Towards Structural Realism

ABSTRACT: In the debate over scientific realism, attention has been given recently to a realist position referred to as structural realism. In this essay, I offer a version of this position and indicate how it addresses two standard forms of underdetermination argument posed by the anti-realist.

1. Introduction
2. Scientific Realism and Underdetermination
3. Epistemic Underdetermination
   3.1. Epistemically Indistinguishable Theories: Examples
   3.2. Non-empirical Criteria
   3.3. Empirical Criteria
   3.4. Global Conventionalism
   3.5. Dynamical Symmetries and Essential Structure
4. Structural Realism
   4.1. Twistors and Fields: Informing the Semantic Component of Structural Realism
   4.2. What is Structure?
5. Semantic Underdetermination
   5.1. The Pessimistic Meta-Induction: Reliability and Warranted Belief
   5.2. Structural Realism and Reliabilism: An Outline

1. Introduction

In the debate over scientific realism, recent attention has been given to a position referred to as structural realism. Worrall (1994, 1989) describes a realism that takes the mathematical equations of certain theories as representing an underlying structure. This structure is to be distinguished from the ontology associated with a given theory T; with the natures of the objects T claims exists. The equations representing this structure are, in relevant contexts, retained across theory change, hence such a realism is robust under attack from anti-realists appealing to the pessimistic meta-induction. Moreover, it satisfies realist intuitions that the success of such theories can best be explained by inferring that they in some way refer. Hence, Worrall claims, it offers the best of both worlds: It agrees with anti-realists that traditional realists face the problem of retention of ontology across theory change, while at the same time upholding realist intuitions that something must in fact be retained in order to account for instances of theoretical success:[1]

> Structural realism encourages an optimistic induction from the history of theory-change in science, but an optimistic induction concerning the discovery of mathematical structure rather than individual ontology (Worrall 1994, pg. 336).

Ladyman (1998) has insightfully pointed out that Worrall's view is ambiguous as to whether it is an epistemological thesis, the central claim being that we should restrict belief to the structure of successful theories; or a metaphysical thesis, the central claim

[1]Variants of these views also appear in Stein (1989) and Redhead (1999), and in the natural kind realism of Carrier (1993).

being that we should interpret successful theories as, at most, referring to structure. He maintains that, taken as an epistemological thesis, structural realism fails, facing problems similar to earlier versions advocated by Russell and Maxwell. On the other hand, Psillos (1999) has claimed that Worrall's structural realism fails to make precise the underlying notion of structure; hence it fails as a metaphysical thesis as well.

In this paper, I will attempt to flesh out a version of structural realism that makes explicit what its metaphysical and epistemic commitments are. In Section 2, I will articulate what I take to be the key features of the debate over scientific realism and where structural realism fits in. In Sections 3 and 4, I will flesh out a metaphysical version of structural realism that builds on the work of Worrall, Ladyman, and others. Finally, in Section 5, I will attempt to assuage some of the epistemological concerns of Ladyman (1998) by laying out a program that seeks to provide structural realism with an epistemological leg to stand on.

2. Scientific Realism and Underdetermination

Following Earman (1993) and Horwich (1982b), I will identify two components of scientific realism: a semantic component and an epistemic component. The semantic component characterizes the realist's desire to read scientific theories literally. It maintains that the theoretical claims of certain theories refer directly and unambiguously to an autonomous body of facts the discovery of which is independent of methodology. In the words of Horwich:

> Theoretical claims] are not to be understood either as mere assertions of verifiability, as covert, complex reports on observation, or as meaningless devices for the systematization of data (Horwich 1982b, pg. 182).

The epistemic component characterizes the realist's contention that there can be good reason to believe the theoretical claims of certain theories. In short:

Semantic Component: The theoretical claims of certain theories are to be interpreted literally.

Epistemic Component: There are good reasons to believe the theoretical claims of certain theories.

In general, the task the scientific realist faces is to provide explicit accounts of both of these components. For the semantic component, what do theoretical claims amount to; in particular, how should they be read literally? For the epistemic component, what constitutes good reasons for belief?

This distinction, as I see it, is meant to follow the distinction between the two separate enterprises of interpretation and confirmation. For any theory T, the former enterprise endeavors to give an account of what the world would be like if T were true. The latter enterprise endeavors to give an account of the conditions under which we are justified in believing T's claims. The scientific realist must be able to give accounts of both endeavors. For example, under one construal, to be a semantic realist about classical electrodynamics (CED) is to accept the claims it makes about electrons, say, as literal descriptions of actual physical particles. To be an epistemic realist about CED is to claim that its predictive success, say, warrants our belief that the descriptions of electrons it gives are accurate descriptions of real physical entities.[2]

The anti-realist attempts to demonstrate that these components are incompatible: Realism with respect to one component precludes realism with respect to the other. Different variants of anti-realism present themselves as objections to either or both components. The debate thus divides into the following camps:

(a) Semantic realism/Epistemic realism.

(b) Semantic realism/Epistemic anti-realism.

(c) Semantic anti-realism/Epistemic realism.

(d) Semantic anti-realism/Epistemic anti-realism.

I take (a) to be the core position adopted by a scientific realist. Position (b) is adopted by van Fraassen's (1980) constructive empiricist form of anti-realism. Position (c) is adopted by conventionalists who contend that, while theoretical claims can in principle accrue evidential support, in practice they turn out to be semantic short-hand for empirical claims. Finally, I take (d) to be the core position of the instrumentalist.

Most anti-realist objections to (a) can be seen as taking the general form of an underdetermination argument:

(UA) For any version of semantic (epistemic) realism, there are theories T and T' such that if we are semantic (epistemic) realists about T and T', we cannot be epistemic (semantic) realists about T and T'.

In other words, semantic realism underdetermines belief; and epistemic realism underdetermines ontology. From UA, one can extract two forms of underdetermination argument: an epistemic form and a semantic form. Both have appeared in the literature in various guises. One form of epistemic underdetermination argument (EUA) attempts to undermine (a) with the notion of epistemic indistinguishability: If T and T' are distinct epistemically indistinguishable theories, then semantic realism forces us into epistemic anti-realism (T and T', read literally, make different, possibly conflicting, claims, with respect to which we cannot be epistemic realists). Anti-realists

[2]There are other ways to flesh out the semantic and epistemic components of a scientific realist construal of CED. An alternative semantic stance would be to read literally the claims CED makes about field quantities. An alternative epistemic stance would be to base epistemic warrant on unifying power (as opposed to predictive power): We are warranted to believe the claims CED makes because it unifies a wide range of diverse phenomena from electrostatics and magnetostatics.

employing arguments of this type march under the slogan, *Autonomy undermines credibility.* I shall look at this form of argument in Section 3.

One form of semantic underdetermination attempts to undermine (a) with what has been called the Pessimistic Meta-Induction (PMI):[3] Epistemic realism, it is claimed, forces us into semantic anti-realism. There are theories in the history of science whose claims, according to epistemic realists, were entirely justified at the time, and yet we cannot be semantic realists about them: They simply do not refer to real things. Anti-realists employing arguments of this type march under the slogan, *Credibility undermines autonomy.* I shall look briefly at this form of argument in Section 5.

The task, then, for the structural realist, as a scientific realist, is to show that UA is wrong; in particular, that both the EUA and the PMI can be avoided. In the following, I shall attempt to flesh out one direction in which this task might proceed. Briefly, I will take the semantic component of structural realism to be the claim that the structural features that are essential to the reliability of a theory T are to be read literally. I will suggest that these features may be identified as those that minimally determine the class of dynamically possible models of T. I will elaborate on this thesis in Sections 3.5 and 4.2.

The construal of warranted belief that I will suggest the structural realist adopt holds that a necessary condition for warranted belief is the existence of a reliable method of inquiry that produces the belief in question. Both the EUA and the PMI can be viewed as questioning the reliability of theoretical claims. It thus seems reasonable to adopt reliability as a criterion of warranted belief acceptable to both the realist and the anti-realist and thereby avoid charges of question-begging. In Section 5 I shall sketch an approach to demonstrating how structural realism coupled with such a reliabilist epistemology might avoid the PMI.

3. Epistemic Underdetermination

In this section, I begin with a formulation of the epistemic underdetermination argument (EUA) and indicate the options for the epistemic realist. My conclusions are (a) the EUA is a non-starter unless the realist and anti-realist agree on a criterion for warranted belief; and (b) if the realist concedes the criterion to the anti-realist in the form of some notion of empirical warrant, then there is a viable foil to EUA; namely, structural realism.

EUA claims that semantic realism undermines epistemic realism. It takes the following general form, where (ER) and (SR) are Epistemic Realism and Semantic Realism respectively and (EI) denotes what I shall refer to as the Epistemic Indistinguishability thesis:

(ER) Belief in some class C of theories is justified.

(EI) For any theory $T \in C$, there is a theory $T' \in C$ such that,

(i) Any reason to believe T is a reason to believe T' and *vice versa*;

[3]See Putnam (1978, pp. 24-25) for the original statement. Another form of semantic underdetermination argument is the modification of EUA that the conventionalist makes use of (see below footnote 4).

(ii) If T' and T are read literally, they make contradictory claims.

(SR) For all $T \in C$, T is to be read literally.

$\therefore$ ((SR) $\wedge$ (EI)) $\Rightarrow$ $\sim$(ER).

