acquaintance with a cross category notion such as spacio-temporal contiguity or causality; but in the light of the difficulties of that theory we now find that we *can* assume this! [footnote omitted] We are not saying that one *cannot* resolve the issue in this way. But it seems quite clear that without a considerable advance in the theoretical articulation of this rather elusive Russellian concept [i.e. acquaintance], no such resolution of the difficulty can be very compelling. (Demopoulos and Friedman, 1985, p. 632, original emphasis)

I would go further: it seems that, given the supposedly radical difference between external and internal objects, it is very unlikely that the issue could be satisfactorily resolved this way. Moreover, if Russell makes this concession then he seems to be left at the top of a slippery slope: if we can assume that external objects can be "co-punctual" with one another in the same way that internal objects sometimes are why can't we assume that they can be "bigger than" one another in the same way? It thus does seem (as Cruse suggests) that SESRists (such as Russell) are committed to a internal term/external term distinction of the form Ketland suggests (so they cannot evade the Newman/Ketland objection this way).

However, Cruse's main point (translated into the current idiom) is that the WESRist is not committed to anything like Ketland's form of the observational term/theoretical term distinction. He proposes an alternative form of the distinction according to which:

> observational predicates refer to, broadly speaking, perceptible, or observable properties such as redness or squareness. A natural understanding of this would be that these observable properties are unproblematic not because they are *always* observable, but simply because we can in *at least some* cases observe them (Cruse, 2005, p. 565, original emphasis)

This is supposed to capture the intuition that, "we can meaningfully (and for a realist, truly) assert the existence of red blood cells, or microscopic square grids, for example" (Cruse, 2005, p. 564) i.e. that observational predicates can be applied to unobservable objects. A natural interpretation of this suggestion (natural in the light of the foregoing discussion, at any rate) is that rather than taking an observational predicate to be one whose intended extension is a set consisting *only* of (tuples of) observable objects (as Ketland suggests) we are to take an observational predicate to be one whose intended extension is a set consisting of *at least some* (tuples of) observable objects. Theoretical predicates would then be those whose intended extension is a set consisting entirely of (tuples of) unobservable objects.

However, interpreted this way, Cruse's suggestion is also deeply counter-intuitive, because it classes as observational a number of predicates that are, intuitively speaking, theoretical. It is true that "being a superstring" is on this account a theoretical predicate, because no superstring is observable, but "being a collection of superstrings" is observational, because it applies to some (in fact, if the theory is correct, all) observable objects. And it is (to say the least) counter-intuitive to classify "being a collection of superstrings" as observational.

However, this is not the only possible interpretation of Cruse's suggestion. In fact, this interpretation assumes that observability of properties and relations must be defined in terms of observability of objects. But it has already been noted (in chapter 2) that there is no compelling reason to assume this. It is quite conceivable that the extensions of observational predicates can contain both observable and unobservable objects, *and* that the extensions of theoretical predicates can contain both observable and unobservable objects, depending on what the criteria of observability are.[7]

If the WESRist adopts an observational term/theoretical term distinction along these lines then it is true that Ketland's proof of theorem 2 does not go through (Ketland's proof crucially assumes that observational predicates apply only to sets of [tuples of] observables, which is not the case with this characterisation of the observational term/theoretical term distinction). The WESRist can thereby evade the conclusion that:

Theorem 2: The Ramsey-sentence of a theory A is true if and only if A has a model which is empirically correct and T-cardinality correct. (cf. Ketland's "Theorem 6" [2004, p. 298])

[7] It has been argued that there are problems with any attempt to draw the observable/unobservable distinction in a way that is suitable for the WESRists purposes. The point here is that there is no need to assume that, if the observable/unobservable distinction can be suitably drawn, it must be such that the observability of properties and relations is defined in terms of the observability of objects.

However, the WESRist would be well advised to leave the champagne on ice, if not in the cellar. Even using this more liberal characterisation of the observational term/theoretical term distinction we can still prove a theorem that casts doubt on the view that knowledge of a theory's Ramsey-sentence is the sort of knowledge that the WESRist wants to claim that we have.

We assume that we have a language containing a number of observational predicates (construed as above, so that they may apply to observable and unobservable objects) and a number of theoretical predicates (construed as above, so that they too may apply to observable and unobservable objects). The structure associated with the intended interpretation of the observational predicates is:

$<D_A, O_1, O_2...>$

where D_A is the domain of (observable and unobservable) objects in the world that instantiate some observable property or relation that is referred to by one of the observational predicates of the language and each O_i is the intended extension of an observational predicate of the language. Now, given a theory, A, the Ramsey-sentence of A is obtained by Ramseyfying away the theoretical predicates. As before, it is obvious that:

Theorem 4: The Ramsey-sentence of a theory A is true if and only if there is some sequence of relations, $R_{2.1}$, $R_{2.2}...$ such that $<<D_A, D_B>, O_1, O_2..., R_{2.1}, R_{2.2}...> \models A$

where D_B is the domain of objects in the world that do not instantiate any observable property or relation that is referred to by one of the observational predicates of the language (depending on the choice of these predicates, D_B may well be the empty set, e.g. if one of the observational predicates is "larger than" D_B will presumably be empty as every object is surely on at least one side of this relation to at least one other object). We will say of an arbitrary structure for the language, $<<D_1, D_2>, R_{1.1}, R_{1.2}...R_{2.1}, R_{2.2}...>$, (where $<D_1, D_2>$ is an arbitrary two-sorted domain, each $R_{1.i}$ is an arbitrary

interpretation of an observational predicate over D_1, and each $R_{2.i}$ is an arbitrary interpretation of a theoretical predicate over $D_1 \cup D_2$) that:

Definition 3: $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}\ldots R_{2.1}, R_{2.2}\ldots\rangle$ *gets the extensions of the observational predicates right* if and only if its reduct $\langle D_1, R_{1.1}, R_{1.2}\ldots\rangle$ is isomorphic to $\langle D_A, O_1, O_2\ldots\rangle$

We will also say that:

Definition 4: $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}\ldots R_{2.1}, R_{2.2}\ldots\rangle$ is *B-cardinality correct* if and only if $|D_2| = |D_B|$.

It is easy to prove:

Theorem 5: The Ramsey-sentence of a theory A is true if and only if A has a model which gets the extensions of the observational predicates right and which is B-cardinality correct.

The proof is in two steps:

[1] Left-to-right: Suppose the Ramsey-sentence of A is true. Then, by Theorem 3, there is some sequence of relations $R_{2.1}$, $R_{2.2}$... such that $\langle\langle D_A, D_B\rangle, O_1, O_2\ldots, R_{2.1}, R_{2.2}\ldots\rangle \models A$, i.e. $\langle\langle D_A, D_B\rangle, O_1, O_2\ldots, R_{2.1}, R_{2.2}\ldots\rangle$ is a model of A. Clearly, $\langle D_A, O_1, O_2\ldots\rangle$ is isomorphic to $\langle D_A, O_1, O_2\ldots\rangle$, so $\langle\langle D_A, D_B\rangle, O_1, O_2\ldots, R_{2.1}, R_{2.2}\ldots\rangle$ gets the extensions of the observational predicates right (by definition 3). Equally clearly, $|D_B| = |D_B|$, i.e. $\langle\langle D_A, D_B\rangle, O_1, O_2\ldots, R_{2.1}, R_{2.2}\ldots\rangle$ is B-cardinality correct (by definition 4).

[2] Right-to-left: Suppose A has a model, $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}\ldots R_{2.1}, R_{2.2}\ldots\rangle$, which gets the extensions of the observational predicates right and which is B-cardinality correct. As $\langle\langle D_1, D_2\rangle, R_{1.1}, R_{1.2}\ldots R_{2.1}, R_{2.2}\ldots\rangle$ gets the extensions of the observational predicates right, $\langle D_1, R_{1.1}, R_{1.2}\ldots\rangle$ is

isomorphic to $<D_A, O_1, O_2...>$, i.e there is a bijection f ($D_1 \rightarrow D_A$), such that, for every $R_{1.i}$ and every n-tuple, $<x_1, x_2...x_n>$, of elements of D_1:

$<x_1, x_2,...x_n> \in R_{1.i}$ if and only if $<f(x_1), f(x_2),...f(x_n)> \in O_i$

As $<<D_1, D_2>, R_{1.1}, R_{1.2}...R_{2.1}, R_{2.2}...>$ is B-cardinality correct, there is a bijection g ($D_2 \rightarrow D_B$). We can use f and g to define a new function, f*g ($<D_1, D_2> \rightarrow <D_A, D_B>$) such that f*g(x) = f(x) if $x \in D_1$ and f*g(x) = g(x) if $x \in D_2$ and we can use f*g to define new relations such that, for every $R_{2.i}$:

$R_{2.i}' =_{df} \{<f*g(x_1), f*g(x_2),... f*g(x_n)> : <x_1, x_2,... x_n> \in R_{2.i}\}$

By the construction of the $R_{2.i}$'s, f*g is an isomorphism between $<<D_1, D_2>, R_{1.1}, R_{1.2}...R_{2.1}, R_{2.2}...>$ and $<<D_A, D_B>, O_1, O_2..., R_{2.1}', R_{2.2}'...>$. We know that $<<D_1, D_2>, R_{1.1}, R_{1.2}..., R_{2.1}, R_{2.2}...> \models A$, so $<<D_A, D_B>, O_1, O_2..., R_{2.1}', R_{2.2}'...> \models A$. So by theorem 3, the Ramsey-sentence of A is true.

QED.

Theorem 5 implies that the Ramsey-sentence of a theory can tell us *something* substantial about the world, beyond what the antirealist typically allows: it tells us about the "observable" properties and relations of the unobservable world. So ESR, construed on these lines, does represent a halfway house between realism and antirealism. However, it doesn't look like the house the ESRist claims to inhabit: the position implies that we have no non-trivial knowledge of unobservable properties and relations (and, in particular, no interesting structural knowledge of such properties and relations). This implies that predicates likes "strangeness" or "being a collection of superstrings" serve only an instrumentalist function in our theories, which seems to be at odds with the traditional ESRist's claims.[8]

[8] Discussing whether or not this position was actually plausible in any detail would be tangential to the aims of this thesis. It would also be a highly scholastic exercise, because it is not a position that anyone holds. However, there is one very obvious objection to this

3.3. Melia and Saatsi's Reply

Melia and Saatsi's (2006) response to Newman's objection is based on the observation that:

> The properties postulated in scientific theories are typically taken to stand in certain intensional relations to various other properties. Some properties *counterfactually depend* on others, some are *correlated in a law-like manner* with others, some are *independent* of others, and some are *explanatory* of others. (Melia and Saatsi, 2006, pp. 579-580, original emphasis)

Melia and Saatsi point out that such relations between properties (i.e. second-order relations) cannot be expressed in standard second-order logic.[9] Moreover, they claim that if we formulate scientific theories and their Ramsey-sentences in a language that is capable of expressing such relations then Newman's objection will be blocked.

As Melia and Saatsi note (2006, p. 580) the obvious way to formulate such relations would be to introduce higher order predicates into one's language. This is not the approach they ultimately favour, but let's consider this approach first. Consider a "theory" which states that there is a click on a Geiger counter if and only if there is an atom in the vicinity of the Geiger counter that radioactively decays. We might attempt to formalise this theory as follows:

---

position, which is that it relies on a sharp distinction between the observable and unobservable and, as argued in chapter 2, drawing such a distinction in a principled way is highly problematic.

[9] The terminology could become confusing here. I call properties and relations *of* properties and relations second-order properties and relations. It is important to note that second-order logic is so-called because it allows for quantification over sets of (tuples of) objects as well as objects and not because it accommodates second-order properties and relations in this sense.

$\exists xCx \leftrightarrow \exists y(Ay \ \& \ Dy)$

where "A" and "D" are "theoretical" predicates such that "Ax" means "x is an atom in the vicinity of a Geiger counter" and "Dx" means "x radioactively decays" and "C" is an "observational" predicate such that "Cx" means "x is a click in a Geiger counter". Melia and Saatsi would say that implicit in the theory is the claim that the correlation between Geiger counter clicks and radioactive decay is a lawful (as opposed to accidental) correlation. Consequently, they would argue that a more faithful formalisation of the theory would be as follows:

$(\exists xCx \leftrightarrow \exists y[Ay \ \& \ Dy]) \ \& \ \mathbf{L}DC$

where **L** is a second-order predicate such that **L**XY means "X is lawfully correlated with Y". Consequently the Ramsey-sentence of the theory is not

$\exists X \exists Y(\exists xCx \leftrightarrow \exists y[Xy \ \& \ Yy])$

which would be contentless[10] but is rather:

$\exists X \exists Y[(\exists xCx \leftrightarrow \exists y[Xy \ \& \ Yy]) \ \& \ \mathbf{L}XC]$

which is not so trivial: it states that *some* property has a lawful correlation with clicks on Geiger counters.

However, this assumes that the second-order predicate **L** does not need to be "Ramseyfied". If it does then we would obtain the following "Ramsey-sentence":

[10] The Ramsey-sentence of this "theory" (formalised this way) is completely contentless: it states that there is a click on a Geiger counter if and only if there is something that has two (not necessarily distinct) properties. This is utterly trivial since a click on a Geiger counter is something that has at least one and hence two (not necessarily distinct) properties.

$\exists \mathbf{X} \exists X \exists Y[(\exists x Cx \leftrightarrow \exists y[Xy \ \& \ Yy]) \ \& \ \mathbf{X}XC]$

which is, again, effectively contentless: it states that there is a click on a Geiger counter if and only if there is something that has two (not necessarily distinct) properties and that there is some (second-order) relation between one of these properties and clicks on Geiger counters. This is contentless because there is always *some* second order relation between *any* two properties: given any two properties P and Q we can construct the second-order relation {<P, Q>} between them.

So this response is only viable if it is reasonable to suppose that second-order relations between properties like "is lawfully correlated with" do not themselves need to be Ramseyfied. But this is surely not something the ESRist could consistently accept: "lawful correlations" and their ilk are surely not observable, much less internal, relations.

As noted, Melia and Saatsi do not advocate the use of higher order predicates to formalise the relations between properties. They suggest that we should instead augment the language with a number of modal operators that express the pertinent relations: "So, for instance, let L_P express 'it is physically necessary that…'. Then $\exists X L_P \forall x(Xx \leftrightarrow Gx)$ says that there is a property which is *lawfully* coextensive with G." (Melia and Saatsi, 2006, p. 581). The claim is that scientific theories, and their Ramsey-sentences, would typically (perhaps invariably) include such modal operators. This sidesteps the problem that undermines the previous approach, because there can be no question of "Ramseyfying" modal operators. This approach is, however, open to the objection that it requires us to take these modal operators as logical primitives and we surely cannot accept that modal operators expressing things like "it is *physically* necessary that…" can be taken as *logical* primitives, since whether or not something happens as a matter of *physical* necessity is an issue that must be decided empirically, not as a matter of logic.

4. Replies that Argue that Some Structures/Relations are Privileged

In this section three variants of Newman's own "absurd" response to his objection are considered. This response is founded on the claim that some relations are more important than others. Newman took this response to be absurd because, as the ESRist believes we have only structural knowledge of the relations in question, he lacks a criterion to distinguish the important relations from the unimportant relations, so "importance" must be left as an unanalysed primitive, a mysterious quality that attaches to some relations but not others. However, the idea of a primitive important/unimportant distinction (or something similar) has not appeared absurd to everyone. The proposal most similar to Newman's own was put forward by Carnap (1967) to resolve an analogue of Newman's problem that faces the theory he puts forward in the *Aufbau.* The adaptation of this proposal to resolve Newman's objection to ESR is discussed in subsection 4.1. The other two variants of this approach discussed here both in some sense deny that "importance" needs to be taken as primitive. Votsis' (2003, 2004, chapter 4) proposal (discussed in subsection 4.2) grounds the importance of a relation on the means by which it is discovered. Psillos' (1999, chapter 3) proposal (discussed in subsection 4.3) takes the importance of a relation to be a contingent physical property of the relation. Psillos himself argues that taking the approach ultimately amounts to abandoning ESR. (Both Merrill [1980] and Lewis [1983] make suggestions similar to Psillos' in response to Putnam's [1977] "model-theoretic" argument against realism – an argument that is very closely related to Newman's objection to ESR.)