(The assumption driving the conditional in the conclusion evidently is that if belief in t entails belief in $\sim t$, then we should refrain from belief in t.) The options for the epistemic realist are then:

(a) Reject SR. Recall that this form of anti-realism is associated with conventionalism.[4]

(b) Reject EI. There are two ways to do this.
  (i) Reject (EIi) by claiming that, for two theories T and T', there are always reasons to prefer one theory over the other.
  (ii) Reject (EIii) by claiming that a literal construal of T and T' need not entail a contradiction.

Before considering these options it is instructive to examine three distinct types of purported epistemically indistinguishable theories that have appeared in the literature. The distinctions rest essentially on the means by which the relation of indistinguishability is cashed out. It may amount to a global isomorphism (Type 1, below), or something weaker.

3.1. Epistemically Indistinguishable Theories: Examples

Type (1): "Linguistically" Indistinguishable Theories

These are globally isomorphic, total theories of the world that differ by exchanging two terms in their theoretical vocabularies ("electron" and "proton", for example, for total theories encompassing atomic physics).

Horwich (1982a) makes examples of this type more precise in the following manner: Let S_1 and S_2 be (sets of) sentences in (possibly different) formal languages. They are potential notational variants if and only if every interpretation of S_1 is isomorphic to every interpretation of S_2 and vice versa. (In other words, if and only if there is a predicate mapping which transforms S_1 into S_2 and whose inverse transforms S_2 into S_1.) There are then two ways of viewing S_2:

---

[4]The conventionalist can equivalently be seen as turning the EUA into a semantic underdetermination argument. She will conclude ((ER) $\wedge$ (EI)) $\Rightarrow$ $\sim$(SR). In general, both the conventionalist and van Fraassen's anti-realist claim that scientific realism (the conjunction of (ER) and (SR)) is incompatible with (EI). They differ on the component of scientific realism they reject as a result.

(i) S_1 and S_2 are sentences in different languages L_1 and L_2. In this case, S_1 and S_2 can be viewed as formalisms of a single theory S_1, with S_2 being a translation of S_1 from L_1 into L_2. Consequently, no underdetermination is involved.

(ii) S_1 and S_2 are sentences in the same language L. In this case, S_2 is said to be homophonically related to S_1. S_2 is literally an interpretation in L of a different theory S_2 that, taken literally, makes different referential claims than S_1.

Underdetermination then occurs, it is claimed, when there are "total" (*viz.* epistemically ideal) theories S_1, S_2 with formulations S_1, S_2 such that S_1 and S_2 are potential notational variants, and S_2 is homophonically related to S_1.

Type (2): In-Principle Observationally Indistinguishable Theories
These are theories that differ in their theoretical structure and not their observational content. Such theories, by their very laws, rule out the existence of evidence that can distinguish between them.

The two examples below assume the distinction between spacetime symmetries and dynamical symmetries in the context of spacetime theories. Briefly, given a generally covariant rendering, all spacetime theories can be characterized as picking out a class of dynamically possible models: (M, A_i, P_i), where M is a differential manifold, the A_i's are absolute geometric object fields characterizing the fixed spacetime structure of M, the P_i's are dynamical geometric object fields characterizing the contents of the spacetime, and the A_i's and P_i's satisfy the field equations of T.[5] One can now distinguish between spacetime symmetries: automorphisms on M that leave the A_i's invariant (the group of all automorphisms ϕ such that $\phi * A_i = A_i$, where $\phi * A_i$ is the drag-along of A_i induced by ϕ), and dynamical symmetries: automorphisms on M that leave the set of physically possible P_i's invariant (the group of all automorphisms ψ such that, if (M, A_i, P_i) is a dynamically possible model of the theory, then so is (M, A_i, $\psi * P_i$)).

Example (a): $ND(0)$ and $ND(v)$, where $ND(0)$ is Newtonian dynamics with the velocity v of the center of mass of the universe equal to zero, and $ND(v)$ is the same theory with $v \neq 0$. The theoretical structure of both is Newtonian dynamics in Newtonian spacetime (the latter in order for absolute velocity to be well-defined). The spacetime symmetries of both examples thus are (Newt), and the dynamical symmetries of both examples are (Gal) (see below).

Example (b): NG_N and NG_C, where NG_N is Newtonian gravitation in Neo-Newtonian spacetime and NG_C is Newtonian gravitation in Newton-Cartan spacetime. The spacetime symmetries of NG_N are (Gal) and the dynamical symmetries are (Max) plus

[5] Such a distinction need not be taken to favor substantivalist interpretations of spacetime theories over relationalist ones. Relationalists may simply view absolute geometrical object fields as encoding spatiotemporal relations between physical objects, and dynamical geometric object fields as encoding (a subset of) non-spatiotemporal relations.

a particular gauge freedom associated with the gravitational potential. The spacetime symmetries of NG_C are (Max) and the dynamical symmetries are the same as NG_N.

Example (a) may be found in van Fraassen (1980), Friedman (1983), Laudan & Leplin (1991) and Earman (1993). (Newt) denotes the Newtonian group which is generated by coordinate transformations between rigid, Euclidean frames adapted to a unique flat affine connection. (Gal) denotes the Galilean group, generated by coordinate transformations between rigid, Euclidean frames adapted to a class of flat affine connections. A dynamically possible model of *ND* in Newtonian spacetime is given by (M, h^{ab}, t_a, u^a, ∇_a) where M is a differentiable manifold, h^{ab} is a (2, 0) symmetric tensor field of rank 3 (the spatial metric), t_a is a covariant vector field (the temporal metric), u^a is a timelike vector field (the integral curves of which define an absolute reference frame), and ∇_a is a smooth differential operator associated with a flat affine connection.

Example (b) is considered by Friedman (1983) and Earman (1993). (Max) denotes the Maxwell group, generated by coordinate transformations between rigid, non-rotating Euclidean frames. Bain (2005) describes the dynamically possible models of NG_N and NG_C and justifies the assignment of symmetries given above (in the case of a particular version of NG_C).

Type (3): In-Practice Observationally Indistinguishable Theories

These are theories that differ in practice in their observational content, but not in principle. Evidence that can distinguish between such theories is not ruled out by said theories; what is ruled out is our physical access to such evidence.

Example: Robertson-Walker models (M, g_{ab}, T^{ab}) of general relativity in which M is simply connected versus models in which M is multiply connected. While it is theoretically possible for evidence to exist that distinguishes between such models, no physically possible observer restricted to analyzing evidence in her past light-cone can ever hope to gain access to such evidence. The spacetime symmetries of these examples consist of *Diff*(M): the group of diffeomorphisms on M. The dynamical symmetries will depend on the particular distribution of matter as encoded by the stress-energy tensor T^{ab}.[6]

[6]Identifying the spacetime symmetries of general relativity (GR) with *Diff*(M) is done for convenience only, and follows simply from the claim that the only absolute object in general relativity is the manifold M. I grant that this is debatable; however, nothing in the subsequent discussion is dependent on it; nor on the claim that the dynamical symmetries of GR depend on the symmetries of T^{ab}. I concede the latter on behalf of those who consider these examples to be legitimate cases of indistinguishable theories. For my own purposes, it would be easier to identify the dynamical symmetries of GR with *Diff*(M) insofar as I believe such examples are not distinct theories, but rather are distinct models of a single theory (namely, GR). In general, for any given dynamically possible model (M, g_{ab}, T^{ab}) of GR (and not just Robertson-Walker models), it is the case that (M, $\psi^* g_{ab}$, $\psi^* T^{ab}$) is also a dynamically possible model, where $\psi \in Diff(M)$.

This example is given originally in Glymour (1977). Malament (1977) extends such examples to include other global properties of M including temporal orientability, spatial orientability, orientability, inextendability, noncompactness, the existence of closed future-directed causal curves, the existence of a global time function, and global hyperbolicity. Earman (1993) and Kelly (1996) consider such examples to be examples of epistemic indistinguishability. In the above, g_{ab} is a Robertson-Walker metric and T^{ab} is an appropriate stress-energy tensor, both of which satisfy the Einstein equations.

3.2. Non-empirical criteria

One can adopt Option (bi) of Section 3 above and deny (EIi) for all three of the above types of purported epistemically indistinguishable theories if one subscribes to a non-empirical criterion of epistemic warrant by means of which the class C of epistemically privileged theories in ER is characterized. Examples of such criteria include simplicity, explanatory power, unifying power, etc. Fine (1986) and Kukla (1994) have charged that forms of inference based on such traits beg the question for the realist in so far as there is no justification for such forms that will satisfy an anti-realist who licenses belief only in inferences based on empirical data (i.e., inferences from known empirical data to claims about as yet unknown empirical data). Notably, the explanatory defense of realism succumbs to this critique. The general form for what has become known as the Miracle Argument (MA, hereafter) is the following:

(MA)

(1) If explanatory power is a reason for belief and realism is the best explanation of the predictive success of science, then we should be realists.

(2) Explanatory power is a reason for belief.

(3) Realism is the best explanation of the predictive success of science.

∴ We should be realists.

Putnam's (1975) original formulation contends that it would be a miracle if successful scientific theories were not approximately true. As Kukla (1996) observes, this entails that the truth of theories is the only explanation of their success. The weaker form above contends merely that realism is the best explanation of success.[7] The rule of inference expressed by premises (1) and (2) is Inference to the Best Explanation (IBE, hereafter). Since the anti-realist will only license belief for inferences based on empirical data, these premises will be unacceptable (regardless of the acceptability of (3)).