4.1. A Carnapian Reply

This reply has some similarities to Melia and Saatsi's (2006) reply (discussed in subsection 3.3). The essence of the proposal is the suggestion that we should take importance (or as Carnap [1967] calls it, "foundedness") as a primitive (second-order) logical property that attaches to some relations (in the way that identity is sometimes taken as a primitive logical relation that holds between some pairs). With a little adaptation of the Ramsey-

sentence approach, this enables the proponent of ESR to evade Ketland's variant of the Newman objection. Let the symbol for the foundedness property be "**Found**". Instead of advocating belief in a theory's Ramsey-sentence, the ESRist who takes this approach should advocate belief in the theory's Ramsey-sentence*, where the latter is just like a normal Ramsey sentence, except that, for each predicate variable, X, we add (in the scope of the quantifier ∃X) the phrase, "& **Found**(X)". For example, consider the toy theory:

$\forall x(O_1x \rightarrow T_1x) \ \& \ \forall y(T_2y \rightarrow O_2y)$

where O_1 and O_2 are observational predicates and T_1 and T_2 are theoretical predicates. This theory yields the Ramsey-sentence:

$\exists X \exists Y(\forall x[O_1x \rightarrow Xx] \ \& \ \forall y[Yy \rightarrow O_2y])$

and the Ramsey-sentence*:

$\exists X \exists Y(\forall x[O_1x \rightarrow Xx] \ \& \ \forall y[Yy \rightarrow O_2y] \ \& \ \mathbf{Found}[X] \ \& \ \mathbf{Found}[Y])$

As "**Found**" is taken as a logical primitive the Ramsey-sentence* contains only logical and observational terms. It is thus hygienic, by the WESRists standards (in that the WESRist's claim does not imply that it is impossible to know the Ramsey-sentence*). However, if Ramsey-sentences are swapped for Ramsey-sentence*s then Ketland's argument no longer goes through. In particular, the relevant analogue of:

Theorem 1: The Ramsey-sentence of a theory A is true if and only if there is some sequence of relations, $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$...,. such that $\langle\langle D_O, D_T \rangle, O_1, O_2..., R_{2.1}, R_{2.2}..., R_{3.1}, R_{3.2}...\rangle \models A$

is:

Theorem 1': The Ramsey-sentence* of a theory A is true if and only if there is some sequence of relations, $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$..,. such that << D_O, D_T >, O_1, O_2..., $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$...> $\models$ A *and such that each member of the sequence is a founded relation.*

This blocks Ketland's proof of theorem 2 at the last step: although given that a theory has some model, << D_1, D_2 >, $R_{1.1}$, $R_{1.2}$..., $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$...>, that is empirically correct and T-cardinality correct we can construct a model of the theory of the form << D_O, D_T >, O_1, O_2..., $R_{2.1}$', $R_{2.2}$'..., $R_{3.1}$', $R_{3.2}$'...> this does not guarantee that the theory's Ramsey-sentence* is true, as there is no guarantee that the relations $R_{2.1}$', $R_{2.2}$'... we have constructed will be founded.

However, even Carnap does not have licence to invent logical predicates at whim; if we are to accept "**Found**" as a new logical term we surely must be given reason to do so. Here is what Carnap says:

> [**Found**] does not belong to any definite extralogical domain, as all other nonlogical objects do. Our considerations concerning the characterisation of the basic relations of a constructional system as founded relation extensions of a certain kind hold for every constructional system of any domain whatever. It is perhaps permissible, because of this generality, to envisage the concept of foundedness as a concept of logic and to introduce it, since it is undefinable, as a *basic concept of logic*. (Carnap, 1967, p. 237, original emphasis)

We *might*, just possibly, think that the following was plausible:

If a property or relation is instantiated in every possible system then it is a logical property or relation.

However, if I understand him correctly, Carnap seems to assume something like the following:

If a property or relation is instantiated in every possible constructional system then it is a logical property or relation.

The notion of a constructional system is central to Carnap's theory in the *Aufbau*, but it is not the case that every possible system is a possible constructional system, so it is hard to see why we should accept this assumption. Compare Carnap's assumption with the following assumption:

If a property or relation is instantiated in every possible system whose domain contains human beings then it is a logical property or relation.

No one would accept this: it would lead to the conclusion that properties like "being a human being" and "being a mammal" are logical properties. So it seems we at least need an argument that shows that constructional systems are more important (founded, as it were) than systems whose domain contains human beings.

It is instructive to contrast foundedness with identity, a relation we might sensibly take to be logical. In the first place we can see that the identity relation will be instantiated by some pairs of objects from any possible non-empty domain; by contrast the foundedness property need not be instantiated by any relations from a given set of relations. In the second place, given a domain we can determine the extension of the identity relation over that domain a priori; by contrast we cannot determine the extension of the foundedness relation over a set of (extensionally specified) relations a priori.

It seems that it is fair to say that taking the notion of the importance of a relation as a primitive logical notion *is* absurd and cannot form the basis of a reasonable response to Newman's objection.

4.2. Votsis' Reply

Votsis claims that Newman's objection purports to show that "the knowledge claims of SR [structural realism] [are] of little worth or

importance" (Votsis, 2003, p. 886) by showing that "the information they offer can also be derived a priori from set theory modulo a cardinality constraint, hence the only important information contained in the structural realist claims concerns the cardinality of the domain" (Votsis, 2003, p. 886).[11] He also claims that the inference from the latter to the former rests on the assumption that, "any information contained in a statement that is also derivable a priori lacks importance" (Votsis, 2003, p. 886). Votsis takes issue with this assumption. There is a sense in which everyone will surely agree that this assumption is false. The statement that there is no largest prime is surely of some importance, at least in some contexts, but the claim is derivable a priori.

However, Votsis does not dispute the assumption that "any information contained in a statement that is also derivable a priori lacks importance" (Votsis, 2003, p. 886) by pointing out that there is a sense in which some results that are obtained a priori are important. Rather, he seems to make the extraordinary suggestion that *how* a claim is arrived at affects its importance. In particular, he seems to suggest that a claim is more important if it is arrived at empirically than if it is arrived at a priori. He claims that, "The *method* of arriving at the abstract structures is at least partly empirical...The fact that set theory also allows me to derive the same structure a priori does not mean that the information I have reached is devoid of importance" (Votsis, 2003, p. 887, original emphasis). But surely this is wrong: if the information is of no importance if it is arrived at by set theory it is surely still of no importance if it is arrived at empirically. A fact that is unimportant, in so far as it can be easily discovered, does not become

[11] This isn't strictly true, at least if we take the modern version of Newman's objection that is directed at the Ramsey-sentence approach to ESR. In this form, what Newman's objection shows is that knowing that a theory's Ramsey-sentence is true is only knowing that the theory has a model that is T-cardinality correct *and empirically correct.* Moreover, it seems that the essence of Newman's objection is not so much that ESR implies that scientific theories are of little worth but rather that it implies that all they tell us about the unobservable world is its cardinality, so ESR is not significantly distinct from antirealism. So it seems that Votsis fails to address the right issue in his reply to Newman's objection. However, this is not too important: it will be argued that he does not satisfactorily address the wrong issue.

more important because it can also be discovered by an unnecessarily difficult route. For example, suppose someone claims to have made the important discovery that "eggs is eggs". We might well reply that their discovery is not in fact important, in so far as it is easy to show (a priori) that "eggs is eggs": the claim follows directly from the a priori principle that everything is self-identical. It would be ridiculous if they were to reply that the discovery *was* important because they arrived at it by an empirical study of eggs. Indeed, it seems that Votsis' contention comes down to the claim that "two" results can be the same but of different importance, a claim that violates the law of the indiscernibility of identicals.

Votsis attempts to support his contention with a thought experiment:

> 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. Despite the similarities, there is an important difference between the two cases. The latter case is one in which 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 main point is quite simple: The fact that arithmetic allows me to do this a priori does not mean that the information I have reached counting objects is of little or no importance. One need only consider the consequences if I had made an error in counting. (Votsis, 2003, p. 886, original emphasis)

This case does not at all support Votsis' contention. In this case the two procedures, the a priori arithmetical procedure of adding 133 and 123 and the empirical procedure of counting the objects in two collections, do not

achieve the same results. The former enables us to determine that 133 + 123 = 256, whereas, as Votsis notes, the latter enables us to determine that there are 256 objects in a particular collection. These results may well be of different importance but we cannot infer from this, as Votsis does, that two procedures yielding the same result can yield results of different importance, as the results in this case are not the same: the former is a theorem of arithmetic, the latter is a contingent fact about the world.

Votsis goes on to note that, "Using the...*a priori method*, set theory allows us to set up any structure we like...No structure is privileged in this sense. The structural realist's *a posteriori method* guarantees that some structures are privileged over others." (Votsis, 2003, p. 887, original emphasis). There is a trivial sense in which the structures that have been arrived at a posteriori are privileged compared with those that have not been arrived at a posteriori, but it is not clear that there is an important sense: it is not obvious why the fact that some structures have been arrived at a posteriori guarantees that these structures are *more important* than those structures that have been arrived at "merely" a priori. Simply being arrived at via an a posteriori method does not seem to be sufficient to make a result important, especially if that result could have been arrived at a priori. After all, if the claim, "eggs is eggs" *had* been discovered to be true a posteriori it would not thereby be more important than those identity claims that had been that had been arrived at "merely" a priori.

Newman argued (in effect) that the ESRist (unlike the conventional realist) does not have the resources to distinguish the important structures instantiated by a system from the unimportant structures, or even to say in what sense one structure could be more important than another. Votsis seems to be suggesting that the ESRist can make the distinction, because a structure is made important simply in virtue of the fact that it has been arrived at via an a posteriori method. This is surely untenable.

4.3. The Merrill/Lewis/Psillos Reply

This proposal was first suggested as a possible response to Putnam's model theoretic argument against realism by Merrill (1980). It was adopted by

Lewis (1983) and has been discussed in connection with Newman's objection by Psillos (1999). The key to the proposal is the suggestion that it is a contingent fact that some relations instantiated by the world are more important than others. The importance, or as proponents of this approach usually put it, "naturalness", of a relation is not a logical property of the relation, nor a property the relation somehow acquires via the method by which it is discovered, but a physical property. Proponents of this view would not, presumably, deny that there is a perfectly good sense in which objects in the domain of the world, D_W, instantiate every relation compatible with the cardinality of D_W. However, they would add that only some of these relations are natural relations. The idea is that the world isn't *just* a collection of objects that can be grouped howsoever we please, rather it is a collection of objects that also have preferred *natural* groupings. The world itself determines that some relations are more important than others and in this way comes pre-structured.

This is just the "natural kinds" doctrine (or something very like it) and so this response is only open to those ESRists who are prepared to buy into this doctrine (or something very like it).[12] However, if one does accept it then Newman's objection misses the point. Let's call a structure a structure of the world if its domain is D_W (the set of objects in the world). Let's call a structure a natural structure of the world if in addition the relations in it are (the sets corresponding to) natural relations. It is true that any structure whose domain has the same cardinality as D_W is isomorphic to some structure of the world, but it is certainly not true that any such structure is isomorphic to some natural structure of the world (and presumably science aims to discover not just any old structures of the world but only the natural structures).

[12] Worrall (personal communication) says that he takes natural kinds to be, by definition, the properties and relations that we refer to with the predicates of our best theories. This response to Newman's objection is also not open to the ESRist who accepts only this form of the natural kinds doctrine: saying that the predicates of our best theories refer to natural kinds does not limit the properties and relations to which these predicates can refer if any property or relation that is referred to by these predicates is, by definition, a natural kind.

In terms of the Ramsey-sentence approach to ESR the claim would be that the relations over which the quantifiers in the Ramsey-sentence range are restricted to the natural relations. It follows that theorem 1:

Theorem 1: The Ramsey-sentence of a theory A is true if and only if there is some sequence of relations, $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$..,. such that << D_O, D_T >, O_1, O_2..., $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$...> $\models$ A

must be replaced by:

Theorem 1'': The Ramsey-sentence of a theory A is true if and only if there is some sequence of *natural* relations, $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$..., such that << D_O, D_T >, O_1, O_2..., $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$...> $\models$ A

because the relations $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$... are now only in the scope of the existential quantifiers if they are natural. This blocks Ketland's proof of theorem 2 at the last step: although given a theory has some model, << D_1, D_2 >, $R_{1.1}$, $R_{1.2}$..., $R_{2.1}$, $R_{2.2}$..., $R_{3.1}$, $R_{3.2}$...>, that is empirically correct and T-cardinality correct we can construct a model of the theory of the form << D_O, D_T >, O_1, O_2..., $R_{2.1}$', $R_{2.2}$'..., $R_{3.1}$', $R_{3.2}$'...> this does not guarantee that the theory's Ramsey-sentence is true, as there is no guarantee that the relations $R_{2.1}$', $R_{2.2}$'... we have constructed are natural. As Merrill puts it:

> So long as we ignore any *intrinsic* structuring of the world, there is nothing to forbid us *imposing* a structure along any lines we chose. But if, as the realist surely must hold, the real world is a *structured* domain, then we are not free to ignore its intrinsic structuring in playing our model-theoretic tricks. (Merrill, 1980, p. 74, original emphasis)

Melia and Saatsi (2006) have objected to this response to Newman's objection as follows:

> the structural realist *shouldn't* take the second order quantifiers to be [sic] the natural properties if he wants his Ramseyfied theories to avoid the pessimistic meta-induction. One way in which new theories can overthrow old theories is by showing that properties that were once thought to be absolutely fundamental are in fact not. *Being green, being hot, being hydrogen* have all turned out to be *disjunctive* properties. Future developments in physics may show that *having a mass* is a disjunctive property. But the structural realist wants his Ramsey sentences to be *preserved* across theory change – they are supposed to capture something that is *constant* between theories, else the structural realist does little better than the full blown realist in dealing with the pessimistic meta-induction. If the intended domain of the second order quantifier is the set of natural properties, then the discovery that the properties postulated by the previous generation of theories are *not* natural will refute the structural realist's Ramsey sentence as much as it refutes the old realist theory. (Melia and Saatsi, 2006, p. 576, original emphasis)

However, it has been argued (in chapter 1) that the WESRist's response to the pessimistic induction is in any case unconvincing, so it seems that the structural realist should not be too worried if he is forced to abandon this response.

It might also be objected that the structural realist could have no warrant for the claim that the second-order variables in the Ramsey-sentence range over only natural properties and relations, because we can have no idea what the natural kinds are independently of our theories. However, the structural realist needn't claim that we do know what natural kinds are *independently* of our theories. He will claim that whatever reason we have to believe that our theories are true is also a good reason to think that the terms

in our theories refer to natural kinds and that the second-order variables in the Ramsey-sentence range over only natural properties and relations.[13]

As noted, Psillos (1999) considers this response to Newman's objection but argues that it is not available to the ESRist:

> in order for them [i.e. ESRists] to distinguish between natural and non-natural classes they have to admit that some non-structural knowledge is possible, viz. that some classes are natural, while others are not. (Psillos, 1999, p. 66)

As long as we do not go down Carnap's route and take naturalness to be a logical property of properties and relations then it is true that the knowledge that some property or relation is natural is not purely structural. However, neither the WESRist nor the SESRist claims that we have only purely structural knowledge (WESRists claim that our knowledge is restricted to claims constructed using logical and observational terms, SESRists claim that our knowledge is restricted to claims constructed using logical and

[13] He might also suggest, as Lewis does, that:

> It takes two to make a reference, and we will not find the constraint [on what properties and relations we refer to] if we look for it always on the wrong side of the relationship. Reference consists in part of what we do in language or thought when we refer, but in part it consists in eligibility of the referent. (Lewis, 1983, p. 371)

Lewis' suggestion is that there is some *feature of the world* that restricts the range of our variables for us. So it doesn't matter whether we *know* that our terms refer to natural kinds or not. It is just a fact that only certain properties and relations are eligible referents of terms, so, like it or not, our (second-order) variables can only range over certain properties and relations. However, this response is incoherent. It amounts to the claim that "there are some properties and relations (the unnatural ones) that lie outside the scope of our quantifiers" and this claim is obviously self-defeating. Suppose (for reductio) that it is true. Then there is something in the scope of the "some properties and relations" that lies outside the scope of the "some properties and relations". So the claim is false. Or, to put it another way, we clearly can refer to unnatural relations, even if we don't typically do so.

internal terms). However, Psillos' point is essentially unaffected by this consideration, because the "naturalness" of a property or relation is surely not an observable (or internal) property, so neither the WESRist nor the SESRist can consistently treat "naturalness" as a primitive second-order non-logical predicate, and if the predicate "naturalness" must itself be "Ramseyfied" this response will not work: cf. subsection 3.3.

5. Chapter Summary

It has been argued that none of the attempts that have been made to evade Newman's objection are successful. Consequently, Newman's objection remains a very serious problem for the ESRist. Of course, one cannot rule out the possibility that ESRist may in the future come up with a satisfactory reply, but in the absence of such a reply it seems that the sensible attitude toward his position is one of considerable scepticism.

Chapter 5

Ontic Structural Realism

The main proponents of OSR are French and Ladyman (see French, 1998, 1999, French and Ladyman, 2003a, 2003b and Ladyman, 1998, 2001). Related views have been suggested by Dipert (1997), Esfeld (2004) and Esfeld and Lam (forthcoming).[1] Ladyman (1998) puts forward OSR as an alternative to ESR. He characterises the latter as the view that "the objective world is composed of unobservable objects between which certain properties and relations obtain; but we can only *know* the properties and relations of these properties and relations, that is the *structure* of the objective world." (Ladyman, 1998, p. 412, original emphasis).[2] Roughly characterised, OSR is the view that the reason that we can only know the structure of the objective world is because there is nothing else to know; the world just *is* a structure.

OSR appears to be strictly stronger than ESR (it appears that it is just ESR with the additional claim that "the world is a structure") so it seems that it must inherit all the problems of ESR. However, this is not so. The OSRist does *not* simply take ESR wholesale and strengthen it: he (more or less explicitly) denies a (widely held) metaphysical presupposition that the ESRist (implicitly) accepts, i.e. that objects are ontologically basic and that in so far as structures exist at all their existence is dependent on the existence of the objects that constitute them. The OSRist reverses the order of ontological priority, arguing that "structure is ontologically basic"

[1] In fact, as will be seen in section 3.5, an effectively OSRist approach to the interpretation of space-time theories was proposed by Earman in 1989 (although he considers the approach only to reject it).