Note that a realist might respond in one of two ways. She might dispute the claim that epistemic criteria must be empirical in nature, and attempt to argue for the

[7]Leplin (1997) argues for the stronger claim that truth is the only explanation of success. I shall not review the arguments here since I shall be entertaining the anti-realist's overarching concern that premise (2) is unacceptable.

inclusion of non-empirical criteria in general, and explanatory virtues in particular, in determinations of belief-worthiness. Or she might agree (implicitly perhaps) with the anti-realist that epistemic virtue should be based solely on empirical criteria, and attempt to ground IBE on the latter. Psillos' (1999) defense of MA appears to take this latter route. Since Psillos has been a vocal critic of structural realism, it will be instructive to examine the argument that underlies his own scientific realist sympathies. I will claim that it is not quite adequate. However, I believe that the reliabilist criterion of epistemic warrant that Psillos adopts is ultimately to be preferred, and, in Section 5, will indicate how it could be employed by the structural realist to provide her with an epistemic leg to stand on.

Psillos claims that the Miracle Argument can be viewed as a rule-circular justification of IBE . Rather than begging the question over what counts as a criterion of epistemic warrant, Psillos sees the Miracle Argument as accepting an empirical criterion; namely, that we are warranted in believing the claims produced by a method of inference just when that method is, as a matter of fact, reliable. This externalist criterion of justification is to be made distinct from the following internalist criterion: We are warranted in believing the claims produced by a method of inference just when we are justified in believing that the method is reliable (Psillos 1999, pp. 84-85). Psillos then characterizes the Miracle Argument as implicitly defending the reliability of IBE in a rule-circular manner, which, under an externalist view, is benign and not vicious. I think the attempt does not quite hit the mark, as I shall now argue.

Psillos' (No) Miracles Argument (NMA) takes the form (Psillos 1999, pg. 78, my reconstruction):

(NMA)

(1) The best explanation for the success of scientific methodology is that the background theories it relies on are approximately true.

(2) Scientific methodology is successful.

====================== [IBE]

∴ The background theories on which successful scientific methodology is based are approximately true.

Psillos describes this abductive argument in the following terms:

> NMA is a philosophical argument which aims to defend the reliability of scientific methodology in producing approximately true theories and hypotheses. Its strength, however, rests on a more concrete type of explanatory reasoning which occurs all the time in science (Psillos 1999, pg. 79).

This more concrete type of explanatory reasoning just is IBE. The claim is that theories in science are the products of IBE: a theory amounts to the best explanation for the observed data. Again, according to Psillos,

> NMA is not just a generalisation over scientists' abductive inferences. Although itself an instance of the method that scientists employ, NMA aims at a broader target: to defend the thesis that Inference

> to the Best Explanation, or abduction (that is, a type of inferential method), is reliable (Psillos 1999, pg. 79).

NMA employs a "meta-IBE"; what Psillos calls a "second-order" IBE, in contrast to the "first-order" IBE's that produce scientific theories.[8] Psillos now characterizes NMA in the following manner:

> ... by means of a meta-IBE, [NMA] concludes that the background theories are approximately true. Since these approximately true theories have been typically arrived at by first-order IBEs, this information together with the conclusion of the meta-IBE entail that IBE is reliable. So, the truth of the conclusion of NMA is (part of) a sufficient condition for accepting that IBE is reliable (Psillos 1999, pg. 83).

The reasoning of Psillos' realist (PR) may be reconstructed in the following manner:

(PR)

(a) The best explanation for the approximate truth of background theories is that the method that produced them is reliable.

(b) These background theories are approximately true.

(c) These background theories are the products of IBEs.

======================== [IBE]

∴ IBE is reliable.

According to Psillos, this is an instance of a "rule-circular" argument in the sense that its conclusion states a property of a method of inference, and this method is implicitly employed in the derivation of the conclusion. In particular, the claim that IBE is reliable is based in part on a premise (b) that itself is the conclusion of argument (namely, NMA) that employs IBE. Psillos contrasts this type of circularity, which he claims is benign, with the more vicious circularity exhibited by a "premise-circular" argument, in which the conclusion is explicitly stated as a premise.

Hence the anti-realist objects to NMA on charges that it begs the question over whether IBE is a reliable method of inference. Psillos's realist responds with PR, which demonstrates to the anti-realist, in a benignly rule-circular manner, the reliability of IBE. Should the anti-realist be convinced?

For a rule-circular argument to establish its conclusion, we should require that its premises be true. The truth of premise (b) in PR is established by recourse to NMA, hence depends on the truth of NMA's premises. Note that an internalist might require, in addition to the truth of the premises, that the reliability of the rule be established as well. Hence, in order for NMA to establish its conclusion, an internalist would have to assume the conclusion of PR. This would make PR explicitly premise-circular and thus suspect. However, on an externalist account, we do not have to establish the reliability of a rule of inference in order to use it (Psillos 1999, pg. 83). Hence premise (b) in PR

[8] Fine (1984, pg. 84) makes a similar distinction.

passes externalist muster to the extent that the premises in NMA are acceptable (and whether or not they are has nothing to do with the issue of the reliability of IBE).

Consider, now, premise (c). Its truth certainly cannot be assumed outright. One way to establish its truth is to argue that IBE is a foundational rule of inference -- the basis on which all forms of inductive inference (and, in particular, those used in science) are built. But there are problems with this approach. First, it is questionable whether the reliability of foundational rules can be established by means of rule-circular arguments. As Salmon (1966, pp. 12-17) points out, rule-circular arguments are only effective in the space of some prior plausibility for the inference rule in question. They cannot be used to justify an inference rule "from scratch". If they could, then any rule of inference could be established by an appropriate rule-circular argument, even intuitively bad ones. For instance, the rule of affirming the consequent could be justified by means of the following rule-circular argument: If affirming the consequent is reliable, then coal is black; Coal is black; Therefore, affirming the consequent is reliable. (On the other hand, rule-circular arguments can be effective in establishing rules that already have some initial plausibility; but whether or not IBE has initial plausibility in the context of scientific inferences just is the point at issue between realists and anti-realists.)

In any event, it is highly questionable whether IBE should be treated as a foundational rule, as Day and Kincaid (1994) have argued. Doing so minimally requires providing an explicit account of what counts as an explanation, and what counts as the best such explanation. Such accounts will be controversial, to say the least. Primarily for this reason, Day and Kincaid suggest treating IBE as a contextual rule of inference; they allow that its use is justified, but only in specific contexts in which the criteria of use have been agreed on beforehand.

If IBE is not foundational, then it is hard to see how premise (c) is justified. At the least, it behooves Psillos' realist to explain why IBE is more privileged than other accounts of theory identification (i.e., accounts based on alternative methods of inference such as demonstrative induction, enumerative induction, eliminative induction, Bayesian inference, or logical reliability, to name a few).

Hence, on an internalist account, PR fails, degenerating into a vicious premise-circular argument. On an externalist account, PR leaves open ("begs", if you will) the question of whether IBE is fundamental in scientific inquiry.

3.3. Empirical criteria

For an anti-realist who licenses belief only in inferences based on empirical evidence, the epistemic indistinguishability thesis becomes an empirical indistinguishability thesis and a distinction between empirical claims and theoretical claims must be made.[9] One way

---

[9]For the purposes of this section, I shall assume the distinction is unproblematic. It can, but does not have to be, based on an observable/unobservable split in the vocabulary of the language in which the theory is expressed. In general, it suffices to allow that theoretical claims are expressed in a language that out-strips the language in which evidence statements are made. If such a syntactic distinction is objectionable, allow a distinction in kind between unobserved observables and in-principle-unobservable unobservables, the latter being the subject of theoretical claims. The basis of the anti-realist's

to effectively engage the anti-realist is to adopt her criterion of warrant and then deny (EIi) by claiming that there are legitimate forms of inference from evidence that uniquely pick out a given theory. Examples include Bayesian updating (while the selection of priors is notoriously subjective, convergence theorems of the Gaifman/Snir type provide some measure of objectivity), as well as inferences based on demonstrative induction and eliminative induction. Advocates of such forms of inference might charge the epistemic anti-realist who subscribes to EUA with a naive hypothetico-deductivist account of confirmation in which evidence supports a theory just when it is entailed by the theory.[10]

Realists of this stripe may deny (EIi) for examples (1), (2a) and (3) above by claiming that the pairs of theories under consideration are not really distinct theories. For example (1), the intuition is clear. For (2a), a realist may claim that $ND(0)$ and $ND(v)$ share the same ideologies, ontologies and dynamics with respect to the nature of spacetime and motion (Earman 1993, pg. 31) (*i.e.*, they have the same spacetime symmetries and the same dynamical symmetries), hence at most, they are distinct models of a single theory; namely, Newtonian dynamics in Newtonian spacetime. For (3), a realist may claim that examples of observationally indistinguishable spacetimes are again different models of a single theory, general relativity; hence, at most, what they motivate is skepticism with respect to global characteristics of spacetime as described by general relativity. They do not indicate that there are epistemically indistinguishable alternatives to general relativity.

However, it is not clear how example (2b) can be handled in this manner. To the extent that the realist claims that $ND(0)$ and $ND(v)$ are the same theory on the basis of shared ideological and ontological commitments with respect to the nature of spacetime (viz., same spacetime symmetries), she must also claim that NG_N and NG_C are different theories with different ideological and ontological commitments. In particular, their commitments diverge over whether spacetime is flat or curved.