[2] This characterisation of ESR may give the impression that the ESRist holds that all knowledge of the world is purely structural, which is not the case. The WESRist maintains that our knowledge of the world consists of claims constructed using logical and observational terms while the SESRist claims maintains that our knowledge of the world consists of claims constructed using logical and internal terms.

(French and Ladyman, 2003a, p. 46) and that in so far as objects exist at all their existence is dependent on the existence of structures. This allows him to evade the most famous objection to ESR, i.e. Newman's objection. If we think that the world is at bottom a collection of objects (as does the ESRist) then it is trivial that these objects instantiate any given structure (consistent with their cardinality) so knowing that the world instantiates some given structure is knowing nothing more than the cardinality of the world. However, if we think that the world is at bottom a structure (as does the OSRist) then it is by no means trivial that the world instantiates (or rather, is isomorphic to) some given structure.[3]

This chapter is divided into three main sections. In section 1, three ways in which the claim that the world is a structure might be understood are suggested and discussed. In section 2, a potentially fatal objection to OSR (in any of its forms) is addressed. Finally, in section 3, the positive arguments adduced in favour of OSR are considered. The upshot of the discussion is that while (one form of) OSR is tenable, the arguments in favour of the doctrine are inconclusive.

1. Three Types of OSR

1.1. OSR1

As noted (in footnote 2) Ladyman's characterisation of ESR might suggest that it is the view that all we know of the world is its (mathematical) structure. The corresponding version of OSR would be that the world just is

[3] Worrall (personal communication) claims that a response to Newman's objection along these lines is available to a proponent of his brand of ESR, because he does not accept that objects are ontologically basic, but is agnostic on this issue. However, it is hard to be truly agnostic on this issue: standard formal semantics treats objects as ontologically basic, so, by accepting standard formal semantics, one is also implicitly accepting this metaphysical assumption. Moreover, no ESRist has responded to Newman's objection by rejecting the framework of standard formal semantics, which it obviously presupposes (at least not in print), so it seems reasonable to assume that most ESRists would accept this framework and therefore (implicitly) the metaphysical assumptions that objects are ontologically basic.

(or at least supervenes on) a mathematical structure. At first sight this claim seems highly implausible: how could the concrete physical world be (or supervene on) an abstract mathematical structure? However, many people find the materialist's claim that mental states supervene on physical states at least somewhat plausible, despite the fact that, at first sight, it too seems highly implausible. So it would be wrong to dismiss this version of OSR out of hand. Let's call it OSR1.

Consider the following structure:

<D, P>

where D = {a} and P = {a}. Clearly, no-one would suggest that this was the complete structure of the actual world: it is far too simple. However, it is the complete structure of a number of (very simple) possible worlds. For example, it is the complete structure of a possible world in which there is just one elementary particle, which has the property of being positively charged (and no other properties). It is also the complete structure of a possible world in which there is just one elementary particle, which has the property of having spin ½ (and no other properties). So a single structure can represent a variety of different worlds. So a world cannot be just a mathematical structure, as OSR1 claims. Worlds have non-structural properties, as is evidenced by the fact that the structurally identical worlds described above are different. The difference between the worlds is reflected in the fact that the property P represents a different physical property in each case.

It has been suggested to me that the OSR1ist might reply that this example begs the question, because his claim is that apparently basic physical properties such as "being charged" or "having spin ½" are in fact not at all basic, but are, rather, complex structural properties. Hence the possible worlds in which there is a single elementary particle with one of these properties are structurally far more complex than has been suggested here and are, moreover, distinct. However, the OSRist claims to be a scientific realist: he believes that our scientific theories reveal the structure of the world. But, according to our scientific theories, properties such as

"being charged" or "having spin ½" *are* basic (and structurally simple) properties. Physics does not suggest that the (apparently) fundamental physical properties of (apparently) fundamental particles are (or supervene on) a complex structure of more fundamental properties and relations. Hence, in so far as the OSR1ist wishes to remain a scientific realist, this response is not available to him.[4]

The lesson is that mathematical properties and relations are abstract with respect to physical properties and relations, in the sense that the same mathematical property or relation can be realised by a number of different physical properties or relations. It is precisely because of this abstractness that a purely mathematical structure has no physical content. This is an obvious but fundamental objection to OSR1. Indeed, it is so obvious and so fundamental that it might seem incredible to suppose that anyone could seriously have proposed OSR1. In fact, it may be that no-one has, but comments such as, "the structural dissolution of physical objects leads to a blurring of the line between the mathematical and the physical" (French and Ladyman, 2003a, p. 41) have led a number of critics of OSR to understand the doctrine this way and put forward this objection. For example, Cao objects to OSR on the grounds that, "no mathematical structure would have physical meaning without interpretation, which itself cannot be structural in nature, but rather, involves qualitative terms" (Cao, 2003, p. 59). Busch (2003, p. 220), Psillos (2004, p. 8) and Simons (2002, p. 38) also make (essentially) this point.[5]

In response to this criticism (as adduced by Cao [2003]) French and Ladyman (2003b) deny that they are committed to a form of OSR in

[4] In fact, in some Bohmian interpretations of quantum mechanics spin is not regarded as a fundamental property (see, Berkovitz, 2000). Clearly, however, a similar argument could be run using properties that these interpretations do take to be fundamental.

[5] In fact, Busch goes further: he claims that the only acceptable way of thinking of mathematical structures is as in re objects. However, if one adopts the in re approach to mathematical structures then one presupposes that physical systems are distinct from, and ontologically prior to, mathematical structures. In effect he argues that the OSR1ist is exactly wrong, in that the mathematical is ultimately reducible to the physical, not vice-versa.

which the structures are purely mathematical. They claim that the structures to which they refer are "physical structures" and that these differ from mathematical structures in that they "can be related – via partial isomorphisms in our framework – to the (physical) 'phenomena'. This is how 'physical content' enters. Less trivially, the mathematical can be distinguished from the physical in that the latter is also *causal*" (French and Ladyman, 2003b, p. 75, original emphasis).

What are we to make of this distinction between physical and mathematical structures? It is reminiscent of the distinction that Redhead (2001a, p. 74) and Frigg (2003, p. 46) draw between "concrete" (or "specific") and "abstract" (or "unspecific") structures. Their intention is to distinguish between structures in which the domain is a set of concrete physical objects and structures in which the domain is a set of abstract mathematical objects.[6] For example the following represents a concrete structure:

<{Brutus, Caesar}; {<Brutus, Caesar>}>

The structure whose domain contains the elements referred to by "Brutus" and "Caesar" (i.e. Brutus and Caesar) and which has a property consisting of a single ordered pair, which contains the elements Brutus and Caesar (in this order).

Abstract structures can then be taken as either in re or ante rem objects. In the former case an abstract structure will be construed as an isomorphism class of concrete structures (i.e. a class whose members are all and only those concrete structures that are isomorphic to some particular concrete structure, which is itself a member of the class, as, trivially, any structure is isomorphic to itself). In the latter case an abstract structure will

[6] Actually, this might not be Redhead's intention. As examples of concrete structures he gives us "a pile of bricks, timbers and slates, which are then 'fitted together' to make a house, or brush strokes which 'relate' to form a picture, or words which string together into meaningful sentences." (Redhead, 2001, p. 74) perhaps suggesting that he takes concrete structures to be actual physical systems (rather than sets – i.e. mathematical objects – whose elements are physical objects).

be viewed as the neo-Platonic form that all these concrete structures instantiate. We can represent an abstract structure (isomorphic to the above concrete structure) as follows:

<{a, b}; {<a, b>}>

as long as "a" and "b" are not taken to denote any (particular) physical object.

However, the OSRist can't intend the difference between physical and mathematical structures to be quite the same as the difference between concrete and abstract structures suggested by Frigg and Redhead or his position would collapse into something like conventional realism.[7] Given what has been said about the radical metaphysics implicit in OSR this may come as a surprise. How can OSR collapse into conventional realism when the OSRist maintains that structures are ontologically prior to objects and the conventional realist maintains that objects are ontological prior to structures? The reason is that, given this understanding of what a physical structure is, it is apparent that we immediately abandon the radical metaphysical claim of OSR, because we are back to the idea that structures are built from objects, i.e. that objects have ontological priority over structures.

To maintain the central claim of OSR, whilst still making a distinction between physical and mathematical structures, we need to take a physical structure to be one in which the properties and relations are physical, while the objects are taken as non-physical. The OSRist will argue that independently existing objects out of which these relations are built are a convenient fiction: physical "objects" exist only as places in the structure of the world, which is built out of relations. This is the line that French seems to be taking when he claims that objects play only "a heuristic role allowing for the introduction of the structures which then carry the

[7] Although not quite conventional realism: the conventional realist probably wouldn't want to say that the world *was* a physical structure, but at most that the world *displayed* a physical structure, relative to some choice of predicates.

ontological weight" (French, 1999, p. 204). So, rather than taking a physical structure to be built out of physical objects, such as Brutus and Caesar, we take it to built out of physical relations, such as "killed". The difference between a physical and mathematical structure is then that in a physical structure the properties and relations are physical properties and relations, whereas in a mathematical structure they are only abstract properties and relations. In general, a representation of a structure, e.g.

<{a, b}; {<a, b>}>

could be a representation of either a physical or a mathematical structure, depending on whether "{<a, b>}" denotes a physical or a mathematical relation in the context in which the representation appears.

However, there are still at least two possible readings of what the OSRist intends:

OSR2: The world is a structure built out of (physical) relations (but not properties).
OSR3: The world is a structure built out of (physical) properties and relations.

It might be noted that these conceptions of OSR do not quite do justice to the claim that it is *structure* that is ontologically basic: in these cases it seems that it is the properties and/or relations that are ontologically basic. However, this seems to be at least in keeping with the spirit, if not the letter, of OSR.

1.2. OSR2 and Metaphysics

In OSR2 the world is taken to be a structure built out of relations, but no properties. This approach seems to be implied by Ladyman's suggestion that "we abandon the attempt to interpret physical theory in terms of underlying objects and properties of which the world is made" and focus on "*structure* and *relations* directly" (Ladyman, 2001, p. 73, original emphasis). OSR2

implies that properties are not physically real. Esfeld (2004) and Esfeld and Lam (forthcoming) are explicitly committed to something like this approach (i.e. they deny the existence of properties) although they call their position "moderate structural realism" (on the grounds that they take objects and relations to be on a par, ontologically, rather than arguing that relations are ontologically prior to objects[8]).

What reason is there supposed to be for making this discrimination between properties and relations? Esfeld and Lam (forthcoming) suggest that the main reason is provided by a well-known argument, which Jackson states as follows:

> When physicists tell us about the properties they take to be fundamental, they tell us about what these properties *do*. This is no accident. We know about what things are like essentially through the way they impinge on us and on our measuring instruments. (Jackson, 1998, p. 23, original emphasis)

This suggests:

> the possibility that (i) there are two quite distinct intrinsic properties, P and P*, which are exactly alike in the causal relations they enter into, (ii) sometimes one is possessed and sometimes the other, and (iii) we mistakenly think that there is just one property because the difference does not make a difference (as the point is put in information theory). An obvious extension of this possibility leads to

[8] Their position appears to be almost deliberately circular on this point. They argue that "It makes no sense to assign an ontological priority to objects, because instead of having fundamental intrinsic properties, there are only the relations in which they stand. In other words, an object as such is nothing but that what [sic] bears the relations. As regards the relations, it makes no sense to attribute an ontological priority to them, for at least insofar as they exist in the physical world, they exist as relations between objects." (Esfeld and Lam, forthcoming, p. 5).

> the uncomfortable idea that we know next to nothing about the intrinsic nature of our world. (Jackson, 1998, pp. 23-24)

In fact there are cases where we know (something like) this has happened. For centuries it was thought that samples of jade were all essentially alike. However, we now know that there are two essentially different types of jade (there are two quite distinct chemical compounds, both of which are green, hard etc.). Before the discovery of the difference between these types of jade we were in a position somewhat like the one described in the quote above. Although the different compounds were never actually exactly alike in the causal relations into which they entered, they were alike as far as we could tell and so we did mistakenly think that they were essentially the same. Similarly, it was initially thought that the atoms of a single chemical element must all be alike, but we now know that elements can have different isotopes (so, for example, some carbon atoms have atomic mass 12, others 13 and others 14 etc.).

The argument purports to demonstrate that we know nothing about the properties of physical objects (and given the OSRist's desire to close the gap between epistemology and metaphysics, the denial that properties are real then becomes inevitable). However, it shows at best something considerably less than this. In effect, the argument rests on the claim that any one-place predicate that we deploy might actually refer to a disjunction of two or more quite distinct natural properties that are "exactly alike in the causal relations they enter into" (Jackson, 1998, p. 23) (rather as "being a carbon atom" turned out to be a disjunction of "being a carbon-12 atom", "being a carbon-13 atom", etc.).[9] But it hardly follows from this that we know *nothing* about the properties of objects in the world. True, we can never be sure that the predicates we deploy refer to the most fundamental natural properties in nature, but (the argument seems to concede) we can have a good reason to think that they refer to properties with at least a

[9] Clearly, the argument must intend to say that any one-place predicate that we deploy might actually refer to a disjunction of *natural* properties, because otherwise the claim is trivial: any property (with more than one instance) can be taken as a disjunction of properties if no restriction is placed on the naturalness of the latter.

derivative naturalness (in that they are disjunctions of the most fundamental natural properties); we might not be able to cut the world at all the joints, but this argument provides no reason to think that we ever cut at anywhere but a joint.

Yet even if it had been demonstrated that the predicates we deploy might well refer to properties that are not in any way natural this would still only indicate that we know nothing about the *natural* properties of physical objects. That would not preclude the possibility that we have knowledge of the unnatural properties (and even these are, after all, properties). For example, as long as people generally come (independently) to the same conclusion as to whether or not a given object has the property of "being jade" then that is surely a good reason to say that they know whether or not that object has that property, irrespective of whether or not the property of "being jade" has any naturalness at all. Indeed, if one does not accept the existence of natural kinds then this would be the only sort of knowledge of properties that would be possible.

1.3. OSR2 and Physics

Esfeld and Lam (forthcoming) present an argument for the claim that there are no properties in nature based on reflections about general relativity. This theory, they claim, entails that space-time points have relational but no monadic properties. Even if this is true, however, what it seems to support is not the claim that all objects lack monadic properties (i.e. the claim that there are no monadic properties), but only the weaker claim that some objects lack monadic properties.[10]

Esfeld (2004) also puts forward an argument for the claim that there are no properties in nature on the basis of the phenomenon of quantum entanglement. A bipartite quantum system is in an entangled state if and

[10] Which is not to say that this weaker claim is of no interest: it is generally assumed that all objects must have at least one monadic property (see, for example, Langton, 1998, p. 22).

only if it cannot be written as the product of two separate states.[11] For example, the following bipartite state (in which particle 1 has spin up and particle 2 has spin down) is *not* entangled:

$|up\rangle_1|down\rangle_2$

as it is the product of the separate states:

$|up\rangle_1$ and $|down\rangle_2$

The paradigmatic entangled bipartite pure state is the singlet state:

$(1/\sqrt{2})(|up\rangle_1|down\rangle_2 - |down\rangle_1|up\rangle_2)$

This represents a superposition of two states, one in which particle 1 has spin up and particle 2 has spin down and one in which particle 1 has spin down and particle 2 has spin up. Esfeld (2004) claims that consideration of such states supports the claim that there are no properties in nature, because entangled quantum particles appear to have relations to other quantum particles, but no monadic properties.

There are two problems with this claim. The first is that it is false: entangled particles still have state-independent monadic properties. An entangled electron, for example, still has mass and charge. Esfeld acknowledges this problem, but his response is unconvincing. He says that since quantum physics is essentially the theory of the state-dependent properties of objects and since it "is our basic physical theory, it would be desirable to derive state-independent properties within the formalism of quantum theory. The idea then is to get state-independent properties such as charge and mass on the basis of state-dependent properties" (Esfeld, 2004,

[11] I will speak indifferently of states, systems and subsystems being entangled. It is to be understood that a system is entangled if and only if it is in an entangled state and a subsystem is entangled if and only if it is part of an entangled system. Similarly I will speak indifferently of systems and states being bipartite. It is to be understood that a state is bipartite if and only if it is the state of a bipartite system.

p. 612). Whether or not this is especially desirable is unclear, but it is clear that there is no reason to believe that it can be done. In fact there are good reasons to suppose it cannot be done: Hamiltonian operators always depend on state-independent monadic properties, so it seems very unlikely that one could do quantum mechanics without taking such properties as basic.

The second problem with Esfeld's claim is that (according to quantum physics) quantum particles needn't always be in entangled states, and particles that are not in entangled states have state-dependent monadic properties. So the fact that particles can be in entangled states shows at most that quantum particles can sometimes lack state-dependent monadic properties. Again, Esfeld acknowledges this problem, but again his response is unconvincing. He says, "What has to be accounted for in quantum theory is not entanglement, but cases of the absence of entanglement, if there really are such cases (if not, it has to be explained why there appear to be such cases)" (Esfeld, 2004, p. 604). This is unconvincing because Esfeld does not explain why it is that such cases need to be accounted for, nor does he give any reason for supposing that such cases might actually be non-existent.