Furthermore, it is not apparent that there are any general principles on which a demonstrative and/or eliminative induction can be based to pick one over the other. In the context of relativistic theories of gravity, there are arguments that rule out non-metric theories. In particular, Schiff's conjecture claims that evidence in support of the weak equivalence principle (*WEP*) (*viz.*, the world line of a test body in a gravitational field is independent of its composition and internal structure) is evidence in support of metric theories. However, such arguments cannot be employed to pick between NG_N and NG_C since neither is a metric theory (although NG_C does satisfy *WEP*).

3.3. Global Conventionalism

---

underdetermination argument is simply that inferences to the former are warranted whereas inferences to the latter are not.

[10] Norton (1994, 1993) indicates how demonstrative induction was used to warrant belief in Planck's quantum hypothesis. A recent argument of Steven Weinberg's amounts to a demonstrative induction for effective local quantum field theory (Bain 1998). Eliminative induction has been championed by Earman (1992, Chap. 7) and Kitcher (1993, Chap. 7). However, to be effective against EUA it must be demonstrated that it leads to a unique hypotheses and not just a reduction of the space of contenders.

Finally, one can deny (EIi) for type (1) hypotheses while maintaining that this type constitutes the only non-trivial case of purported epistemically indistinguishable hypotheses. This is Horwich's (1986, 1982a) strategy. He claims that the choice between type (1) theories is not underdetermined; rather it is always decided uniquely by the manner in which we have decided a priori to express our beliefs. This latter decision is purely conventional and results in what Horwich refers to as global conventionalism.[11]

According to Horwich, the manner in which we express our beliefs is given succinctly by,

$$R(S_i) \Rightarrow S_i \ , \qquad (*)$$

where S_i is an interpreted set of sentences expressing some total theory T_i and $R(S_i)$ is its Ramsey sentence.[12] Horwich considers $(*)$ to be a reference fixing definition for all terms occurring in S_i. For any term t, $(*)$ fixes its reference by fixing the structural role t plays in S_i (Horwich considers $(*)$ an explication of the thesis that use determines meaning). Suppose S_1 and S_2 are incompatible (sets of) sentences in the sense that they are homophonically related potential notational variants (*i.e.*, S_2 is obtained from S_1 by a predicate mapping). If S_1 is the way we have chosen to make explicit our beliefs, then, since $S_2 \Rightarrow \sim S_1$ and $S_2 \Rightarrow R(S_1)$, by $(*)$, we know a priori that S_2 is false. Thus Horwich claims that global conventionalism, *i.e.*, the thesis that the way we choose to make explicit our beliefs is conventional, avoids EUA. Insofar as semantic realism requires that a literal construal of a theory entails a literal construal of the elements that comprise its models and insofar as, by literal construal, is meant, at the most general level, autonomy from human capacities or practices, then global conventionalism is a form of semantic anti-realism, for it entails that how we construe an epistemically ideal theory depends on our belief-fixing practices (see, *e.g.*, Horwich 1986, pp. 178-179).

Horwich (1982a, pg. 76-77) claims that the distinction between empirical and theoretical claims on which examples of empirically indistinguishable theories rest cannot be made. Thus EUA is concerned only with examples of type (1) above, and these are handled by adopting global conventionalism.[13] I submit, however, that example (2b) constitutes a non-trivial instance of empirical indistinguishability for

---

[11]He distinguishes this from the "local" conventionalism of Reichenbach(footnote 4).

[12]Recall the Ramsey sentence for a theory T formulated as a sentence S in a 1st-order language L is obtained by replacing some subset of L-terms that occur in S with variables and then existentially quantifying over these variables. Suppose these L-terms are given by $G_1, \ldots, G_k$. Then the Ramsey sentence of $S(G_1, \ldots, G_k)$ is given by $\exists\phi_1 \ldots \exists\phi_k S(\phi_1 , \ldots \phi_k)$. The conditional (*) is sometimes referred to as the Carnap sentence of T.

[13]Horwich (1991, pg. 11, f. n. 6) backs down a bit on the first claim: "...despite the widespread skepticism regarding the existence of such a distinction [between observable and theoretical phenomenon], there is still something to be said for the view that some version of it is psychologically real, and epistemologically important, and can be adequately characterized." However, he still believes that "it is not at all obvious that there are any cases of alternative, empirically equivalent formulations that could not be reconciled in this way [viz., via global conventionalism]." (1991, pg. 13).

which there exists a clear distinction between empirical and theoretical claims and for which the global conventionalist's reference-fixing device (∗) cannot be applied. In both NG_N and NG_C, empirical claims can be restricted to descriptions of particle trajectories, while theoretical claims can be restricted to claims concerning which geometrical objects are absolute and which are dynamical. Since the theories are not homophonically related, the global conventionalist cannot apply his criterion (∗) to decide between them.

This leaves option (bii), the denial of (EIii), to which I now turn.

3.5. Dynamical Symmetries and Essential Structure

By rejecting (EIii) and upholding (SR), one advocates a semantic realist position I wish to associate with structural realism. The question for structural realism in this context is the following: How can purported epistemically indistinguishable theories be read so that they do not entail contradictions? The intuition I wish to motivate is that, while such theories may differ in their ontologies, they share the same essential structure. Hence, while they may contradict each other at the level of ontology, at the level of essential structure, they agree.

One way to flesh out this intuition is by considering the examples of epistemically indistinguishable spacetime theories given above. Such examples suggest that, if two spacetime theories T, T' share the same dynamical symmetries, then they share the same essential dynamical structure. In the next section, I will try to be a bit more explicit about what this entails about the notion of essential structure. For the moment, I will suppose it is a coherent attitude to adopt and examine what it entails about the examples of epistemically indistinguishable theories given above.

Given such structural realist commitments, epistemically indistinguishable (EI) theories of type (1) are taken for what they naively are -- trivial relabelings. For EI theories of types (2) and (3), structural realism appropriates the same intuitions as the conventionalist insofar as it claims that a literal construal of any given member of an equivalence class of dynamically possible models (*i.e.*, any particular manifestation of the essential structure) of a given theory is conventional. However, what is not conventional just is the essential structure in question. In particular, in both examples (2a) and (2b), the essential structure, as indicated by the dynamical symmetries, is the same. In example (2b), NG_N and NG_C indeed are committed to different ontologies and ideologies with respect to space and time insofar as their spacetime symmetries are different. But their dynamical symmetries are the same; hence the essential structure required to construct dynamically possible models is identical. For example (2a), this is more readily evident.

One initial qualification is immediately apparent: The examples detailed above are all given in a common formalism (namely, the tensor formalism) by means of which the notion of dynamical symmetries was cashed out. A more general notion of essential structure should take into account the existence of alternative formalisms and should hence perhaps be based on a more general notion of dynamical equivalence. In the next section I will attempt to flesh these notions out in a bit more detail.

4. Structural Realism

If the semantic component of structural realism enjoins us to read literally the essential structure of our theories, the question still remains as to what this essential structure amounts to. The suggestion raised in the previous section was that it should depend in some way on the dynamical structure associated with a given theory. In the context of spacetime theories formulated in the tensor formalism, the dynamical structure of a theory perhaps can be identified with its dynamical symmetries. Two spacetime theories expressed in the tensor formalism share the same essential dynamical structure just when they share the same dynamical symmetries. In a more general setting, however, determining dynamical structure is bound to be more difficult. To get an indication of this, in the next section I shall briefly look at an example taken from twistor theory. It turns out that certain Yang-Mills gauge field theories admit a twistor formulation. This alternative formulation, I will claim, makes semantic realism with respect to such Yang-Mills theories problematic at the level of individuals-based ontologies, but not so at the level of dynamical structure. With this example in hand, I will turn in Section 4.2 to an appraisal of various proposals in the literature for what essential structure amounts to in the general context.

4.1. Twistors and Fields: Informing the Semantic Component of Structural Realism

The twistor formalism rests on a correspondence between complex, compactified Minkowski spacetime $\mathbb{CM}^c$ and projective twistor space $\mathbb{PT}$. Based on this correspondence, solutions to certain hyperbolic differential equations in Minkowski spacetime can be encoded in complex-analytic, purely geometrical structures in an appropriate twistor space. Hence, the dynamical information represented by the differential equations in the spacetime picture gets encoded in geometric structures in the twistor picture.[14] Advocates of the twistor formalism emphasize this result -- they observe that, in the twistor picture, there are no dynamical equations; there is just geometry. This suggests that a naive semantic realist may be faced with a non-trivial task in providing a literal interpretation of Yang-Mills theories in the twistor formalism. Before discussing this task, I will briefly describe the mathematics underlying the twistor correspondence and its application to Yang-Mills gauge theories.