1.4. OSR2 and Rigid Designators

Another reason for wanting to deny the reality of properties might be the fear that permitting properties will allow objects back into our ontology (as independent existents) and this would undermine the main claim of OSR (although of course this fear would hardly count as a good reason for discriminating against properties, unless one was already convinced that the main claim of OSR is correct). For example, the property "being *the* Tony Blair" appears to be a property that necessarily only one particular object can instantiate: it seems that, in any world in which the property is instantiated, it must be instantiated by the same object.[12] In that case, however, that object cannot be ontologically secondary to the structures in which it appears (because if it was it would not [necessarily] be the same

[12] At least, that's apparently the way it appears to some people (e.g. Kripke, 1981). Personally, I don't have strong intuitions on this matter.

object in the different worlds [structures]). However, the OSRist has at least two possible responses to this argument that stop short of denying the reality of properties altogether.

Firstly, he can argue that apparent properties that seem to presuppose the existence of particular individuals ("being *the* Tony Blair", "being *this* table" etc.) are not really features of the world. This would not commit him to the view that no properties are really features of the world: he could seemingly still allow the reality of properties such as "being red" or "being a cat" that clearly do not presuppose the existence of particular individuals. However, it is still unclear what reason there could be to deny the reality of properties like "being *the* Tony Blair", unless one is already committed to the main tenet of OSR. After all, there appears to be an instance of the property "being *the* Tony Blair", so it is hard to believe that no such property exists.

Secondly, the OSRist can argue that although apparent properties that seem to presuppose the existence of particular individuals are genuine features of the world they do not really presuppose the existence of particular individuals. It is true that, in any given world, there can only be one instance of "being *the* Tony Blair" but it does not follow from this alone that the object that instantiates this property must be the same in every world: in any given world there can only be one instance of the property "being the tallest man in the world", but it does not follow from this that the object that instantiates this property must be the same in every world. Admittedly, it might seem strange to say that different objects could have the property "being *the* Tony Blair". This is because such properties appear to be disguised names, and names appear to be rigid designators (that is, they appear always to apply to the same object, even in different worlds).[13] OSR would require us to give up the idea that there are any rigid designators, because, according to OSR, objects take their identity from the structure in which they occur, so we cannot have (literally) the same object in two different worlds (structures). It may seem counterintuitive to deny the existence of rigid designators but the central tenet of OSR (that objects are

[13] See footnote 12.

ontologically secondary to properties and relations) is itself deeply counterintuitive, so many of its consequences are bound to seem likewise and pointing out that the counterintuitive central tenet of OSR has some counterintuitive consequences is unlikely to convince a proponent of OSR that he is wrong.

So there seems to be little motivation (even if one already accepts the main tenet of OSR) for discriminating between properties and relations in the way that OSR2 does. OSR2 is also potentially subject to the incoherence objection, which is discussed in section 2.[14]

1.5. OSR3

According to OSR3 the world is a structure built out of properties and relations. This approach is the version of OSR that is the closest to conventional scientific realism (or antirealism, if we take the properties and relations that constitute the world to be exclusively observable) but it still upholds the OSRist doctrine that objects are ontologically dependent on structures (or at least properties and relations) and not vice-versa. As such it is still potentially subject to the ubiquitous incoherence objection. However, there seems to be no other objections to this view that aren't also objections to conventional scientific realism (or antirealism) and this seems to be the most plausible form of OSR: unlike OSR1 it does not fall prey to the objection that a purely mathematical structure has no physical content and unlike OSR2 it does not make the seemingly unmotivated claim that properties (but not relations) have no reality.

---

[14] Moreover, whilst it might seem plausible to deny the reality of unobservable properties (such as "having spin ½" or "being an electron") it seems highly implausible to deny the reality of observable properties (such as "being red" or "being a cat"). This suggests a variant of OSR2, in which the world is taken to be a structure containing (both observable and unobservable) relations and some (exclusively observable) properties. However, this version of OSR2 would face most of the difficulties associated with the original version of OSR2 as well as the difficulties associated with any position that makes essential use of a sharp observable/unobservable distinction.

2. The Incoherence Objection

The two most common objections to OSR are that, (i) the world could not literally be a mathematical structure, because mathematical structures in themselves have no physical content and (ii) it is incoherent to claim that structures (or properties/relations) could be ontologically prior to objects because objects are conceptually prior to structures (and properties/relations). It has already been argued that the first of these objections undermines OSR1, but it clearly does not affect the other versions of OSR that are considered above. The second objection, which I call the incoherence objection, would apply to all forms of OSR, as it denies that the OSRist's main claim is coherent.

Busch puts forward the incoherence objection as follows, "the very idea of structure presupposes some elements that go together to make up that structure. A relation might take *anything* as its relata, but it always takes *something*" (Busch, 2003, p. 213, original emphasis). Busch's point is that nothing (and, in particular, not the world) can be, at bottom, a structure, because structures themselves aren't, at bottom, structures: they're built out of objects (at least according to the standard conception of them). Similarly, properties and relations are sets of (tuples of) objects (according to the standard [formal] conception of them). The objection has also been put forward by Psillos (2004, p. 7).

Chakravartty (2003) agrees that this "is a natural first response [to OSR]" (Chakravartty, 2003, p. 871) but argues that as, "OSR recommends that we *revise* our concepts in such a way as to view relations as ontologically subsistent" then "To argue against OSR on the basis of a violation of conceptual dependence [i.e. on the basis that it violates the notion that objects are prior to relations] is thus, it seems, question begging" (Chakravartty, 2003, p. 872, original emphasis). The incoherence objection undoubtedly is question-begging, but its proponents might argue that the question deserves to be begged, on the grounds that the conceptual revision that the OSRist demands is impossible to carry out.

However, as we shall, there is no reason to think that the revision is impossible to carry out. Taking a Quinean line one might argue more modestly that although the revision is not strictly impossible, it is a revision of a belief that is so central to our web of beliefs that one would need to be presented with very strong reasons to carry it out (cf. Quine, 1951, pp. 39-43); before one tries to replace a plank of Neurath's ship whilst at sea it is better to make sure that it is really rotten, especially if it is well below the waterline (cf. Neurath, 1932, p. 201 and Quine, 1960, pp. 3-4). Taking this approach one would then have to look at the OSRist's positive arguments and decide whether or not they provided good enough reasons to motivate the radical change they demand. This is, in effect, the approach that Chakravartty (2003) takes (he concludes that the OSRist's reasons are not sufficiently compelling to warrant their revolutionary demands).

Whilst this seems to be a sensible approach (it is essentially the approach that is adopted in the remainder of this chapter) a caveat is in order: OSR clearly requires some revision of our "common-sense" metaphysical intuitions (as embodied in standard formal semantics) but the revision that (the more plausible forms of) OSR demands is not as radical as it first appears. The claim that the existence of (physical) objects depends on the existence of structures might be read *either* as (i) the claim that the existence of physical objects depends on the existence of mathematical structures *or* as (ii) the claim that the existence of physical objects depends on the existence of structures built out of physical (properties and) relations. The former claim would be associated with OSR1, which, it has been argued, is implausible on independent grounds. Only the latter claim (or something like it) would be associated with either OSR2 or OSR3. However, the latter claim seems to be almost equivalent to the philosophically familiar "bundle theory" (which claims that objects are bundles of properties and/or relations) because, as noted, in the sort of structures involved in these cases it actually seems to be the properties and/or relations (rather than the structures) that are taken to be ontologically

basic.[15] Although in conflict with common-sense metaphysics and the standard formal construal of the relationship between objects and properties/relations the bundle theory is hardly radically revisionary from a philosophical point of view: it is one of the two main philosophical theories of objects available (the main – and usually favoured – alternative being the "substrata" or "bare particulars" theory, which takes objects to be a propertyless substrata in which properties inhere).[16] In the light of this OSR3 seems to amount to no more than conventional realism combined with an explicit commitment to the bundle theory of objects.

The conclusion of this section is that the more plausible forms of OSR (OSR2 and OSR3) do not demand such a radical conceptual revision as might at first be thought, so the incoherence objection is not as devastating as it might at first appear. However, the main claim of OSR undoubtedly demands some revision of our common-sense metaphysical intuitions (as embodied in formal semantics, which takes objects to be the fundamental units out of which properties, relations and structures are built, rather than vice-versa) and so far none of the arguments in favour of OSR that might motivate us to make such a revision have been discussed. These are the subject of the next section.[17]

---

[15]Compare this with French's (2006) passing observation that, "in the absence of further metaphysical explication of the notion of structure itself, it is not yet clear whether or not such an approach [i.e. OSR] collapses into another form of the well-known conception of objects as bundles of properties" (French, 2006, pp.10-11).

[16] In fact there are at least two forms of the bundle theory. The first takes properties as universals and so takes objects to be bundles of universals. The second takes properties to be collections of (individual) instances of properties (tropes) and take objects to be bundles of tropes. For present purposes, however, there is no need to distinguish these versions of the theory. The point is merely that the OSRist's claim that objects are ontologically secondary to properties/relations is not without precedent.

[17] Van Fraassen makes a related, but distinct objection to OSR, which runs as follows: "If God had decided not to create anything concrete, then that statement [i.e. the statement 'X is multiply instantiated', where X is a property] would have been false. Therefore, taking the contrapositive, if such a statement is true, then there exist concrete entities, therefore entities other than properties and relations" (van Fraassen, 2006, p. 294). This is supposed to tell against OSR, since the OSRist would agree that some starements of the form "X is multiply instaniated" are true, but wants to deny the existence of entities other than

3. Arguments in Favour of OSR

3.1. The Argument from Coherence/Parsimony

Esfeld and Lam (forthcoming) claim that there are two general philosophical arguments in favour of OSR: the argument from coherence and the argument from parsimony. The argument from coherence is that epistemology should cohere with metaphysics, i.e. that there shouldn't be a gap between epistemology and metaphysics. If this argument is a good one it seems to tell against a metaphysics that posits the existence of substrata that can only be described as "something, I know not what" (cf. Locke, 1690, book 2, chapter 23, paragraph 2, pp. 208-209) in which properties are supposed to inhere, and thus (indirectly) in favour of the bundle theory of objects (which, it has been claimed, is the essence of the more plausible forms of OSR). So far, however, the argument from coherence seems more like an assertion that an argument: why shouldn't there be a gap between epistemology and metaphysics? The basis for this claim seems to be the argument from parsimony. This is just an application of Occam's razor, and runs as follows. If there is more in our metaphysics than is required for it to cohere with our epistemology then we should trim off the excess. In particular (provided it coheres with our epistemology) a metaphysics that posits only properties and relations is to be preferred to one that posits properties, relations *and* substrata.

This is essentially the traditional central argument for the bundle theory of objects. Advocates of the substrata theory are likely to reply that although parsimony is, in general, a virtue, this is taking it too far, because we cannot properly make sense of a metaphysics without substrata in which properties inhere (or they might say, more modestly, that although such a

properties and relations. However, this seems unconvincing. If by "concrete entities" van Fraassen means "ontologically primitive objects" then the OSRist will deny the premise. On the other hand, if by "concrete entities" van Fraassen means objects in the sense of the OSRist (i.e. bundles of properties) then it does not follow from the fact that concrete entities exist that entities other than properties and relations exist.

metaphysics makes sense it leaves a worrying explanatory gap, because bundle theories don't explain why it is that properties are bundled together). Whether one thinks that substrata are a redundant metaphysical excess or that they are necessary to make our metaphysics intelligible (or at least fulfil a useful explanatory role) seems to depend on one's intuitions about what is intelligible (or what needs explaining). Consequently, although the argument from parsimony does lend some support to OSR, it is far from decisive.[18]

3.2. The Argument from Quantum Physics

In classical statistical mechanics there are four ways we can distribute two particles over two states (four possible microstates of the combined system):

[1] Particle 1 in state 1, particle 2 in state 1.
[2] Particle 1 in state 1, particle 2 in state 2.
[3] Particle 1 in state 2, particle 2 in state 1.
[4] Particle 1 in state 2, particle 2 in state 2.

Note that, assuming the particles are alike in all their intrinsic properties, states 2 and 3 are indistinguishable. There are thus three possible macrostates of the combined system:

[A] Both particles in state 1.
[B] One particle in state 1 and one particle in state 2.
[C] Both particles in state 2.

[18] It should be noted that when Esfeld and Lam put forward this argument they suggest that it tells against a metaphysics that posits the existence of monadic properties (not necessarily against one that admits substrata) on the grounds that such a metaphysics does not cohere with our epistemology in that we (allegedly) do not know the monadic properties of objects. It has already been argued that this discrimination against monadic properties (as opposed to relations) is unwarranted.

In the classical case, these macrostates are given the weighting 1:2:1 (Maxwell-Boltzmann statistics) as one would expect (assuming that the underlying microstates are equally weighted). In the quantum case, the weighting of the macrostates depends on the type of particles involved. If the particles are bosons then the macrostates occur in the ratio 1:1:1 (Bose-Einstein statistics). If the particles are fermions then states A and C are forbidden (by the Pauli exclusion principle) and the macrostates occur in the ratio 0:1:0 (Fermi-Dirac statistics).[19]

At first sight, what this seems to imply is that permuting bosons doesn't make any difference; that microstates 2 and 3 are in fact the very same state differently described. Thus labelling the particles, "particle 1" and "particle 2" in the first place is misleading. As Weyl put it:

> ...the possibility that one of the identical twins Mike and Ike is in the quantum state E_1 and the other in the quantum state E_2 does not include two differentiable cases which are permuted on permuting Mike and Ike; it is impossible for either of these individuals to retain his identity so that one of them will always be able to say "I'm Mike" and the other "I'm Ike". Even in principle one cannot demand an alibi of an electron! (Weyl, 1931, p. 241)

This view holds that electrons are objects that can be aggregated (so that it makes sense to say "there are two of them") but not counted (if it was possible to count the particles then it would be possible to pair them off/label them with the natural numbers, so it would make sense to say "this is electron 1"). This is undoubtedly counterintuitive. And this counterintuitiveness is reflected and reinforced by the fact that methods for

[19] Actually, this is somewhat misleading, because it is not really clear that one can say (prior to measurement) that quantum systems are ever in these macrostates. In particular, it is doubtful that they can be said to be in macrostate B. This point is discussed in more detail shortly.

dealing with such objects in standard set-theory are rather contrived (see French and Krause, 2006, chapter 6).[20]

However, French (2006) and French and Redhead (1988) have argued that the conclusion that quantum particles are non-individuals on the basis of quantum statistics is too hasty. In the quantum case there *are* four distinct possible quantum states underlying the three macrostates, but they are not:

[5] $|S_1>_1|S_1>_2$ (i.e. particle 1 in state 1 and particle 2 in state 1)

[6] $|S_1>_1|S_2>_2$

[7] $|S_2>_1|S_1>_2$

[8] $|S_2>_1|S_2>_2$

as one would expect from analogy with the classical case. Rather they are the states:

[9] $|S_1>_1|S_1>_2$

[20] However, Krause and French (1995) and French and Krause (2006, chapter 7) (following Krause [1992]) argue that a form of quasi-set theory provides a suitable formal framework for dealing with objects of this kind, so we should not be too perturbed by them. The essence of quasi-set theory is that, instead of having just one kind of atom (urelemente) as in standard set theory we have two: m-atoms and M-atoms. M-atoms are like the atoms of standard set theory, and are intended to represent individual objects (the sort of objects found in the macroscopic and classical realms). By contrast, the concept of identity does not apply to m-atoms, and these are intended to represent non-individual objects (the sort of objects found in the microscopic quantum realm). This means that expressions of the form "x = y" are not well formed if x and y denote m-atoms. On the other hand, there is an "indistinguishability" predicate (≡) in quasi-set theory which can apply to pairs of m-atoms, M-atoms and quasi-sets (quasi-sets being entities built out of m-atoms and M-atoms, much as standard sets are built out of the atoms of standard set theory). (Where both apply, the indistinguishability predicate coincides with the identity predicate, i.e. if x and y are not m-atoms, then x = y if and only if x ≡ y.) Because m-atoms can be indistinguishable without being identical it follows that they cannot be counted (see Krause and French, 1995, p. 206). Nonetheless, quasi-sets are taken to have quasi-cardinalities; the concept of a quasi-cardinal is taken as a primitive. This reflects the idea that quantum particles cannot be counted, but can form aggregates containing a definite number of particles.

[10] $(1/\sqrt{2})(|S_1>_1|S_2>_2 + |S_2>_1|S_1>_2)$

[11] $(1/\sqrt{2})(|S_1>_1|S_2>_2 - |S_2>_1|S_1>_2)$

[12] $|S_2>_1|S_2>_2$

States 9, 10 and 12 are symmetric (meaning that if the particle labels are exchanged throughout the states are unchanged). State 11 is antisymmetric. The reason quantum statistics seem unusual is because bosons are restricted to symmetric states and fermions are restricted to antisymmetric states.

This might be taken to suggest that quantum particles are individuals after all: there are two distinct quantum states that give rise to macrostate B, it is just that only one of them is ever accessible to a given quantum system. (However, it is not at all clear that one would describe states 10 and 11 as actually giving rise to macrostate B: the implications of this point will be considered shortly.) French and Redhead (1988) conclude that what this demonstrates is "not that quantum particles *must* be individuals but rather that it is *possible* for them to be individuals, despite the peculiarities of quantum statistics." (French and Redhead, 1988, pp.237-238, original emphasis). Hence two metaphysical pictures are in fact available: one in which quantum particles are individuals and one in which they are not.

It is from this point that the OSRist launches the argument for his doctrine on which he generally lays the most stress. It is claimed that it would be "an ersatz form of realism that recommends belief in the existence of entities that have such ambiguous metaphysical status" (Ladyman, 1998, p. 420). Therefore (so the argument goes) we should not take objects to be the fundamental building blocks of reality, but should "shift to a different ontological basis altogether" (Ladyman, 1998, p. 420): we should be realists about the existence of relations (and properties) but regard objects as having – at most – a derivative form of existence. The argument can be summarised as follows:

Premise: The question of whether quantum objects are individuals or non-individuals is not resolved by quantum theory.

Conclusion: We should abandon the underlying metaphysics that takes objects as ontologically basic and move to the OSRist's metaphysics in which objects are not regarded as ontologically basic.[21]

No one would regard this argument as wholly compelling, but perhaps some people would regard it as lending some support to OSR. In subsection 3.3 the premise of this argument will be bought into question. In subsection 3.4 the legitimacy of the form of the argument will be questioned.

3.3. Objections to the Premise of the Argument from Quantum Physics

It might be thought that one could resolve this underdetermination (in favour of the non-individuals picture) (and thereby undermine the premise of the argument) by considering entangled quantum states. As French and Redhead show (1988, pp. 240-241), two particles in an entangled state possess all the same monadic and relational properties (including space-time position) in the sense that, (i) the probability that the measurement of a property, P, on particle 1, p_1, has outcome x is equal to the probability that the measurement of P on particle 2, p_2, has outcome x, i.e.

$$\text{Prob}\,(P[p_1] = x) = \text{Prob}\,(P[p_2] = x)$$

for every property P and every value x, (ii) the probability that a measurement of a property P on particle 1 has outcome x, given that a measurement of a property Q on particle 2 has outcome y is equal to the

[21] The unresolved underdetermination is between two metaphysical pictures that both take objects as ontologically basic: one that takes objects to be individuals in which properties inhere (these are the traditional bare particulars or substrata, which will be called "individual substrata" here) and one that takes objects to be non-individuals in which properties inhere (these will be called "non-individual substrata" here). But note that the second of these is already somewhat non-conventional, because traditionally objects are regarded as individuals.

probability that a measurement of a property P on particle 2 has outcome x, given that a measurement of a property Q on particle 1 has outcome y, i.e.

$$\text{Prob}\,(P[p_1] = x \mid Q[p_2] = y) = \text{Prob}\,(P[p_2] = x \mid Q[p_1] = y)$$

for every property P, every property Q, every value x and every value y. In standard interpretations of quantum mechanics these probabilities tell us all there is to know: there are no facts about physical properties that the quantum mechanical description of the system omits (although, as their name suggests, this is not the case in hidden variable interpretations). Hence, French and Redhead claim, according to standard interpretations of quantum mechanics, the particles are indistinguishable.

Saunders (2003, pp. 293-294) argues that there is a legitimate sense in which entangled particles *are* discernible. The claim is that entangled particles are "weakly discernible": although they have exactly the same properties and stand in exactly the same relations to other objects and to each other, there is some relation, R, which each stands in to the other but which neither stands in to itself. Thus the formula $\forall x(Rp_1x \leftrightarrow Rp_2x)$ is false (it is falsified when $x = p_1$ or $x = p_2$). For example, consider the relation which x bears to y if and only if: x and y are parts of a system in a superposed state and in each component of the superposition x is in the opposite state to y. For particles in states 10 and 11, each particle bears this relation to the other, but not to itself. Hence Saunders would classify such particles as weakly discernible.

French (2006, p. 6) argues that the appeal to this type of relation is question-begging, because in order to determine that each particle stands in this relation to the other, but not to itself, one must presuppose that the particles can be individuated (i.e. that labelling the particles 1 and 2 is legitimate). But irrespective of the cogency of this reply, Saunders' argument shows at most only that the normal examples of indistinguishable quantum particles are poor ones: Saunders himself notes that although fermions will always be at least weakly discernible in his sense, two or more bosons may be in exactly the same state without any irreflexive relation of

the above sort holding between them. Two bosons may therefore be entirely indistinguishable.

The principle of the identity of indiscernibles is essentially the claim that if "two" things are indistinguishable (i.e. if they have *all* the same properties) then they are in fact identical.[22] If we wish to uphold the principle and the standard interpretation of quantum mechanics in this case, then we will, it seems, conclude from these considerations that quantum particles (in entangled states) are not individuals and are thus not subject to the principle (it makes no sense to say that they are either identical or not identical).[23] If we wish to uphold the principle of the identity of indiscernibles and the standard interpretation of quantum mechanics, the underdetermination is thus resolved (in favour of the non-individuals picture). But we can of course retain the view of particles as individuals by abandoning the principle (in which case they can be regarded as distinct despite the fact they are indistinguishable) or by adopting a hidden variables interpretation of quantum mechanics (in which case they will be regarded as having different [hidden] properties). So the question of the metaphysical nature of the particles (as individuals or non-individuals) remains underdetermined by the theory.[24]

Nonetheless, there are problems with the premise of the argument from quantum physics, and we do not have to consider entangled quantum systems to reveal them. Although it is true that there are two distinct

---

[22] The principle comes in a variety of forms. The strongest claims that it is not possible for two distinct objects to share all the same monadic properties. An intermediate version claims that it is not possible for two distinct objects to share all the same properties and relations, excluding space-time location. The weakest version claims that it not possible for two distinct objects to share all the same properties and relations. Only the weakest is under scrutiny here – the stronger versions are clearly violated by both quantum and classical particles.

[23] Obviously we would not conclude this if "being particle 1" and "being particle 2" are taken to be properties. Admitting such properties effectively presupposes that the particles are distinguishable individuals.

[24] However, Morganti (2004) argues that this isn't a very interesting case of underdetermination, because the principle of the identity of indiscernibles is highly dubious (so the obvious choice is to abandon it and retain the notion that particles are individuals).

quantum states that give rise to macrostate B, this does not appear really to reopen the door to the possibility of viewing quantum particles as individuals. This is because, (i) the two states are not the two that one would expect if the particles were individuals (they are states 10 and 11, not states 6 and 7) and (ii) the argument that quantum particles can be treated as individuals seems to rest on the claim that for any given quantum system (of the sort being discussed) there are four possible states, but not all of these are accessible. It is not clear that this claim is legitimate: it seems to be a thinly-veiled way of making the obviously illegitimate claim that for any given quantum system (of the sort being discussed) there are four possible states, but not all of these are possible. Surely what we should say here is that for a system of bosons there are three possible states, for a system of fermions one, and the one possible state for a system of fermions is not the same as any of the three possible states for a system of bosons. Thus there are four possible states in total, but not four for any given system. What this seems to support is the old-fashioned view (as put forward by Weyl) that a consideration of quantum statistics does tell strongly in favour of the view that quantum objects are non-individuals. If this is so then the premise of the OSRist's argument from quantum physics, i.e. the claim that the question of whether quantum objects are individuals or non-individuals is not resolved by quantum theory, is undermined.

3.4. Objections to the Form of the Argument from Quantum Physics

However, even supposing that there is an interesting underdetermination between a metaphysics that takes quantum particles as individuals and a metaphysics that takes quantum particles as non-individuals one may wonder how moving to OSR is supposed to remove this underdetermination (or, if it is not supposed to remove the underdetermination, what the advantage of moving to OSR is supposed to be), because the OSRist does not deny that objects exist, he merely denies that they are the most basic ontological units (he views objects as bundles of properties). It thus seems that in the OSRist framework one can still ask whether objects are individuals or non-individuals. Perhaps the point is that for the OSRist only

one answer is possible: objects are non-individuals. To see why this might be so, consider the structures <D = {a, b}, E = {a, b}, S_1 = {a}, S_2 = {b}> and <D = {a, b}, E = {a, b}, S_1 = {b}, S_2 = {a}> where D is the domain, E is the property of being an electron, S_1 is the property of being in state S_1 and S_2 is the property of being in state S_2. Consider first the case where we interpret these structures (as is most natural) using a metaphysics in which objects are ontologically primitive. There are two sub-possibilities: either we can use a metaphysics of individual substrata or we can use a metaphysics of non-individual substrata. In the first sub-case the situations represented differ because the individual substrata (a and b) take on different properties in the two cases. In the second sub-case, the situations do not differ because, as non-individual substrata lack individuality, giving them different names in the first place is misleading. Now consider what happens if we interpret these structures using a metaphysics in which objects are taken to be bundles of properties. In that case there is again no difference between the "two" systems represented here. The system represented in each case is a system consisting of two objects, one that is a bundle of the properties "being an electron" and "being in state S_1" and one that is a bundle of the properties "being an electron" and "being in state S_2". In this case the labels "a" and "b" can only be arbitrary names for these bundles, so swapping them around has no significance. It seems that OSR automatically implies that objects are non-individuals.[25] OSR thus resolves the underdetermination in quantum mechanics by deciding in favour of the non-individuals picture.

However, (assuming that the premise of the OSRist's argument is true, i.e. assuming that there is an interesting underdetermination between the picture of quantum objects as individuals and non-individuals) there would be, in all, three underdetermined options on the table:

[25] But note that if objects are taken to be bundles of tropes, and tropes are taken to be individuals then objects ought to be individuals as well. OSR seems to be more akin to a version of the bundle theory that takes objects to be bundles of universals (or non-individual tropes).

[1] Take indistinguishable objects as individuals with an underlying metaphysics in which objects are individual substrata in which properties inhere.
[2] Take indistinguishable objects as non-individuals with an underlying metaphysics in which objects are non-individual substrata in which properties inhere.
[3] Take indistinguishable objects as non-individuals with an underlying metaphysics in which objects are bundles of properties (i.e. adopt an OSRist metaphysics).

The OSRist's argument appears to run as follows. Suppose we ask, "Are objects substrata in which properties inhere or are they bundles of properties?". If we give the former answer then there is a further metaphysical question to answer: "Are objects individuals or non-individuals?" (because, after answering the first question in this way, there are still two underdetermined options on the table: options [1] and [2]). On the other hand, if one gives the latter answer then there are no further metaphysical questions to answer (because only one option remains on the table: option [3]). For those who dislike metaphysics, this may be a good enough reason to give the latter answer. As giving the latter answer is tantamount to adopting OSR, for those who dislike metaphysics this may be a good enough reason to adopt OSR (although, phrased this way, it sounds like a purely pragmatic reason).[26]

However, we don't have to start with the question "Are objects substrata in which properties inhere or are they bundles of properties?". We can instead start with the question, "Are objects individuals or non-individuals?". Suppose we do start with this question, then we might reason as follows. If we give the latter answer then there is a further metaphysical question to answer: "Are objects substrata in which properties inhere or are

[26] Of course, if one really dislikes metaphysics one might just try to adopt an agnostic attitude to all metaphysical issues. The problem with this is that it is hard to avoid (implicitly) making metaphysical assumptions. For example, as has already been noted, standard formal semantics effectively embodies the substrata view of objects, so whenever we use standard formal semantics we are implicitly assuming this view.

they bundles of properties?" (because, after answering the first question in this way, there are still two underdetermined options on the table: options [2] and [3]). On the other hand, if one gives the former answer then there are no further metaphysical questions to answer (because only one option remains on the table: option [1]). For those who dislike metaphysics, this may be a good enough reason to give the former answer. As giving the former answer is tantamount to adopting a metaphysics of individual substrata, for those who dislike metaphysics this may be a good enough reason to adopt a metaphysics of individual substrata. So, by changing the order of the questions, but using reasoning parallel to the OSRist's, we we end up not with OSR but with an underlying metaphysics of individual substrata. Thus the OSRist's own line of reasoning can equally well be used to support a position very different from OSR. The OSRist's reasoning here is *at best* inconclusive.[27, 28]

---

[27] It has also been claimed that the fundamental metaphysical character of quantum entities is underdetermined in quantum field theory, because the physics is equally compatible with "both the view of fields as substances whose properties are instantiated at space-time points (or regions) and the view of fields as nothing but properties of those space-time points (or regions)" (French and Ladyman, 2003a, p. 46). It might be noted that this underdetermination is no worse in quantum field theory than it is in any other field theory (e.g. classical electromagnetism). Moreover, it does not appear to support OSR: in this case choosing OSR appears to be simply choosing one of the two underdetermined options (the view that fields are collections of properties). It might be objected that this is not so, because in the second interpretation one still has objects as basic ontological units, because fields are taken to be properties of space-time points and space-time points are objects. However, whether or not one takes space-time points as objects depends on one's interpretation of one's space-time theory. It is true that if one is a thorough OSRist there is no underdetermination, because one will give an OSRist interpretation of both quantum field theory and one's space-time theory, but it is equally true that if one is a thorough "objectivist" there is no underdetermination, because one will interpret both quantum field theory and one's space-time theory in terms of objects.

[28] Note that if, as argued, there is not in fact an interesting underdetermination, because a consideration of quantum physics does strongly suggest that quantum objects are non-individuals then there are only two underdetermined metaphysical options on the table: [2] and [3]. In this case OSR is just one of two underdetermined options.

3.5. The Argument from Space-Time Theories

There are two main metaphysical interpretations of space-time theories: substantivalism and relationalism. Broadly speaking, substantivalism holds that space-time exists alongside, but independently of, matter, whereas relationalism holds that space-time is ontologically secondary to matter, in the sense that facts about space-time are reducible to facts about material objects.[29] In the modern context, the debate is about the interpretation of general relativity. The main argument in favour of substantivalism is that there is a solution of the field equations of general relativity (the de Sitter solution) in which space-time is empty, so general relativity itself suggests that space-time can exist without matter. More fundamentally, the very fact that models of the field equations of general relativity are constructed with a manifold of space-time points encourages us to regard space-time points as real, at least if we also find appealing Quine's contention that we should regard as real that which we quantify over (in theories that we believe to be true) (see Quine, 1939, 1948). Both of these arguments suggest that taking general relativity seriously requires us to adopt substantivalism.

However, adopting substantivalism (or at any rate manifold substantivalism, i.e. the view that space-time is a manifold of space-time points) seems to lead to a problem. Given some model of the field equations of general relativity <M, g, T> (where M is the manifold, g is the metric field tensor and T is the stress energy [matter] field tensor) one can generate infinitely many different models via the infinitely many possible diffeomorphisms of M onto itself (a diffeomorphism of M onto itself is a differentiable bijection whose domain and range are each the set of points in M): if <M, g, T> is a model then so is <M, d*g, d*T> for any diffeomorphism d (where, for every point p in M, $d^*g[p] = g[d^{-1}(p)]$ and $d^*T[p] = T[d^{-1}(p)]$). The metric and matter fields in these models are

[29] These are the two main positions, but not the only positions. Esfeld and Lam attribute to Descartes and Spinoza a view according to which space-time and matter "are ontologically identical and form the same substantival entity" (Esfeld and Lam, forthcoming, p. 16). This may be a forerunner of the position that holds that matter can be reduced to space-time, which is now known as geometrodynamics.

essentially unchanged, the models differ only with regard to which particular points in M underlie which parts of the fields. This difference is entirely indiscernible (as space-time points are intrinsically indiscernible), but if we commit ourselves to the existence of space-time points then it does nonetheless seem that these mathematically distinct models represent physically distinct universes (universes in which the space-time points underlying the fields are permuted); we seem to be committed to the view that there are infinitely many distinct but indistinguishable possible worlds. This seems ontologically profligate.

In fact, Earman and Norton (1987) argue, via the so-called "hole argument", that this ontological profligacy has a still worse consequence. They put forward a version of the argument that can be applied to a range of space-time theories. Earman summarises the argument in the context of general relativity as follows:

> Choose any general relativistic model <M, g, T> you like, except...suppose that M, g possess a Cauchy surface. This supposition implies that there is a global time function t: M $\rightarrow$ **R** such that t increases as one moves in the future direction along any timelike curve and such that the level surfaces of t are all Cauchy surfaces...choose a diffeomorphism d such that d = id [i.e. the identity function] for all $t \leq 0$ and $\neq$ id for $t > 0$ and such that there is a smooth join at t = 0. By general covariance, <M, d*g, d*T> is also a model, and M, d*g also possesses a Cauchy surface. By construction, this new model is identical to the first for all $t < 0$: $g|_{t \leq 0} = d^*g|_{t \leq 0}$ and $T|_{t \leq 0} = d^*T|_{t \leq 0}$. But if d is chosen properly, the models will differ for $t > 0$ (Earman, 1989, p. 179)

The conclusion is that if (as the substantivalist has it) <M, g, T> and <M, d*g, d*T> represent physically distinct possible worlds, then even the weakest sort of determinism fails in general relativity, because the worlds

are exactly alike up to t = 0 but different thereafter (although, as noted, the difference is entirely indiscernible). Earman and Norton argue that this is unacceptable, since although it may be that the universe is indeterministic it should not be possible to decide this on purely metaphysical grounds:

> If a metaphysics, which forces all our theories to be deterministic, is unacceptable, then equally a metaphysics, which automatically decides in favour of indeterminism, is also unacceptable. Determinism may fail, but if it fails it should fail for a reason of physics, not because of a commitment to substantival properties which can be eradicated without affecting the empirical consequences of the theory." (Earman and Norton, 1987, p. 524)

There are at least two problems with the hole argument. Firstly, as Earman (1989) concedes, the hole argument does not show that substantivalism automatically forces all (space-time) theories to be indeterministic, because the hole argument only works for space-time theories with particular properties (in particular, generally covariant field equations and a background space-time that has almost no intrinsic structure, i.e. almost no structure that is not imposed by the fields). So substantivalism is only unacceptable if a metaphysics that forces all theories *of this sort* to be indeterministic is.