The correspondence can be encoded most succinctly in what is called a double fibration. This consists of a correspondence space $\mathbb{F}$ which acts as an intermediary between two other spaces. In this case, these spaces are $\mathbb{CM}^c$ and $\mathbb{PT}$. $\mathbb{CM}^c$ is complex, compactified Minkowski spacetime obtained by attaching a null cone at infinity to

---

[14]Besides the correspondence for the anti-self-dual (ASD) Yang-Mills field equations considered below, correspondences also exist for the ASD vacuum Einstein equations (the "non-linear graviton"), the ASD Maxwell equations (the "twisted photon" -- a specialization of the ASD Yang-Mills result), and the zero rest mass field equations for arbitrary spin. Ward (1981) indicates that twistor correspondences can be given in general for (quasi-)linear hyperbolic differential equations in Minkowski spacetime that satisfy one form of Huygen's Principle.

complex Minkowski spacetime (η_{ab}, $\mathbb{C}^4$) (where η_{ab} is the Minkowski metric).[15] $\mathbb{PT}$ is projective twistor space, which is the 3-complex-dimensional space of 2-spinor pairs (ω^A, $\pi_{A'}$), up to a complex constant, that satisfy the twistor equation $\nabla_{B'}{}^{B}\omega^C(x) = -i\varepsilon_B{}^C\pi_{A'}$.[16] The correspondence space $\mathbb{F}$ is the primed spinor bundle over $\mathbb{CM}^c$ consisting of pairs (x^a, $\pi_{A'}$) where x^a is a point in $\mathbb{CM}^c$ and $\pi_{A'}$ is a primed 2-spinor. The double fibration then takes the form,

where the projection maps μ, ν are given by,

$$\nu: \ (x^a, \pi_{A'}) \rightarrow x^a$$

$$\mu: \ (x^a, \pi_{A'}) \rightarrow (ix^{AA'}\pi_{A'}, \pi_{A'})$$

These maps are constructed so that they give the correspondence between elements of $\mathbb{CM}^c$ (spacetime points) and elements of $\mathbb{PT}$ (projective twistors) by the following relation

$$\omega^A = ix^{AA'}\pi_{A'} \qquad \text{(KC)}$$

known as the Klein correspondence. It expresses the condition for the twistor (ω^A, $\pi_{A'}$) $\in \mathbb{T}$ to be incident with the point $x^a \in \mathbb{CM}^c$.[17] Based on this correspondence, the maps allow structures in $\mathbb{PT}$ to be pulled up to $\mathbb{F}$ and then pushed down to $\mathbb{CM}^c$, and *vis*

---

[15]Compactified Minkowski spacetime $\mathbb{M}^c$ is the carrier space for matrix representations of the 4-dimensional conformal group $\mathcal{C}(1, 3)$. $\mathcal{C}(1, 3)$ is generated by conformal mappings of Minkowski spacetime $\mathbb{M}$ to itself.

[16]The 2-spinor ω^A is an element of a complex 2-dimensional vector space $\mathbb{S}$ endowed with a bilinear anti-symmetric 2-form (the spinor "metric") ε_{AB}. ($\mathbb{S}$ is the carrying space for representations of the group $SL(2, \mathbb{C})$, which is the double-covering group of the Lorentz group $SO(1, 3)$.) The 2-spinor $\pi_{A'}$ is an element of the Hermitian conjugate vector space $\mathbb{S}'$. Here and below the abstract index notation for 2-spinors and for tensors is used. In particular, 2-spinor indices are raised and lowered *via* the metrics ε_{AB}, $\varepsilon_{A'B'}$, and tensor indices b can be exchanged for pairs of spinor indices BB'.

Non-projective twistor space $\mathbb{T}$ is the space of solutions (ω^A, $\pi_{A'}$) of the twistor equation, a general solution having the form $\omega^A(x) = \omega_0{}^A$ - $ix^{AA'}\pi_{A'}$, where $\omega_0{}^A$ and $\pi_{A'}$ are constant 2-spinors. $\mathbb{T}$ is a 4-dimensional complex vector space with a Hermitian 2-form (a "metric") of signature $(++--)$. It is the carrying space for representations of the group $SU(2, 2)$, which is the double-covering group of $SO(2, 4)$. Hence (univalent) twistors are the "spinors" of the group SO(2, 4).

[17](KC) gives the locus of points in $\mathbb{CM}^c$ where solutions to the twistor equation vanish (see footnote 16).

versa. In particular, the copy in $\mathbb{PT}$ of the fiber $\nu^{-1}(x^a)$ is obtained directly from (KC) by holding $x^{AA'}$ fixed and varying $(\omega^A, \pi_{A'})$. One obtains a complex linear 2-dimensional space in $\mathbb{T}$, which defines a line in $\mathbb{PT}$. The copy in $\mathbb{CM}^c$ of the fiber $\mu^{-1}(\omega^A, \pi_{A'})$ is obtained in a similar manner by holding the twistor $(\omega^A, \pi_{A'})$ fixed and varying the spacetime point $x^{AA'}$. This defines a complex null 2-dimensional plane in $\mathbb{CM}^c$ referred to as an α-plane. Hence under (KC), points in $\mathbb{CM}^c$ (spacetime points) correspond to "twistor lines", and points in $\mathbb{PT}$ (projective twistors) correspond to α-planes.

Under the double fibration, a Yang-Mills gauge theory on $\mathbb{CM}^c$ can be lifted to $\mathbb{F}$ and it will then depend trivially on the $\pi_{A'}$ variable in the larger correspondence space; *i.e.*, it will be trivial (*i.e.*, constant) on each fiber of μ (each α-plane). Likewise, a gauge theory on $\mathbb{PT}$ can be lifted to $\mathbb{F}$ and will depend trivially on the fibers of ν (twistor lines). Hence in order for gauge theories on $\mathbb{CM}^c$ and $\mathbb{PT}$ to be equivalent, the former have to be trivial on α-planes and the latter have to be trivial on twistor lines. This is the basis for Ward's (1977) theorem, which states that an anti-self-dual Yang-Mills gauge field theory is equivalent to a holomorphic vector bundle over $\mathbb{PT}$ which is trivial on twistor lines:

WARD'S THEOREM: Let U be an elementary open set[18] in $\mathbb{CM}^c$ and U' the corresponding open region in $\mathbb{PT}$ under (KC), which maps points $x \in \mathbb{CM}^c$ into lines $L_x \subset \mathbb{PT}$. There is a 1-1 correspondence between

(a) anti-self-dual $GL(n, \mathbb{C})$ Yang-Mills gauge fields F_{ab} on U; and,

(b) rank n holomorphic vector bundles B over U', such that the restriction $B|_{L_x}$ of B to the line $L_x \subset U'$ is trivial for all $x \in U$.

A full proof is given in Ward and Wells (1990, pp. 374-381). The theorem rests primarily on the fact that a Yang-Mills field F_{ab} is anti-self-dual[19] *if and only if*, for every α-plane Z' that intersects U, the restriction of the covariant derivative D_a to $U \cap Z'$ satisfies $n_a D_a \psi = 0$, for any vector field n_a tangent to Z' and any section ψ of the vector bundle associated with F_{ab} (Ward and Wells 1990, pg. 373). Put simply, F_{ab} is anti-self-dual *if and only if* its associated covariant derivative $D_a = \partial_a - ieA_a$ is flat on α-planes.

Needless to say, anti-self-dual Yang-Mills gauge theories represent a somewhat restricted subclass of Yang-Mills theories. Importantly, the former are sourceless on α-planes, on which they satisfy $D_a F_{ab} = 0$. However, for this restricted class, Ward's theorem indicates that the twistor formulation is expressively equivalent with the standard "spacetime" formulations. Hence a semantic realist interested in taking anti-self-dual Yang-Mills gauge theories at their face value must contend with the twistor

[18] U is an elementary set just when its intersection with any α-plane is connected and simply-connected (Ward and Wells 1990, pg. 372).

[19] An ASD Yang-Mills field is defined by the ASD Yang-Mills equations: $*F_{ab} = -iF_{ab}$, where * is the Hodge-dual operator.

formulation. Finally, it should be noted that versions of Ward's theorem exist for all subgroups of $GL(n, \mathbb{C})$ (Ward and Wells 1990, pg. 386).

How can a semantic realist take the twistor formulation of an anti-self-dual Yang-Mills gauge theory at its face value? Note first that there are significant differences between the bundle that appears in the twistor formulation and the bundle that appears in the standard "spacetime" fiber bundle formulation of ASD Yang-Mills theory. The fiber bundle description of an ASD Yang-Mills theory consists of a principle G-bundle with a connection 1-form over a base manifold identified as spacetime. It is the connection that encodes the dynamics in so far as it is the covariant derivative defined by the connection that acts on matter fields and Yang-Mills fields to evolve them in time. In the twistor picture, one has a vector bundle with no connection. The vector bundle consists of fibers, each one of which is a vector space of ASD Yang-Mills fields; i.e., Yang-Mills fields that satisfy the ASD Yang-Mills equations (see footnote 19). Hence, in a literal sense, the local dynamics in the spacetime formulation gets encoded in a global geometrical structure in the twistor description, as twistor advocates like to point out:

> Note that in the Ward construction the local 'field' information in the space time description is coded in the global structure of the twistor description, whereas there is no local (differential) information in the twistor description... This way in which local space-time field equations tend to 'evaporate' into global holomorphic structure is a characteristic (and somewhat remarkable) feature of twistor descriptions (Penrose and Rindler 1984, pg. 168).

Another difference is that the bundle in the twistor formulation is over projective twistor space $\mathbb{PT}$, the elements of which are projective twistors. The bundle in the fiber bundle formulation is over a base manifold generally taken as representing spacetime, the elements of which are spacetime points. These differences in formulation present a problem for a semantic realist, as I shall now argue.

In the spacetime formulations, traditional semantic realists have tended to read literally the mathematical fields that quantify over the points of the manifold (be it the base space or a bundle space); the resulting literal interpretation describes physical fields that quantify over spacetime points, and that are evolved in time by means of partial differential equations that incorporate the derivative operator associated with a connection. This basic ontology of fields and spacetime points is by no means unproblematic; however, arguably, it informs the most literal construal of Yang-Mills theory in the spacetime formulation. Two observations to qualify this claim are appropriate. First, there are a variety of specific interpretive options for the field realist in the context of Yang-Mills theory, even given a basic field/point ontology. Such options divide over what field to take as fundamental (the gauge potential A_a or the gauge field F_{ab}); and also the nature of the manifold objects such fields quantify over (e.g., points or loops). However, these options still subscribe to a literal reading of the mathematical fields that appear in the spacetime formulation, and the underlying manifold on which they are defined.