Secondly, it is a very peculiar type of indeterminism that is forced on us, even in these cases, and it is arguably a harmless type of indeterminism, because it has no empirical consequences. It would certainly be problematic if a metaphysical theory stipulated that physical theories had to be indeterministic at the empirical level (if, in fact, some successful ones were not) but the form of indeterminism forced on us by substantivalism is consistent with a physical theory being deterministic at the empirical level. It is true that there can be two models of the field equations of general relativity, <M, g, T> and <M, d*g, d*T> that are alike up to t = 0 but differ for t > 0, and that according to the substantivalist these represent different possible worlds that are physically identical up to t = 0 but differ thereafter.

However, as noted, the supposed physical distinction between the worlds is not supposed to be observable, so the indeterminism does not surface at the empirical level (cf. Hoefer, 1996, p. 10). Unless one has a metaphysical prejudice in favour of determinism this kind of metaphysical indeterminism that (according to substantivalism) underlies the empirical determinism of general relativity will seem no more objectionable than the kind of metaphysical determinism that (according to Bohmian mechanics, or the many minds interpretation) underlies the empirical indeterminism of quantum mechanics (and this metaphysical determinism is rarely taken to be a flaw, much less a fatal flaw, of these approaches).

However, although bringing up the substantivalist's commitment to an innocuous form of indeterminism with respect to a certain class of theories seems to be something of a red herring, the substantivalist's ontological profligacy, i.e. his commitment to infinitely many distinct but indistinguishable physical possibilities seems troubling enough in itself. The claim has been made that OSR offers a third way that has all the advantages but none of the disadvantages of substantivalism about space-time (see, for example, Ladyman, 2001, p. 69). (Esfeld and Lam [forthcoming] make a similar claim on behalf of their moderate structural realism.) An OSRist interpretation of general relativity would imply that space-time points exist (so we can take general relativity seriously), but that they exist only as places in a model of the field equations of general relativity, not as objects traditionally conceived (so there is no difference between diffeomorphic models: it makes little sense to say that we can permute space-time points, i.e. put one in the place of another in a new model, if space-time points just are places in a structure: the most we could do would be to swap the names of the points). OSRists might well claim to be in good company here: something like this position seems to be implied by Einstein's observation that "Space-time does not claim existence on its own, but only as a structural quality of the field" (Einstein, 1954, p. 155).[30]

[30] Although Newton is usually regarded as the arch substantivalist there is some evidence that he too might also have held something like OSR with regard to space-time: "the parts of space derive their character from their positions, so that if any two could change their

Earman (1989) considers and rejects this approach because a problem seems to arise if there is more than one isomorphism between two models, W and W':

> If ψ_1: W $\rightarrow$ W' and ψ_2: W $\rightarrow$ W' are relevant isomorphisms...and if i is an individual of W, it follows that i is identical with $\psi_1(i)$ and with $\psi_2(i)$. And so by transitivity of identity, $\psi_1(i) = \psi_2(i)$, which gives a contradiction if $\psi_1(i)$ and $\psi_2(i)$ are distinct (Earman, 1989, pp. 198-199)

The problem is that, as the isomorphisms are supposed to tell us which objects in W are to be identified with which objects in W' (i.e. which places in W correspond to which places in W') then, as there are two isomorphisms between W and W' at least some of the objects in W are each to be identified with two distinct objects in W', which makes a mockery of the supposition that these objects are distinct in W'.

However, this does not appear to be fatal to the OSRist approach. Note that in the envisaged situation the function $\psi_1^{-1}\psi_2$ (where $\psi_1^{-1}\psi_2[x] = \psi_1^{-1}[\psi_2(x)]$) is a (non-trivial) automorphism on W (and $\psi_1\psi_2^{-1}$ is a [non-trivial] automorphism on W'). Hence the models W and W' have two or more equivalent places (i.e. there are two or more indistinguishable objects in each model). For example, consider the structures S = <{a, b}, {a, b}> and S' = <{c, d}, {c, d}>. There are two isomorphisms between these structures, f (where f[a] = c and f[b] = d) and g (where g[a] = d, g[b] = c). The functions $f^{-1}g$ (where $f^{-1}g[a] = b$ and $f^{-1}g[b] = a$) and fg^{-1} (where $fg^{-1}[c] = d$ and $fg^{-1}[d] = c$) are non-trivial automorphisms on S and S' respectively. It seems that what the OSRist should say in this case is that the existence of

---

positions, they would change their character at the same time and each would be converted numerically into the other. The parts of duration and space are only understood to be the same as they really are because of their mutual order and position; nor do they have any hint of individuality apart form that order and position which consequently cannot be altered." (Newton, ca 1670, p. 136).

these automorphisms implies that a and b are just arbitrary names for two indistinguishable places in S and c and d are just arbitrary names for two indistinguishable places in S'. Hence in this case there is no need to say whether a should be identified with c and b with d or vice-versa. That is, the OSRist should deny what Earman assumes that he maintains: that it must always be possible uniquely to decide which object in a given structure is to be identified with which in another (isomorphic) structure.

However, OSR is not the only way in which one can eat one's substantivalist cake without suffering from ontological indigestion. As Hoefer points out, the indigestion arises because of the "ascription of primitive identity to space-time points" (Hoefer, 1996, p. 11). As we have already seen, one can view objects as non-individuals (i.e. avoid ascribing them primitive identity) without going so far as to claim that objects must be completely reconceptualised as bundles of properties à la OSR. One can also take objects as non-individual substrata (in which properties inhere) that can be aggregated (so that it makes sense to say "there are two of them") but not counted (so that it makes no sense to say "this is number 1"). Viewing space-time points as objects of this sort has all the advantages but none of the disadvantages of substantivalism: it implies that space-time points exist (so we can take general relativity seriously), but that they are not individuals. (Thus, there is no difference between diffeomorphic models: it makes little sense to say that we can permute space-time points, i.e. put one in the place of another in a new model, if space-time points cannot be individuated one from another, even in principle.)

So, assuming that we are compelled by the arguments both for and against substantivalism, we are faced with a choice between (at least) two metaphysical approaches that accommodate these arguments. Both approaches agree that space-time points are non-individual objects. The difference between the approaches is that whereas one regards objects as a primitive ontological category, the other (OSR) regards objects as ultimately bundles of properties. As in the case of quantum physics, an OSRist interpretation is one of the viable metaphysical options, but it is not the only one. Thus the argument from space-time theories, like the argument from quantum physics, is inconclusive.

4. Chapter Summary

Three interpretations of OSR have been proposed. It has been argued that the first two of these are untenable. With respect to the third interpretation it has been argued that (i) it is not, in fact, a genuinely novel position, being essentially scientific realism (or antirealism) combined with an explicit commitment to the bundle theory of objects and that (ii) while the most common criticisms of OSR do not tell (decisively) against it the arguments in favour of it are also inconclusive.

Conclusion

In this thesis three principal forms of structural realism have been distinguished: WESR, SESR and OSR. The main conclusions that have been reached with regard to each are summarised below. WESR:

[1] The positive argument in favour of WESR, i.e. the claim that it can accommodate the no miracles argument and the pessimistic induction is unconvincing, because, (i) the no miracles argument is flawed, so it is no particular virtue of WESR that it can accommodate it and, (ii) it is not clear that WESR really can accommodate the pessimistic induction (see chapter 1).

[2] There are unresolved difficulties in drawing the observable/unobservable distinction (or an appropriate alternative distinction) in a way that is suitable for the WESRist's purposes (see chapter 2).

[3] The WESRist has not provided a convincing response to Newman's objection (see chapter 4).

SESR:

[1] Russell's semantic argument for SESR, and its descendents, is unconvincing, because it is based on the principle of acquaintance, a principle for which no substantial argument has been given and which has absurd consequences (see chapter 3 and appendix 2).[1]

[2] The SESRist has not provided a convincing response to Newman's objection (see chapter 4).

[1] As noted (in chapter 3) a variant of the semantic argument, based on a "principle of observation" works somewhat better as an argument for WESR than does the original argument as an argument for SESR. It works better in so far as the principle of observation, unlike the principle of acquaintance does not have absurd consequences (due to the different characteristics of the observational/theoretical and internal/external distinctions). However, there is still no particular warrant for the principle of observation, and the difficulties that face the WESRist (outlined above) would remain even if one gave some weight to the semantic argument for WESR.

OSR:

[1] Depending on how one interprets the doctrine, OSR is either (i) wholly untenable or (ii) conventional scientific realism (or antirealism) combined with the traditional metaphysical view that objects are bundles of properties. In the latter case the doctrine has some plausibility, but the arguments that have been adduced in favour of it are inconclusive.

It thus seems that WESR, SESR and OSR (if it is interpreted as a substantially novel position in the scientific realism debate) are all unattractive.

However, there remain a plethora of other positions in the scientific realism debate; which of these is to be preferred? To have any chance of satisfactorily arguing for an answer to this question would require several more theses. What I am going to do instead is state my opinion and roughly outline why I hold this opinion.

With some reservations, I am a conventional realist. Boyd characterises conventional scientific realism as the conjunction of four central claims:

> (i) "Theoretical terms" in scientific theories (i.e. nonobservational terms) should be thought of as putatively referring expressions; scientific theories should be interpreted "realistically".
> (ii) Scientific theories, interpreted realistically, are confirmable *and in fact often confirmed* as approximately true by ordinary scientific evidence interpreted in accordance with ordinary methodological standards.
> (iii) The historical progress of mature sciences is largely a matter of successively more accurate approximations to the truth about both observable and unobservable phenomena. Later theories typically build upon the (observational and theoretical) knowledge embodied in previous theories.

> (iv) The reality which scientific theories describe is largely independent of our thoughts or theoretical commitments. (Boyd, 1983, p. 195, original emphasis)

I do not wholeheartedly endorse all of these, but I do at least partially endorse each of them.

Claim (i) is a semantic claim. I would endorse the following variant on this claim:

There is no difference in kind between "theoretical" (or external) and "observational" (or internal) terms and a theory of reference should not treat them as different in kind. If we think of the latter as referring expressions then we should think of the former in the same way.

The reason for the rephrasing is due to the worry that reference might fail for *both* theoretical (or external) *and* observational (or internal) terms – i.e. the worry that reference might be a completely fictitious notion. In the face of the arguments of Quine (1960) and Putnam (1981) there seem to me to be two options: (i) abandon the notion of reference and with it the correspondence theory of truth and (ii) take reference as a primitive notion. However, in view of the usefulness of the correspondence theory of truth (in formal semantics), if pushed I probably would incline towards the latter route (with a slightly guilty conscience), so, if pushed, I probably would endorse Boyd's formulation of the conventional realist's semantic claim.

Claim (ii) is an epistemic claim. I would endorse the claim, as stated above. However, I would not endorse the claim that it is rational to believe that every mature (successful) theory is approximately true, a claim not put forward by Boyd here, but a claim that realists often do propose. My belief is that there is no difference in kind between "scientific" claims and "everyday" claims. Consider an everyday theory. Suppose Lara intends to catch the 4pm train from Brighton to London Victoria, which arrives in Victoria and 5pm. I have arranged to meet Lara at a cafe in the centre of town at 5.30pm. At 4.10pm I notice that I have missed a call from Lara on my mobile phone. I try to ring her back but can't get through. I hypothesise

that she was ringing to tell me that she missed her train (she often does). A prediction of this hypothesis is that she will not be at the cafe at 5.30pm. With nothing better to do, I go to meet her anyway. The prediction turns out to be correct. But, despite the fact that this hypothesis has scored a novel predictive success (it is a "mature" hypothesis) I would still be very wary of the claim that the hypothesis is (approximately) true. There could easily be another explanation for her ringing and not being at the cafe at 5.30pm (perhaps she was on the train and was ringing to tell me she had to go elsewhere first and would meet me at 6pm, perhaps she was ringing for a completely unrelated reason and would have been at the cafe but the bus from the train station to the cafe got stuck in traffic, etc.). Nonetheless, as evidence in favour of the hypothesis amasses (e.g. Lara rings me again and tells me that she did indeed miss her train[2]), I become more and more confident that it is (approximately) true.

The same attitude is, I think, appropriate towards scientific theories: a theory being mature (successful) does not in itself provide sufficient warrant for believing that it is approximately true. But it does provide some evidence for believing this. And the more evidence that amasses in favour of the theory the higher degree of belief it becomes reasonable to assign the theory. The realism I advocate is thus attenuated in two respects: it is both piecemeal and graduated. It is piecemeal in so far as it does not support the view that we should regard all successful (mature) scientific theories as approximately true. It is graduated in so far as it takes confirmation as a matter of degree. This latter point may seem to be uncontroversial. But realists often talk as if confirmation is an all or nothing matter (e.g. suggesting that we have warrant to believe that a theory is approximately true if and only it correctly predicts an empirical phenomena that it was not constructed to yield).

Claim (iii) is a historical claim (albeit one that is impossible to assess). I would endorse the following variant on this claim:

[2] Notice that this provides very strong confirmation of the theory despite the fact that it is not a prediction (novel or otherwise) of the theory.

There is such a thing as scientific progress. Even though we cannot say that all successful theories are (probably) approximately true, it is reasonable to assume that the proportion of approximately true theories that we hold is increasing.

This variant is proposed in light of the response to the pessimistic induction outlined in chapter 1, subsection 2.6.

Claim (iv) is an ontological claim, which I endorse.

Given that I reject the no miracles argument, one might ask why I am a conventional realist (in so far as I am). One reason is my suspicion that there is no way to draw a sharp and principled observable/unobservable distinction, which so many forms of antirealism (and, as we have seen, at least one form of non-conventional realism) depend on. Another (not entirely distinct) reason is my endorsement of Popper's claim that "science is common sense 'writ large'" (Popper, 1959, p. 22). This reason has been elaborated by Fine:

> I certainly trust the evidence of my senses, on the whole, with regard to the existence and features of everyday objects. And I have similar confidence in the system of "check, double-check, triple-check" of scientific investigation, as well as the other safeguards built into the institutions of science. So, if the scientists tell me that there really are molecules, and atoms, and ψ/J particles and, who knows, maybe even quarks, then so be it. I trust them and, thus, must accept that there really are such things, with their attendant properties and relations. Moreover, if the instrumentalist (or some other member of the species "non-realistica") comes along to say that these entities, and their attendants, are just fictions (or the like), then I see no more reason to believe him than to believe that *he is* a fiction made up (somehow) to do a job on me; which I do not believe. It seems, then, that I had better be a realist. One

> can summarize this homely and compelling line as follows: it is possible to accept the evidence of one's senses and to accept, *in the same way*, the confirmed results of science only for a realist; hence, I should be one (and so should you!). (Fine, 1984, pp. 1197-1198, original emphasis)

Appendix 1

Chakravartty's Semirealism

1. Chakravartty on Entity Realism

Chakravartty (1998, 2004) argues that the most reasonable form of entity realism and the most reasonable form of ESR are actually the same doctrine, which he calls "semirealism". Entity realism (see, for example, Hacking, 1982) is realism with respect to the existence of unobservable entities postulated by successful scientific theories, such as electrons, but agnosticism with regard to the truth of the theories themselves (and even the approximate truth of the theories, otherwise the position would collapse into conventional realism). The pessimistic induction suggests an objection to this view: there are a number of theoretical entities, such as ether and phlogiston, which have been postulated by successful scientific theories, but since entirely rejected. However, this objection is not considered further here. An objection to the view that *is* considered further is that it appears to be nonsense to say that one is a realist about electrons but one doubts that our theories about electrons are even approximately true, i.e. that one doubts that electrons are anything like our theories tell us: it is absurd to say that you believe in electrons but think that our theories about them might be so completely wrong that electrons might be dog-like mammals found in Australia, for example.[1]

On the other hand there must be some room for disagreement over what properties electrons have, otherwise all theoretical claims (other than existential claims) about electrons become analytically true: if, by definition, electrons have all the properties that our theories ascribe to them then by definition they have a charge of $1.60217646 \times 10^{-19}$ C and the claim that

[1] Moreover, as Psillos (1999, pp. 255-258) points out, our belief in theoretical entities is (according to many entity realists) supposed to be based on our ability to manipulate them but it is clear that a physicist only believes that he is manipulating electrons (for example) because he accepts that certain theories about electrons are (at least approximately) true.

"electrons have a charge of $1.60217646 \times 10^{-19}$ C" is as uncontestable as the claim that, "all bachelors are unmarried". This seems wrong. Suppose we discover that there are no particles exactly like the ones described by our theory of electrons but that there is a species of particle that is like the one described by our theory of electrons except that it has a charge of $1.60217645 \times 10^{-19}$ C. Surely, what we discover is not that there are no electrons but that our theory of electrons is (slightly) wrong: that the claim that "electrons have a charge of $1.60217646 \times 10^{-19}$ C" is false. Chakravartty (1998, p. 394) makes this point in a rather different way: he points out that it is possible to argue with the president of the flat earth society, which implies that his hypothesis is not *analytically* nonsense.

This suggests that objects have some properties essentially ("essentially" here is not used in a thick metaphysical sense, but in the sense of "by definition": in the sense in which it is used in the claim that bachelors are essentially unmarried) and others accidentally[2] and that whilst it is possible that objects do not have the accidental properties that our theories ascribe to them it is not possible that they do not have the essential properties ascribed to them[3] (although it is possible that the purported objects don't exist): anything lacking the essential properties of an electron (for example) just isn't an electron. If one accepts this, then the entity realist cannot be a *complete* agnostic about scientific theories. He must accept that our theories about electrons at least ascribe to electrons the right essential properties.