In the twistor picture, the mathematical field has vanished, as have the dynamical equations, and both have been replaced by a vector bundle over projective twistor space $\mathbb{PT}$. Literally, this bundle is a collection of vector spaces labeled by the points of $\mathbb{PT}$. In a sense, such vector spaces quantify over the points of $\mathbb{PT}$, these points being projective twistors. Recall that projective twistors correspond to complex null surfaces (α-planes) in complex compactified Minkowski spacetime $\mathbb{CM}^c$. It turns out that when the Klein correspondence is restricted to real compactified Minkowski spacetime $\mathbb{M}^c$, projective twistors correspond, in general, to twisted congruences of null geodesics referred to as Robinson congruences.[20] One option, then, for a traditional semantic realist is to view such null geodesics as the individuals in the ontology of anti-self-dual Yang-Mills gauge theories formulated in the twistor formalism. Note, however, that such individuals are very different from those in the literal construal of the spacetime formulation. The "twistor individuals" might be said to be inherently globally dynamical. Each projective twistor has associated with it an entire vector space of solutions to the anti-self-dual Yang-Mills equations. This space might be thought of as encoding possible internal properties associated with the given twistor.

Hence the semantic realist committed to an individuals-based ontology has to decide between two seemingly incompatible literal construals of Yang-Mills theory: The tensor formalism suggests a commitment to local fields and spacetime points, whereas the twistor formalism suggests a commitment to globally twisted null geodesics.[21]

The traditional realist might respond by claiming that Ward's theorem just shows that solutions to the anti-self-dual Yang-Mills equations behave in spacetime as if they were rank n vector bundles that quantify over twisted null geodesics. In other words, we should not read the twistor picture literally - it merely amounts to a way of encoding the behavior of the real objects, which are fields in spacetime, and which are represented more directly in the tensor formalism. In other words, we should only be semantic realists with respect to the tensor formalism. This strategy smacks a bit of ad hocness. All things being equal, what, we may ask, privileges the tensor formalism over the twistor formalism?[22] From a conventionalist's point of view, tensor fields on a manifold are just as much devices that encode the data provided by measuring apparatii as are vector bundles over PT. If the semantic realist is to be genuine about

---

[20]This is the origin of the term "twistor". Roughly, the real correlates of projective twistors correspond to the intersections of a-planes and their duals, referred to as b-planes and defined with respect to the Hermitian twistor "metric" Σ_{ab}. For a null twistor Z^α that satisfies $\Sigma_{ab}Z^\alpha Z^\beta = 0$, this intersection is given by a null geodesic. For non-null twistors, the intersection is given by a Robinson congruence -- a collection of null geodesics that twist about the axis defined by the null case.

[21]The headaches for the traditional semantic realist do not stop here. Other interpretations of twistors-as-individuals: encoding angular momentum and helicity of zero rest mass particles; spinor fields; charges for spin 3/2 fields.

[22]It turns out that all things are not exactly equal when it comes to formulating other types of field theories. In particular, no consistent twistor descriptions have been given for field theories with sources, nor for field theories in generally curved spacetimes. However, for the particular context of anti-self-dual Yang-Mills theories (and those detailed in footnote 14), expressive equivalence between the tensor and twistor formalisms holds.

her semantic realism, it appears that she must be willing to give up commitment to individuals-based ontologies and seek the basis for her literal construal at a deeper level.

4.2. What is Structure?

Armed primarily with the above twistor example, in this section I will consider three proposals that attempt to flesh out the notion of structure: those given by Worrall (1994, 1989); Ladyman (1998) and French (1999); and Bain and Norton (2001).

Worrall (1994, 1989) has suggested that the structure of a theory be identified with the theory's dynamical equations. In the relevant cases, it is the equations that are retained across theory change. For instance, in the case of Fresnel's wave theory of light, Worrall sees the equations of defraction and reflection retained in Maxwell's later electromagnetic theory:

> Thus if we restrict ourselves to the level of mathematical equations - not notice the phenomenal level - there is in fact complete continuity between Fresnel's and Maxwell's theories (Worrall 1989, pg 119).

Moreover, Worrall distinguishes between the *structure* represented by such equations, and the *natures* of individuals that populate traditional semantic realist ontologies. With respect to Fresnel's treatment of light as a wave propagating in an immaterial ether, Worrall claims:

> ...Fresnel clearly misidentified the nature of light, but his theory nonetheless accurately described not just light's observable effects, but also its structure (Worrall 1994, pg. 340).

This approach appears to suffer from two defects. First, we have seen in the twistor formulation of anti-self-dual Yang-Mills theories an “evaporation” of partial differential dynamical equations into global holomorphic geometric structures. Certainly, dynamical equations are one way to encode dynamical structure, but evidently they are not the only means of doing so. Admittedly, the issue at this stage may revolve around exactly what a dynamical equation is. The only point being made here is that care must be taken in identifying these objects if Worrall's thesis is to be maintained.

Second, Psillos (1999) claims that Worrall's distinction between structure and nature, when applied to entities, cannot be maintained:

> To say what an entity is is to show how this entity is structured: what are its properties, in what relations it stands to other objects, etc. An exhaustive specification of this set of properties and relations leaves nothing left out. Any talk of something else remaining uncaptured when this specification is made is, I think, obscure. I conclude, then, that the 'nature' of an entity forms a continuum with its 'structure', and that knowing the one involves and entails knowing the other (Psillos 1999, 156-157).

Psillos claims that to say that Fresnel got the structure of light right is just to say that he got some basic aspects of its nature right. In particular, Psillos' preferred reading of

Fresnel is that Fresnel got some of the fundamental properties of light right. And this can be said without recourse to talk of structure.

Ladyman (1998) has responded to Psillos' critique by explicitly divorcing structural realism entirely from commitments to "individuals-based ontologies" (pg. 422).[23] Ladyman (1998) and French (1999) have suggested an approach to structural realism that identifies the "invariants" underlying various expressively equivalent formulations of a theory as the theory's structure:

> Here there are no unknowable objects lurking in the shadows and objectivity is understood structurally, in terms of the relevant set of invariants (French 1999, pg. 203).

> The idea then is that we have various representations which may be transformed or translated into one another, and then we have an invariant state under such transformations which represents the objective state of affairs (Ladyman 1998, pg. 421).

French suggests that such structure is initially arrived at by considering the "elements" on which it is predicated as individuals. Such individuals, however, only play a heuristic role and it is the structure they enter into that carries the ontological weight (French 1999, pg. 204). Once this structure is identified, the underlying "individuals" can be discarded. French furthermore suggests that group theory is a means of identifying such structure.

French and Ladyman's rejection of "individuals-based ontologies" squares with the discussion of the twistor formulation of anti-self-dual Yang-Mills theories in Section 4.1 above. There it was found that literal interpretations of the twistor and spacetime formulations of such theories diverge at the level of individuals. However, identifying the essential dynamical structure of Yang-Mills theories with their group structure is a bit problematic. An initial problem with this program comes in identifying the group structure that is essential to a given theory. On a literal construal of the spacetime formulations of Yang-Mills theories, the groups that appear are the Poincaré group $ISO(1, 3)$ (which might be seen as encoding the metrical structure of Minkowski spacetime) and the relevant infinite Lie gauge group (referred to in the physics literature as the "local" symmetry group). There are two reasons why we should perhaps be wary of identifying the structure of Yang-Mills gauge theories with such groups. First, the fundamental relevance of the "local" symmetry group is suspect, based as it is on a literal construal of what is known in the literature as the Gauge Argument.[24] Suffice it to say here that, in so far as there are alternative treatments of Yang-Mills gauge theories that do not prioritize "local" symmetry groups, viewing such groups as encoding the fundamental structure of Yang-Mills gauge theories is perhaps a bit premature. Second, in the twistor formulation, the groups that appear are really

---

[23]Redhead (1999) holds a similar view: "My own view is that the best candidate for what is "true" about a physical theory is the abstract structural aspect. The main reason for this is that there is significantly greater continuity in structural aspects than there is in ontology." (Redhead 1999, pg. 34).

[24]For a critique of literal construals of the Gauge Argument, see Martin (2002).

$SU(2, 2)$ and a given gauge group (see, e.g., footnote 16). Intuitively, the twistor formulation privileges the conformal group $\mathcal{C}(1, 3)$ over the Poincaré group. (It privileges the conformal structure of Minkowski spacetime over its metrical structure.) Hence, at least in the case of anti-self-dual Yang-Mills gauge theories, it is unclear what group should be considered fundamental, and subsequently, which "invariants" should be considered as constituting the essential structure.

On the other hand, a case could be made that the move from the spacetime formalism to the twistor formalism involves a trimming of extraneous group structure. This is evident in the sequences of isomorphisms between group structures that appear in the spacetime and twistor formalisms. In the spacetime formalism (for field theories in Minkowski spacetime), one has the sequence

$$SL(2, n) \xrightarrow{2-1} SO(1, 3) \longrightarrow \mathcal{C}(2)$$

where $\mathcal{C}(2)$ is the 2-dimensional conformal group (comprised of automorphisms on the Riemann sphere that preserve the sphere metric analogue of the Minkowski metric). In the twistor formalism, one has the sequence

$$SU(2, 2) \xrightarrow{2-1} SO(2, 4) \xrightarrow{2-1} \mathcal{C}(1, 3)$$

where $SO(2, 4)$ is the restricted 6-dimensional pseudo-orthogonal group (a 6-dimensional extension of the Poincaré group). The move to twistors can thus be seen as an enlargement of group structure (from metrical to conformal structure), and not as a choice between different incompatible group structures. A group-theoretic enthusiast might thus claim that the essential group to be associated with ASD-YM theories is the twistor group $SU(2, 2)$, and the invariants that encode the essential structure of ASD-YM theoris are just twistors.

However, it seems that there is more to say about anti-self-dual Yang-Mills theories than can be encapsulated solely in the "invariants" of $SU(2, 2)$; in other words, in twistors. In particular, the more to be said has to do with the dynamics of the theory. This dynamics can be encoded not simply in the "invariants" of appropriate groups (i.e., in carriers of representations of such groups), but in spaces related to the carrying space of representations of these groups. The dynamics of anti-self-dual Yang-Mills theories is encoded not just in $SU(2, 2)$ invariants (i.e., twistors), but in geometric structures defined over the projective carrying space of such invariants. Thus, to the extent that essential structure involves dynamics, we should perhaps be wary of identifying essential structure solely with group-theoretic invariants.

An advocate of the group-theoretic approach to structure may respond to this by citing Poincaré invariant quantum field theory. Weinberg (1995), for instance, indicates explicitly how the dynamical equations governing free quantum fields can be derived solely from their properties as irreducible representations of $ISO(1, 3)$. The Klein-Gordon equation falls out of an analysis of the properties of scalar representations of $ISO(1, 3)$, for instance, as does the Dirac equation from an analysis of the properties of

4-spinor representations of $ISO(1, 3)$.[25] Hence, in the absence of interactions, the "invariants" of $ISO(1, 3)$ have their dynamics "built-in", so to speak. On the surface, this seems to lend additional credence to the group-theoretic approach to structure. Again, however, I think trouble lurks below the surface. It seems to me that, at least from an interpretational point of view, this treatment of quantum fields places undue emphasis on the free theory. If we are to take interacting quantum field theory seriously, then we must take seriously its fundamental claim that interactions can never be completely "turned off"; hence free field theory is, at best, an idealization. Issues surrounding Haag's Theorem then indicate that an emphasis on "invariants" (*i.e.*, irreducible representations of $ISO(1, 3)$) leads to conceptual problems.[26] Of course this cuts both ways: if we are to take interacting QFT seriously, then we should not take seriously anti-self-dual Yang-Mills theory, insofar as the fields it describes are sourceless. The point to make here is simply that we should be open to other ways of representing essential structure than as group-theoretic invariants.

Finally, in Section 3.1 above, I claimed that the dynamical symmetries of NG_N (Newtonian gravity in Neo-Newtonian spacetime) and NG_C (Newtonian gravity in Newton-Cartan spacetime) were the same, namely elements of the Maxwell group. But the dynamical objects of both theories are different. But such dynamical objects intuitively seem to be what we would consider the dynamical "invariants" of these theories. Under a group-theoretic invariant notion of structure, then, we would have to admit that NG_N and NG_C do not possess the same essential structure.[27]

A final notion of structure that has appeared in the literature is that given in Bain and Norton (2001), who suggest that the Hamiltonian and Lagrangian formalisms are instructive methods of encoding structure. In particular, the structure of the electron, say, is best given by the Hamiltonian or Lagrangian formulation of the particular electron theory. According to Bain and Norton, evidence of such structure comes in the form of "historically stable properties" that the electron exhibits (spatio-temporal properties like spin, as well as electromagnetic, electroweak, and chromodynamic properties that govern how the electron couples with other objects). This might be construed as a return to a view of structure as a property predicated on individual

---

[25]Weinberg also demonstrates how the Maxwell equations result from an analysis of the properties of vector representations of $ISO(1, 3)$. This analysis also produces the $U(1)$ "local" gauge symmetry, and again demonstrates that the claim of fundamentality for "local" symmetries is a suspect thesis.

[26]Haag's Theorem states that, if two irreducible representations of $ISO(1, 3)$ are related by a unitary transformation at a given time t, then for all other times, both of them are free if one of them is free. Simply put: if quantum fields are given by $ISO(1, 3)$ invariants, then the dynamics of free fields cannot be linked with the dynamics of interacting fields in a way that preserves unitarity.

[27]How can the dynamical symmetries of NG_N and NG_C be the same - (Max) - yet their dynamical objects be different? The difference lies in the connections associated with both theories. The NG_N connection is invariant under (Gal) and not under (Max), whereas the NG_C connection is invariant under (Max). Hence the NG_N connection is considered absolute, in so far as the spacetime symmetries of Neo-Newtonian spacetime are (Gal), whereas the NG_C connection is considered dynamical. (The NG_C connection has a part that is dynamic and a part that is absolute (namely, its spatial part). The move to NG_C essentially involves trimming off part of the absolute NG_C connection and making it dynamic. For further details, see Bain (2005).

systems. However, the more appropriate reading would abstract the structure from the individuals *a la* French and Ladyman. Again, historically stable properties are indicators of underlying structure, but should not perhaps be identified uniquely with such structure. But it still remains unclear as to how the Hamiltonian or Lagrangian for a given theory explicitly encodes its structure.

One clue is given by Ruetsche's (2002, pg. 200) description of an interpretation of a theory. After Ruetsche, let an interpretation be given by a 4-tuple {state space, observables/beables, dynamics, semantics}.[28] Elements of the state space correspond to dynamically possible worlds. The observables and beables of the theory are given by functions on the state space. Specifying the beables determines if the interpretation is deterministic, while specifying the observables determines if the interpretation is predictive.[29] These specifications are made via the semantics, which explicitly states what the state space, observables, and beables are. Specifying the dynamics involves specifying a map from the state space to itself that determines how initial dynamically possible worlds evolve in time. In the Hamiltonian formalism, for instance, such a dynamics is given by a Hamiltonian function on the state space that assigns to every dynamically possible world an energy. Given that the state space is appropriately chosen (choosing it as a symplectic manifold, for instance), the Hamiltonian will generate a flow that determines the dynamical trajectories.

Such a description motivates identifying the dynamical structure associated with a theory as consisting abstractly of the triple {state space, dynamics, symmetries}. This object may be considered the solution space associated with the theory (technical correspondence here between state space and solution space). It specifies equivalence classes of dynamically possible trajectories, which may be thought of as solutions to the theory's dynamical equations. Both the spacetime and twistor formulations of ASD Yang-Mills theories share this basic structure (and, assumedly, the set of observables). What they differ on explicitly is the specification of the set of beables. More precisely, Ward's Theorem sets up a 1-1 correspondence between holomorphic vector bundles over $\mathbb{PT}$ and the solution space of the ASD Yang-Mills field equations on Minkowski spacetime. <more discussion>

Dynamical structure as universality class.

The above discussion may be considered a modest first attempt at coming to grasps with the notion of essential dynamical structure. Again, the semantic component of structural realism would have us ontologically commit ourselves only to the essential dynamical structure associated with a given theory. Note finally that this is not to

[28]Reutsche does not explicitly refer to beables in her description (see footnote 29 below).

[29]This distinction follows Belot (1995). A beable in this sense is an individuating property. Intuitively, such beables demarcate the physically possible worlds of a given theory, whereas observables demarcate the observationally distinguishable worlds (thus, depending on one's notion of individuation, beables in this sense may or may not be observables). This notion of beable should be made distinct from another notion that sometimes appears in interpretations of quantum mechanics and is attributable to the physicist John Stewart Bell. A beable in Bell's sense is a property that possesses a definite value at all times. Such Bell-beables always count as observables.

eliminate all talk of "individuals-based" ontologies completely. From a heuristic point of view, we definitely do not want to give up such ontologies. How you interpret a given theory -- which ontology you associate it with -- will certainly influence how you approach the task of extending it (as the interplay between interpretations of general relativity and approaches to quantum gravity demonstrates). All the structural realist claims in this context is that, at most, what you should ontologically commit to is essential structure.

5. Semantic Underdetermination

In this section, I will briefly review the Pessimistic Meta-Induction (PMI) and claim that what is at stake is the reliability of methods of identification for theoretical claims. This suggests the following criterion of warranted belief: A necessary condition for warranted belief is the existence of a reliable method of inquiry that produces the belief in question. I adopt this as a criterion of warranted belief that anti-realist and realist can agree on. I then offer a sketch of how a demonstration, based on this criterion, that structural realism avoids the PMI might proceed.

5.1. The Pessimistic Meta-Induction: Reliability and Warranted Belief

The PMI claims in general that epistemic realism undermines semantic realism. Schematically,

(1) There are theories in the past which, according to the epistemic realist, warrant our belief.

(2) We cannot be semantic realists with respect to these theories (since the semantic claims of such theories are wrong).

∴ We should not be semantic realists with respect to current theories which, according to the epistemic realist, warrant our belief.

The PMI is usually interpreted as being aimed specifically at the explanatory defense of scientific realism (see, *e.g.*, Psillos 1999, Chap 5). Its intent, it is claimed, is to sever the explanatory link between success and truthlikeness on which the explanatory defense is based. I think a better construal of the PMI is that it is questioning the reliability of theoretical claims; in particular, it questions the reliability of semantic realist interpretations of successful theories. Hence its target extends beyond the explanatory defense of realism to semantic realism in general.