On the face of it, this is not much of a concession: it seems that the entity realist is really only agreeing to use words in the way that our theories (implicitly) define them (for example, he agrees to refrain from calling an object an electron unless it is a negatively charged micro-particle) and that he remains sceptical about the substantial claims that our theories make about these entities. However, this appearance is deceptive, because

[2] Or some variant of this view, perhaps the view that objects must (by definition) have a weighted most of the properties that our theories ascribe to them.

[3] Or that whilst it is possible that objects do not have all the properties that our theories ascribe to them, it is not possible that they do not have a weighted most of the properties that our theories ascribe to them.

although the entity realist is sceptical about our theories' attributions of accidental properties to entities he already agrees that our theories are right to make the claim that these entities exist. In itself each claim that the entity realist concedes to the conventional realist might be taken to be fairly vacuous, but taken together they are certainly substantial: the claim that electrons exist is nearly vacuous (*if* one believes that electrons might have or lack *any* property) and the claim that electrons must, by definition, have certain properties is completely vacuous (*if* one is not committed to the existence or non-existence of electrons) but the combination of these claims, the claim that electrons exist and (by definition) have certain properties is substantial.

This is essentially the brand of realism that Chakravartty calls entity realism: realism with regard to the existence of theoretical entities and with regard to the truth of the theories describing these entities, but only in so far as they describe their essential properties. However, Chakravartty gives the doctrine a twist. Rather than leaving the potentially problematic essential/accidental distinction a primitive he defines it as co-extensive with (or perhaps replaces it with) a distinction he draws between "detection properties" and "auxiliary properties":

> let us thus define *detection properties* as those upon which the causal regularities of our detectors depend, or in virtue of which these regularities are manifested. *Auxiliary properties*, then, are those associated with the object under consideration, but not essential (in the sense that we do not appeal to them) in establishing existence claims. Attributions of auxiliary properties function to supplement our descriptions, helping to fill out our conceptual pictures of objects under investigation. Theories enumerate both detection and auxiliary properties of entities, but only the former are tied to perceptual experience. (Chakravartty, 1998, pp. 394-395, original emphasis)

It will be argued shortly that the distinction between detection and auxiliary properties, as drawn here, cannot be made independently of the theories that we believe to be true and that this undermines Chakravartty's project, but first Chakravartty's approach to ESR will be examined.

2. Chakravartty on ESR

One might formulate a version of ESR in which the Ramsey-sentence is formed by Ramseyfying predicates referring to auxiliary properties, but this is not Chakravartty's approach to ESR. He proposes (what he calls) a form of ESR that is very different from those discussed in the main body of this thesis. He agrees with Worrall that what is preserved through theory change (and what we should be realists about) are structures, but he disagrees with the ESRist in two important respects: (i) he argues that preserved structures are not preserved partially reinterpreted but retain all of the same (minimal) interpretation, (ii) he does not think that the entirety of a theory's structure is typically (even approximately) preserved, but only the parts of a theory's structure involving detection properties.[4]

We may contrast his position with "normal" ESR as follows: while the normal ESRist believes in the Ramseyfied version of the sentence expressing the theory, Chakravartty believes in *part* of the unRamseyfied sentence expressing the theory. In particular, he believes in the part that deals with detection properties, but he does not believe in (even a Ramseyfied form of) the part that deals with auxiliary properties. This is very different to conventional ESR. However, given this understanding of

[4] Cf. "not *all* structures have to do with causal relations involving the detection properties of entities. Clearly we can imagine different kinds of structures, such as ones linking auxiliary properties. But for the SRist [structural realist], such flights of fancy are not particularly helpful, for not just any structure will do. SR requires *stable* structures—ones which are, in fact, likely to be preserved. Confidence in theoretical structures is achieved by means of their success in mirroring the phenomenal world, and structures which map phenomenal regularities are those composed of relations between detection properties." (Chakravartty, 1998, p. 400, original emphasis)

ESR, Chakravartty's claim that ESR is equivalent to (Chakravartty's understanding of) entity realism becomes plausible: both views advocate belief only in what our theories say about the essential (detection) properties of entities. This is the view Chakravartty calls "semirealism".

However, a caveat is in order. Although Chakravartty would deny that there are parts of retained structures that get completely reinterpreted he maintains only that each part of a retained structure keeps the same *minimal* interpretation. An example may clarify this matter. Recall the Fresnel/Maxwell equation discussed earlier:

$R/I = \tan(i-r)/\tan(i+r)$

As noted, Worrall argues that whereas both Fresnel and Maxwell kept the interpretation of the observables (i and r) the same (being the angles made by the incident and refracted beams with the normal to the boundary between the media) they offered completely different interpretations of the unobservables (I and R): Fresnel suggesting they represented amplitudes of vibration of the ether, Maxwell suggesting they represented amplitudes of vibration of a disembodied electromagnetic field. By contrast Chakravartty suggests that they offered the same *minimal* interpretation of all the terms in the equation: both minimally interpreted I and R as amplitudes of vibration. In what sense, if any, this can be regarded as a "minimal interpretation" is discussed in the next subsection.

3. An Objection to Semirealism

How can we tell whether a property is a detection property or an auxiliary property? According to Chakravartty:

> In order to distinguish these properties we must turn to the equations with which we attempt to capture phenomenal regularities, and ask: what do these mathematical relations *minimally demand*[?] We must consider not what possible

> metaphysical pictures are consistent with these equations, but rather what kinds of property attributions are *essential* to their satisfaction – i.e. consider not what is possible, but what is *required.* For only these properties are tied directly to the detections we employ so as to construct mappings of natural phenomena in terms of mathematical relations. (Chakravartty, 1998, p. 396, original emphasis)

In particular, Fresnel's equations:

> demand some kind of influence, propagated rectilinearly and resolvable into two components, oscillating at right angles to one another and to the direction of the propagation. The property or properties of light in virtue of which such influences are realized are detection properties. By way of contrast, consider an auxiliary property attributed to light by Fresnel's optics: here we find not merely amplitudes, but amplitudinal displacements in an elastic solid medium. (Chakravartty, 1998, p. 396)

However it is hard to see in what sense Fresnel's equations *demand* this interpretation. As pure mathematics they of course demand *no* interpretation. It would be reasonable to think that when Chakravartty asks what interpretation an equation demands, he means to ask what it demands in order to retain empirical adequacy. But the minimal interpretation required for Fresnel's equations to retain empirically adequacy is far less than the one suggested: we need only interpret I and R as the square roots of the intensities of the incident and reflected light beams respectively; empirical adequacy does not demand that we interpret I and R as amplitudes of vibration. Moreover, if Chakravartty does mean by "demand", "demand in order to retain empirical adequacy" then his position would collapse into instrumentalism (because, unlike the ESRist, he does not think that uninterpreted theoretical structure reveals anything about the world).

Chakravartty must therefore mean something else by "demand", but it is not clear what this might be.

Recall that Chakravartty defines detection properties as properties that are causally linked to our detectors (and auxiliary properties as properties that are not so linked). It seems that it must be impossible to distinguish these properties independently of the theories that we believe to be true: suppose Fresnel's theory was true, and that light is a vibration in an all pervading elastic solid medium (the ether), then such vibrations *are* causally linked to our detections of light intensity, as surely as light is causally linked to our detections of light intensity, *because they are light!* But in that case vibrations of the ether would be detection properties, by Chakravartty's criteria.

This points to a serious problem for semirealism: scientists very rarely speculate about properties that would be causally redundant, if the theory in question was true (an exception may be Newton's speculation that space is the sense organ of God) and so there are very few properties attributed to objects in scientific theories that can be deemed auxiliary (in Chakravartty's sense) before we know that the theory in question is false. In that case, however, the semirealist's advice to believe in what theories tell us about detection properties seems to come down to the following: believe completely in the theories that we believe to be true, and in other theories, in so far as they agree with the former.

Appendix 2

Theorem:

Given a domain with two classes of objects, A and B, and a language of second-order logic whose non-logical terms refer exclusively to objects in class A (or sets of objects in class A, or sets of tuples of objects in class A), it is impossible to create a definite description of an object in class B, if B contains more than one object.

A definite description of an object, $ix(\varphi)$, contains a formula, φ, with one free variable, x, that is satisfied iff the object is assigned to x.

To prove the theorem it suffices to show that as long as there are at least two objects, b_1 and b_2, in class B then any formula, φ, that has one free variable, x, would be satisfied when b_1 is assigned to x iff it would be satisfied when b_2 is assigned to x. This is equivalent to the claim that φ is satisfied by some g (such that $g[x] = b_1$) with some G iff φ is satisfied by some g' (such that $g'[x] = b_2$) with some G', where g and g' are functions assigning individuals to individual variables, and G and G' are functions assigning sets of appropriate tuples to predicate variables.

This claim follows from the claim that φ is satisfied by some g (such that $g[x] = b_1$) with some G iff φ is satisfied by g' with G', where for all y:

[1] If $g(y) = b_1$ then $g'(y) = b_2$ (so, in particular, $g'(x) = b_2$)
[2] If $g(y) = b_2$ then $g'(y) = b_1$
[3] Otherwise $g'(y) = g(y)$

and G'(X) is the set of tuples obtained from G(X) by replacing each occurrence of b_1 in each tuple in G(X) with an occurrence of b_2 and replacing each occurrence of b_2 in each tuple in G(X) with an occurrence of b_1. Note, that (given 1, 2 and 3) this implies that $\langle g'(y_1),\ldots g'(y_n)\rangle \in G'(X)$ iff $\langle g(y_1),\ldots g(y_n)\rangle \in G(X)$. This claim can be proved by mathematical

induction (in the proof it is always assumed that g' and G' are obtained from g and G as indicated).

For the basis case, we need to prove that for any atomic formula, ψ, ψ is satisfied by some g (such that $g[x] = b_1$) with some G iff ψ is satisfied by g' with G'. There are three cases to consider:

[1] $\psi = Xy_1,\ldots y_n$, where X is an n-place predicate variable:

(i) ψ is satisfied by some g (such that $g[x] = b_1$) with some G iff $\langle g(y_1),\ldots g(y_n)\rangle \in G(X)$.

(ii) This is the case iff $\langle g'(y_1),\ldots g'(y_n)\rangle \in G'(X)$ (by definition of G').

(iii) This is the case iff ψ is satisfied by g' with G'.

(iv) So ψ is satisfied by some g (such that $g[x] = b_1$) with some G iff ψ is satisfied by g' with G'.

[2] $\psi = Py_1,\ldots y_n$, where $g(y_i) \neq b_1$ or b_2 for any i and P is an n-place predicate:

(i) ψ is satisfied by some g (such that $g[x] = b_1$) with some G iff $\langle g(y_1),\ldots g(y_n)\rangle \in$ the extension of P.

(ii) This is the case iff $\langle g'(y_1),\ldots g'(y_n)\rangle \in$ the extension of P (because g' is such that $g'[y_i] = g[y_i]$ for every i, as $g[y_i] \neq b_1$ or b_2 for any i).

(iii) This is the case iff ψ is satisfied by g' with G'.

(iv) So ψ is satisfied by some g (such that $g[x] = b_1$) with some G iff ψ is satisfied by g' with G'.

[3] $\psi = Py_1,\ldots y_n$, where $g(y_i) = b_1$ or b_2 for some i and P is an n-place predicate:

(i) ψ is not satisfied by some (any) g (such that $g[x] = b_1$) with some (any) G and ψ is not satisfied by g' with G' (because g and g' both assign a member of class B to the y_i such that $g[y_i] = b_1$ or b_2, and by definition of the language no such object is in the extension of any predicate).

(ii) So ψ is satisfied by some g (such that $g[x] = b_1$) with some G iff ψ is satisfied by g' with G'.

Induction Hypothesis: α and β are formulae such that α is satisfied by some g (such that $g[x] = b_1$) with some G iff α is satisfied by g' with G' and β is satisfied by some g (such that $g[x] = b_1$) with some G iff β is satisfied by g' with G'.

As is well known, the logical connectives can all be defined using only "¬" and "&" and the quantifiers are interdefinable. So there are four inductive cases to consider:

[1] $\varphi = \neg\alpha$:

(i) φ is satisfied by some g (such that $g[x] = b_1$) with some G iff α is not satisfied by g with G (by the semantics of "¬").

(ii) This is the case iff α is not satisfied by g' with G' (by the induction hypothesis).

(iii) This is the case iff φ is satisfied by g' with G' (by the semantics of "¬").

(iv) So φ is satisfied by some g (such that $g[x] = b_1$) with some G iff φ is satisfied by g' with some G'.

[2] $\varphi = \alpha \,\&\, \beta$:

(i) φ is satisfied by some g (such that $g[x] = b_1$) with some G iff α is satisfied by g with G and β is satisfied by g with G (by the semantics of "&").

(ii) This is the case iff α is satisfied by g' with G' and β is satisfied by g' with G' (by the induction hypothesis).

(iii) This is the case iff φ is satisfied by g' with G' (by the semantics of "&").

(iv) So φ is satisfied by some g (such that $g[x] = b_1$) with some G iff φ is satisfied by g' with G'.

[3] $\varphi = \exists y\alpha$ ($y \neq x$, as x is free in φ):

(i) φ is satisfied by some g (such that $g[x] = b_1$) with some G iff α is satisfied by some y variant of g, g^y, with G (by the semantics of "$\exists y$").

(ii) If this is the case then α is satisfied by some g (such that $g[x] = b_1$) (i.e. g^y) with some G (from (i) and the fact that g is a y variant of itself).

(iii) This is the case iff α is satisfied by g' (the g' formed from g^y) with G' (by the induction hypothesis).

(iv) If this is the case then φ is satisfied by g' with G' (by the semantics of "$\exists y$" and the fact that g' is a y variant of itself).

(v) So if φ is satisfied by some g (such that $g[x] = b_1$) with some G then φ is satisfied by g' with G' (by (i) to (iv)).

(vi) φ is satisfied by g' with G' (where g' and G' are formed from some arbitrary g [such that $g(x) = b_1$] and G as indicated) iff α is satisfied by some y variant of g', g'^y, with G' (by the semantics of "$\exists y$").

(vii) This is the case iff α is satisfied by g (where, for all z, if $g'^y[z] = b_1$ then $g[z] = b_2$, if $g'^y[z] = b_2$ then $g[z] = b_1$ and otherwise $g[z] = g'^y[z]$) with G (the G from which the G' in (vi) is formed) (because, the g' formed from this g is the g'^y used in (vi) and the G' formed from this G is the G' used in (vi) so it follows from the induction hypothesis that α is satisfied by this g'^y with this G' iff α is satisfied by this g with this G).

(viii) If this is the case then α is satisfied by some g (such that $g[x] = b_1$) with some G.

(ix) If this is the case then φ is satisfied by some g (such that $g[x] = b_1$) with some G (by the semantics of "$\exists y$" and the fact that g is a y variant of itself).

(x) So if φ is satisfied by g' with G' then φ is satisfied by some g (such that $g[x] = b_1$) with some G (by (vi) to (ix)).

(xi) So φ is satisfied by some g (such that $g[x] = b_1$) with some G iff φ is satisfied by g' with G' (by (v) and (x)).

[4] $\varphi = \exists X\alpha$:

(i) φ is satisfied by some g (such that $g[x] = b_1$) with some G iff α is satisfied by g with some X variant of G, G^X (by the semantics of "$\exists X$").

(ii) If this is the case then α is satisfied by some g (such that $g[x] = b_1$) with some G (i.e. G^X) (from (i) and the fact that G^X is an X variant of itself).

(iii) This is the case iff α is satisfied by g' with G' (the G' formed from G^X) (by the induction hypothesis).

(iv) If this is the case then φ is satisfied by g' with G' (by the semantics of "$\exists X$" and the fact that G' is an X variant of itself).

(v) So if φ is satisfied by some g (such that $g[x] = b_1$) with some G then φ is satisfied by g' with G' (by (i) to (iv)).

(vi) φ is satisfied by g' with G' (where g' and G' are formed from some arbitrary g [such that $g(x) = b_1$] and G as indicated) iff α is satisfied by g', with some X variant of G', G'^X (by the semantics of "$\exists X$").

(vii) This is the case iff α is satisfied by g (the g used to form the g' mentioned in (vi)) with G (where G is such that G[X] is the set of tuples obtained by replacing every occurrence of b_1 in a tuple in $G'^X[X]$ with an occurrence of b_2 and every occurrence of b_2 in a tuple in $G'^X[X]$ with an occurrence of b_1) (because, the g' formed from this g is the g' used in (vi) and the G' formed from this G is the G'^X used in (vi) so it follows from the induction hypothesis that α is satisfied by this g' with this G'^X iff α is satisfied by this g with this G).

(viii) If this is the case then α is satisfied by some g (such that $g[x] = b_1$) with some G.

(ix) If this is the case then φ is satisfied by some g (such that $g[x] = b_1$) with some G (by the semantics of "$\exists X$" and the fact that G' is an X variant of itself).

(x) So if φ is satisfied by g' with G' then φ is satisfied by some g (such that $g[x] = b_1$) with some G (by (vi) to (ix)).

(xi) So φ is satisfied by some g (such that $g[x] = b_1$) with some G iff φ is satisfied by g' with G' (by (v) and (x)).

QED.

References

Balzer, W., C.U. Moulines and J.D. Sneed, 1987, *An Architectonic for Science: The Structuralist Program*, D. Reidel, Dordrecht.

Berkeley, G, 1710, *A Treatise Concerning the Principles of Human Knowledge*, reprinted 1998, Oxford University Press, Oxford.

Berkovitz, J, 2000, "The Nature of Causality in Quantum Phenomena", *Theoria*, 15(1), 87-122.

Boyd, R, 1983, "On the Current Status of Scientific Realism", reprinted in R. Boyd, P. Gasper and J.D. Trout (eds), 1991, *The Philosophy of Science*, MIT Press, Cambridge (Massachusetts), 195-222.

Bueno, O, 1999, "What is Structural Empiricism? Scientific Change in an Empiricist Setting", *Erkenntnis*, 50, 59-85.

Busch, J, 2003, "What Structures Could not be", *International Studies in the Philosophy of Science*, 17(3), 211-223.

Cao, T.Y, 2003, "Can we Dissolve Physical Entities into Mathematical Structures?", *Synthese*, 136, 57-71.

Carnap, R, 1956, "The Methodological Character of Theoretical Concepts" in H. Feigl and M. Scriven (eds), 1956, *Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, University of Minnesota Press, Minneapolis, 38-76.

Carnap, R, 1966, *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*, Basic Books, New York.

Carnap, R, 1967, *The Logical Structure of the World: Pseudoprobelms in Philosophy*, Routledge and Kegan Paul, London (translated by R.A. George).

Chakravartty, A, 1998, "Semirealism", *Studies in the History and Philosophy of Science*, 29(3), 391-408.

Chakravartty, A, 2003, "The Structuralist Conception of Objects", *Philosophy of Science*, 70, 867-878.

Chakravartty, A, 2004, "Structuralism as a Form of Scientific Realism", *International Studies in the Philosophy of Science*, 18(2&3), 151-171.

Chisholm, R, 1957, *Perceiving: A Philosophical Study*, Cornell University Press, Ithaca.

Churchland, P.M, 1982, "The Ontological Status of Observables: In Praise of the Superempirical Virtues" reprinted in P.M. Churchland and C.A. Hooker (eds), 1985, *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen*, University of Chicago Press, Chicago, 35-47.

Cruse, P, 2005, "Ramsey Sentences, Structural Realism and Trivial Realization", *Studies in the History and Philosophy of Science*, 36, 557-576.

Cruse, P. and D. Papineau, 2002, "Scientific Realism without Reference", in M. Marsonet (ed.), *The Problem of Realism*, Ashgate, Aldershot, 174-189.

Demopoulos, W. and M. Friedman, 1985, "Bertrand Russell's *The Analysis of Matter*: Its Historical Context and Contemporary Interest", *Philosophy of Science*, 52(4), 621-639.

Dipert, R.R, 1997, "The Mathematical Structure of the World: The World as Graph", *The Journal of Philosophy*, 94(7), 329-358.

Ducasse, C.J, 1942, "Moore's Refutation of Idealism" in P. Schlipp (ed.), 1942, *The Philosophy of G.E. Moore*, Northwestern University Press, Chicago.

Duhem, P, 1906, *The Aim and Structure of Physical Theory*, reprinted 1991, Princeton University Press, Princeton (translated by P.P. Wiener).

Earman, J, 1989, *World Enough and Space-Time: Absolute Versus Relational Theories of Space and Time*, MIT Press, Cambridge, Massachusetts.

Earman, J. and J. Norton, 1987, "What Price Space-Time Substantivalism? The Hole Story", *The British Journal for the Philosophy of Science*, 38, 515-525.

Einstein, A, 1954, *Relativity: The Special and the General Theory (15th Edition)*, Methuen, London (translated by R.W. Lawson).

Elsamahi, M, 2005, "A Critique of Localized Realism", *Philosophy of Science*, 72(5), 1350-1360.

Enderton, H.B, 2001, *A Mathematical Introduction to Logic (2nd Edition)*, Harcourt, San Diego.

Esfeld, M, 2004, "Quantum Entanglement and a Metaphysics of Relations" *Studies in History and Philosophy of Modern Physics*, 35, 601-617.

Esfeld, M. and V. Lam, forthcoming, "Moderate Structural Realism About Space-Time", *Synthese*. Currently available at http://philsci-archive.pitt.edu/archive/00002778/.

Fine, A, 1984, "The Natural Ontological Attitude" reprinted in M. Curd and J.A. Cover (eds), 1998, *Philosophy of Science: The Central Issues,* W.W. Norton, New York, 1186-1208.

Fitting, M. and R.L. Mendelsohn, 1998, *First-Order Modal Logic*, Kluwer, Dordrecht.

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

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

van Fraassen, B.C, 2006, "Structure: Its Shadow and Substance", *The British Journal for the Philosophy of Science*, 57, 275-307.

French, S, 1998, "On the Withering Away of Physical Objects" in E. Castellani (ed.), 1998, *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, Princeton University Press, Princeton, 93-113.

French, S, 1999, "Models and Mathematics in Physics: The Role of Group Theory" in J. Butterfield and C. Pagonis (eds), 1999, *From Philosophy to Physics*, Cambridge University Press, Cambridge, 187-207.

French, S, 2006, "Identity and Individuality in Quantum Theory", *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu.

French, S. and D. Krause, 2006, *Identity in Physics: A Historical, Philosophical, and Formal Analysis*, Clarendon Press, Oxford.

French, S. and J. Ladyman, 2003a, "Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure", *Synthese*, 136, 31-56.

French, S. and J. Ladyman, 2003b, "The Dissolution of Objects: Between Platonism and Phenomenalism", *Synthese*, 136, 73-77.

French, S. and M. Redhead, 1988, "Quantum Physics and the Identity of Indiscernibles", *The British Journal for the Philosophy of Science*, 39(2), 233-246.

Fresnel, A.J, 1818, "Memoire Couronné sur la Diffraction", reprinted in A.J. Fresnel, 1865, *Oeuvres Complètes, I*, Paris.

Fresnel, A.J, 1822, *De La Lumière*, reprinted in A.J. Fresnel, 1865, *Oeuvres Complètes, I*, Paris.

Frigg, R, 2003, *Re-presenting Scientific Representation*, Ph.D. Thesis, London School of Economics and Political Science.

Goodman, N, 2006, *Fact, Fiction and Forecast (4th edition)*, Harvard University Press, Cambridge, Massachusetts.

Hacking, I, 1982, "Experimentation and Scientific Realism" reprinted in M. Curd and J.A. Cover (eds), 1998, *Philosophy of Science: The Central Issues,* W.W. Norton, New York, 1153-1168.

Halbach, V, 1999, "Conservative Theories of Classical Truth", *Studia Logica*, 62, 353-370.

Hempel, C, 1966, *Philosophy of Natural Science*, Prentice-Hall, Princeton.

Hoefer, C, 1996, "The Metaphysics of Space-Time Substantivalism", *Journal of Philosophy*, 93, 5-27.

Howson, C, 2000, *Hume's Problem: Induction and the Justification of Belief*, Clarendon Press, Oxford.

Jackson, F, 1998, *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Clarendon Press, Oxford.

Ketland, J, 2004, "Empirical Adequacy and Ramsification", *British Journal for the Philosophy of Science*, 55, 287-300.

Kitcher, P, 1993, *The Advancement of Science*, Oxford University Press, Oxford.

Krause, D, 1992, "On a Quasi-Set Theory", *Notre Dame Journal of Formal Logic*, 33, 402-411.

Krause, D. and S. French, 1995, "A Formal Framework for Quantum Non-Individuality" *Synthese*, 102, 195-214.

Kripke, S, 1981, *Naming and Necessity (2nd Edition)*, Blackwell, Oxford.

Kuhn, T.S, 1996, *The Structure of Scientific Revolutions (3rd Edition)*, University of Chicago Press, Chicago.

Ladyman, J, 1998, "What is Structural Realism?" *Studies in the History and Philosophy of Science*, 29(3), 409-424.

Ladyman, J, 2001, "Science, Metaphysics and Structural Realism", *Philosophica*, 67, 57-76.

Langton, R, 1998, *Kantian Humility: Our Ignorance of Things in Themselves*, Oxford University Press, Oxford.

Laudan, L, 1981, "A Confutation of Convergent Realism" reprinted in M. Curd and J.A. Cover (eds), 1998, *Philosophy of Science: The Central Issues,* W.W. Norton, New York, 1114-1135.

Lipton, P, 2001, "Quests of a Realist" *Metascience*, 10(3), 347-353.

Lewis, D, 1970, "How to Define Theoretical Terms", *The Journal of Philosophy*, 67(13), 427-446.

Lewis, D, 1983, "New Work for a Theory of Universals", *Australasian Journal of Philosophy*, 61(4), 343-377.

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

Locke, J, 1690, *An Essay Concerning Human Understanding*, reprinted 1995, Prometheus Books, Amherst (New York).

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

Maxwell, G, 1962, "The Ontological Status of Theoretical Entities" reprinted in M. Curd and J.A. Cover (eds), 1998, *Philosophy of Science: The Central Issues,* W.W. Norton, New York, 1052-1063.

Maxwell, G, 1968, "Scientific Methodology and the Causal Theory of Perception" in I. Lakatos and A. Musgrave (eds), 1968, *Problems in the Philosophy of Science: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965, volume 3*, North-Holland Publishing, Amsterdam, 148-160.

Maxwell, G, 1970a, "Structural Realism and the Meaning of Theoretical Terms" in S. Winokur and M. Radner (eds), 1970, *Analyses of Theories and Methods of Physics and Psychology*, University of Minnesota Press, Minneapolis, 181-192.

Maxwell, G, 1970b, "Theories, Perception, and Structural Realism" in R. Colodny (ed.), 1970, *Nature and Function of Scientific Theories*, University of Pittsburgh Press, Pittsburgh, 3-34.

Melia, J. and J. Saatsi, 2006, "Ramseyfication and Theoretical Content", *British Journal for the Philosophy of Science*, 57, 561-585.

Merrill, G.H, 1980, "The Model-Theoretic Argument Against Realism" *Philosophy of Science*, 47, 69-81.

Morganti, M, 2004, "On the Preferability of Epistemic Structural Realism", *Synthese*, 142, 81-107.

Neurath, O, 1932, "Protocol Sentences" reprinted in A.J. Ayer (ed.), 1959, *Logical Positivism*, The Free Press, Glencoe (Illinois), 199-208.

Newman, M, 2005, "Ramsey Sentence Realism as an Answer to the Pessimistic Meta-Induction", *Philosophy of Science*, 72(5), 1373-1384.

Newman, M.H.A, 1928, "Mr. Russell's 'Causal Theory of Perception'" *Mind*, 37(146), 137-148.

Newton, I, 1962, "De Gravitatione et Aequipondio Fluidorum" in A.R. Hall and M.B. Hall (eds), 1962, *Unpublished Scientific Papers of Isaac Newton*, Cambridge University Press, Cambridge, 89-156.

Poincaré, H, 1903, *Science and Hypothesis*, reprinted in S. J. Gould (ed.), 2001, *The Value of Science: Essential Writings of Henri Poincaré*, Random house, New York.

Poincaré, H, 1905, *The Value of Science*, reprinted in S. J. Gould (ed.), 2001, *The Value of Science: Essential Writings of Henri Poincaré*, Random house, New York.

Popper, K. R, 1959, *The Logic of Scientific Discovery*, reprinted in 1980, Unwin Hyman, London.

Psillos, S, 1999, *Scientific Realism. How Science Tracks the Truth*, Routledge, London.

Psillos, S, 2001a, "Is Structural Realism Possible?", *Philosophy of Science*, 68(3), S13-S24.

Psillos, S, 2001b, "Quests of a Realist: Author's Response" *Metascience*, 10(3), 366-370.

Psillos, S, 2004, "*The* Structure, the *Whole* Structure and Nothing *but* the Structure?", *Proceedings of the Philosophy of Science Association, 19th Biennial Meeting - PSA2004: PSA 2004 Symposia*, http://philsci-archive.pitt.edu/view/confandvol/92.html.

Putnam, H, 1962, "What Theories are Not" in H. Putnam, 1975, *Mathematics, Matter and Method: Philosophical Papers, Volume I*, Cambridge University Press, Cambridge, 215-227.

Putnam, H, 1977, "Realism and Reason", *Proceedings and Addresses of the American Philosophical Association*, 50(6), 483-498.

Putnam, H, 1979, *Mathematics, Matter and Method: Philosophical Papers, volume 1 (2nd Edition)*, Cambridge University Press, Cambridge.

Putnam, H, 1981, *Reason, Truth and History*, Cambridge University Press, Cambridge.

Quine, W.V.O, 1939, "Designation and Existence", *Journal of Philosophy*, 36, 701-709.

Quine, W.V.O, 1948, "On What There Is", *Review of Metaphysics*, 2, 21-38.

Quine, W.V.O, 1951, "Main Trends in Recent Philosophy: Two Dogmas of Empiricism", *The Philosophical Review*, 60(1), 20-43.

Quine, W.V.O, 1960, *Word and Object*, MIT Press, Cambridge (Massachusetts).

Quine, W.V.O, 1968, "Comment on Maxwell's Paper" in I. Lakatos and A. Musgrave (eds), 1968, *Problems in the Philosophy of Science: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965, volume 3*, North-Holland Publishing, Amsterdam, 161-163.

Redhead, M, 2001a, "The Intelligibility of the Universe", in A. O'Hear (ed.), 2001, *Philosophy at the New Millennium: Royal Institute of Philosophy Supplement: 48,* Cambridge University Press, Cambridge, 73-90.

Redhead, M, 2001b, "Quests of a Realist", *Metascience*, 10(3), 341-347.

Russell, B, 1911, "Knowledge by Acquaintance and Knowledge by Description", reprinted in B. Russell, 1917, *Mysticism and Logic*, Routledge, London.

Russell, B, 1912, *The Problems of Philosophy*, Oxford University Press, Oxford.

Russell, B, 1914, "The Relation of Sense-data to Physics", reprinted in B. Russell, 1917, *Mysticism and Logic*, Routledge, London.

Russell, B, 1915, "The Ultimate Constituents of Matter", reprinted in B. Russell, 1917, *Mysticism and Logic*, Routledge, London.

Russell, B, 1919, *Introduction to Mathematical Philosophy*, George Allen & Unwin, London.

Russell, B, 1927, *The Analysis of Matter*, Kegan Paul, Trench, Trubner and co., London.

Russell, B, 1968, *The Autobiography of Bertrand Russell, II*, Allen and Unwin, London.

Russell, B, 1969, *The Autobiography of Bertrand Russell, III*, Simon & Schuster, New York.

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

Salmon, W, 1966, *The Foundations of Scientific Inference*, University of Pittsburgh Press, Pittsburgh.

Saunders, S, 2003, "Physics and Leibniz's Principles" in K. Brading and E. Castellani (eds), 2003, *Symmetries in Physics: Philosophical Reflections*, Cambridge University Press, Cambridge, 289-307.

Simons, P, 2002, "Candidate General Ontologies for Situating Quantum Field Theory" in M. Kuhlmann, H. Lyre and A. Wayne (eds) *Ontological Aspects of Quantum Field Theory*, World Scientific, Singapore.

Suppe, F, 1977, "The Search for Philosophic Understanding of Scientific Theories" in F. Suppe (ed.), 1977, *The Structure of Scientific Theories (2nd edition)*, University of Illinois Press, Urbana, 3-241.

Suppes, P, 1957, *Introduction to Logic*, D. Van Nostrand Company, New York.

Votsis, I, 2003, "Is Structure Not Enough?", *Philosophy of Science*, 70, 879-890.

Votsis, I, 2004, *The Epistemological Statue of Scientific Theories: An Investigation of the Structural Realist Account*, Ph.D. Thesis, London School of Economics and Political Science.

Votsis, I, 2005, "The Upward Path to Structural Realism", *Philosophy of Science*, 72(5), 1361-1372.

Weyl, H, 1931, *The Theory of Groups and Quantum Mechanics,* reprinted 1950, Courier Dover Publications, New York (translated by H. P. Robertson).

Whittaker, E.T, 1951, *A History of the Theories of Aether and Electricity: The Classical Theories*, Thomas Nelson, London.

Wittgenstein, L, 1953, *Philosophical Investigations (3rd Edition)*, translated by G.E.M. Anscombe, 1967, Blackwell, Oxford.

Worrall, J, 1989, "Structural Realism: The Best of Both Worlds?", *Dialectica*, 43, 99-124.

Worrall, J, 1991, "Feyerabend and the Facts" in Gonzalo Munévar (ed.), 1991, *Beyond Reason: Essays on the Philosophy of Paul Feyerabend*, 329-353.

Worrall, J, 1994, "How to Remain (Reasonably) Optimistic: Scientific Realism and the 'Luminiferous Ether'" *PSA 1994*, 1, 334-342.

Worrall, J, forthcoming, "Miracles, Pessimism and Scientific Realism"

Zahar, E.G, 1973, "Why did Einstein's Programme Supersede Lorentz's?" *British Journal for the Philosophy of Science*, 24, 95-123.

Zahar, E.G, 1996, *Henri Poincaré: Science and Philosophy*, Akademie Verlag, Berlin.

Zahar, E.G, 2001, *Poincaré's Philosophy: From Conventionalism to Phenomenology*, Open Court, Chicago.

Zahar, E.G, 2004, "Ramseyfication and Structural Realism" *Theoria: Revista de Teoria, Historia y Fundamentos de la Ciencia*, 19(49), 5-30.