I interpret the above form as contending that theoretical claims, read literally, are not reliable. Furthermore, the EUA can also be interpreted as a reliability critique of realism. In general, the anti-realist claims that theoretical claims are not reliable. The EUA in particular concludes that we should not believe them (they are evidentially unreliable); whereas the PMI concludes that we should not read them literally (they are referentially unreliable).

In general, there seem to be two ways to respond to the PMI. A specific response attempts to show historically that referential stability is exhibited. Such demonstrations generally require a distinction to be made between relevant and irrelevant parts of a theory.[30] Such demonstrations weaken the inductive base for the PMI and prime the realist's intuition pump. A more general response allows that, even if such stability can be demonstrated in specific examples, the inductive skeptic who licenses belief only in inferences to in-principle observables will not be convinced. To fully convince such a skeptic requires a demonstration that inferences to structure are just as reliable as the types of inference such a skeptic is willing to allow. Such a demonstration, if successful, would counter both forms of underdetermination argument in one fell swoop. While a full treatment of this is beyond the scope of the present essay, in the next section I will sketch one way such a demonstration might proceed.

5.2. Structural Realism and Reliabilism: An Outline

The PMI contends that theoretical claims do not have referential stability. Above, I interpreted this as: There are no reliable methods of identification for theoretical claims. My take of the PMI thus has it adopting the following reliabilist epistemic criterion: A necessary condition for warranted belief is the existence of a reliable method of inquiry. In this section, I indicate how this might be fleshed out in terms of formal learning theory. To begin, I shall distinguish informally between three types of inference:

(A) inferences to unobserved observables;

(B) inferences to in-principle unobservables;

(C) inferences to structure.

I take (A) to underlie empirical claims in general, while (B) underlies theoretical claims. The anti-realist in general contends that inferences of type (A) are warranted while inferences of type (B) are not. In the reliabilist paradigm, this is the claim that reliable methods exist for making inferences of type (A), but do not exist for making inferences of type (B). Now suppose inferences of type (C) have the following properties:

(i) They may legitimately be said to underlie theoretical claims;

(ii) they can be defined by means of reliable methods; and,

---

[30]See, *e.g.*, Kitcher's (1993) distinction between working posits and presuppositional posits, or Psillos' (1999) related distinction between idle and essentially contributing constituents. Bain and Norton (2001) make the distinction in terms of essential *vs* surplus structure.

(iii) they are reliable in contexts in which inferences of type (A) are and type (B) are not.

If these properties could be supported, then the epistemic anti-realist would be wrong when she infers from the non-existence of reliable inferences of type (B) to the non-existence of reliable inferences to theoretical claims in general. The task, then, for a reliabilist defense of epistemic structural realism is to make good on these distinctions. What follows is one way this task could be approached.

Before I begin, a few qualifications are in order. Again, the goal is simply to indicate that inferences to structure (i.e., that which purported EI hypotheses have in common) can be defined in a manner that should be unobjectionable to the anti-realist who employs the EUA and the PMI; *i.e.*, to the anti-realist who holds that inferences of type (A) are warranted and inferences of type (B) are not. I am not advocating formal learning theoretic techniques in general; nor am I claiming such techniques are the only means by which the relevant distinctions can be articulated. I claim merely that they provide a means of illustrating these distinctions. The main results are given in Theorems 2 and 4 below. Theorem 2 establishes a criterion of reliability under which inferences of type (A) are warranted and inferences of type (B) are not. Theorem 4 establishes a criterion of reliability under which inferences of types (A) and (C) are warranted and inferences of type (B) are not. I submit that these results substantiate the structural realist's contention that terms referring to structure are at least as referentially stable as the terms the anti-realist who employs the EUA and the PMI is willing to accept as referentially stable.

The set up is the following:[31] I will consider a set of possible worlds $\mathcal{W} = \{w_1, w_2, ...\}$; an evidence matrix Po$\mathcal{W} \supseteq \mathcal{E} = \{E_1, E_2, ...\}$ which consists of sets of worlds (here and below, "Po" denotes "power set"); and a set of hypotheses $\mathcal{H} = \{H_1, H_2, ...\}$ which is a partition of $\mathcal{W}$ (*i.e.*, the individual H_i's are mutually exhaustive and pair-wise consistent with respect to $\mathcal{W}$). From $\mathcal{E}$ one can construct a set of infinite data sequences $e^w = (E_1{}^w, E_2{}^w, ...)$, where $E_k{}^w$ stands for E_k or $\sim E_k$ depending on whether $w \vDash E_k$ or $w \vDash \sim E_k$. Let $e_n{}^w$ denote the finite sequence $(E_1{}^w, ..., E_n{}^w)$. Furthermore, for any world $w \in \mathcal{W}$, let $D_{\mathcal{E}}(w) = \{E \in \mathcal{E} : w \vDash E\}$ be the set of all evidence statements that are true in w, and let $\mathcal{DT}_{\mathcal{E}}(w) = \{e^w : \text{rng}(e^w) = D_{\mathcal{E}}(w)\}$ be the set of all data sequences that are true in w (here rng(e^w) is the set of evidence statements occurring in e^w). Finally, a *method of hypothesis identification* F is a map from finite evidence sequences to hypotheses, $F : \{e_n{}^w\} \rightarrow \mathcal{H}$. F takes a finite data sequence and outputs the hypotheses for which the sequence holds true.

The goal now is to explicate the sense in which a method F can be said to be reliable and the conditions under which reliability is achievable. In doing so, I shall assume that the following restrictions hold for the above formal description of hypotheses and evidence.

[31] Here I follow Earman (1993) and Kelly (1996, Chap. 12).

(1) The evidence is perfect: All true data are presented and no false datum is presented.

(2) Evidence sequences may present themselves in any order.

(3) All possible worlds involve at most countably many events or objects to be observed, and each is eventually described in the evidence.

These are made for the interests of simplicity and, as far as I am aware, do not bare on the discussion below. Constraints (1) and (2) have been addressed by Kelly and Glymour (1992), Kelly, Juhl and Glymour (1994), and Kelly (1996, Chap. 15) in extensions of the formal learning paradigm to cases in which the data stream is "infected" by the actions of the scientist. Constraint (3) can be overcome by adopting the topological paradigm (Kelly 1996, xx).[32]

$\mathcal{E}$-separation, $\mathcal{H}$-separation and Empirical Distinguishability (ED)
I first describe two ways in which worlds may differ. Based on these, a notion of empirical distinguishability for hypotheses is defined, and a necessary and sufficient condition for empirical distinguishability is given. This will be essential in the definition of reliability to follow.

The first way in which worlds may differ occurs when there is a piece of evidence on which they disagree:

Def. 1. Two worlds $w_i, w_j \in \mathcal{W}$ are *$\mathcal{E}$-separated* ($\mathcal{E}$-*sep*(w_i, w_j)) just when $\exists E_k \in \mathcal{E}$ such that $E_k^{w_i} \neq E_k^{w_j}$.

A second way in which worlds may disagree is over hypotheses:

Def. 2. Two worlds $w_i, w_j \in \mathcal{W}$ are *$\mathcal{H}$-separated* ($\mathcal{H}$-*sep*(w_i, w_j)) just when $\exists H_i, H_j \in \mathcal{H}$, $(i \neq j)$, such that $w_i \in H_i$, $w_j \in H_j$.

It is now possible to define a notion of empirical distinguishability.[33]

[32]Kelly (1996, pg. 270) further restricts the learning context to "empirical" hypotheses; i.e., hypotheses stated in a language that does not out-strip the evidence language. (He also considers hypotheses that only involve first-order quantification.) His concern is with mapping out the levels of inductive skepticism associated with inferences of type (A). He makes a distinction between local underdetermination at this level and global underdetermination. It is the latter that is the concern of the scientific realism debate.

[33]Other types of empirical distinguishability can be defined. Earman (1993) gives three. Def. 3 in the text above corresponds to his ED_3. A more restrictive version occurs when hypotheses in $\mathcal{H}$ are themselves $\mathcal{E}$-separated from each other. Let $\mathcal{E}^*$-*sep*(H_i, H_j) just when $\mathcal{E}$-*sep*(w_i, w_j), for $\forall w_i \in H_i$, $\forall w_j \in H_j$. Then $\mathcal{H}$ is ED_1 *w.r.t.* $\mathcal{E}$ and $\mathcal{W}$ just when $\forall H_i, H_j \in \mathcal{H}$, $(i \neq j)$, $\mathcal{E}^*$-*sep*(H_i, H_j). The piece of evidence that separates H_i and H_j may be interpreted as a crucial test. The halfway house ED_2 occurs when $\forall w_i \in \mathcal{W}$, $\exists w_j \in \mathcal{W}$ such that $\mathcal{H}$-*sep*$(w_i, w_j) \Rightarrow \mathcal{E}$-*sep*$(w_i, w_j)$. My concern is with the weakest version ED_3 insofar as its negation is the strongest form of empirical indistinguishability.

Def. 3. $\mathcal{H}$ is *empirically distinguishable (ED) with respect to* $\mathcal{E}$ *and* $\mathcal{W}$ just when
$$\forall w_i, w_j \in \mathcal{W}, \mathcal{H}\text{-}sep(w_i, w_j) \Rightarrow \mathcal{E}\text{-}sep(w_i, w_j).$$

A necessary and sufficient condition for a set of hypotheses to be ED can now be stated. Note first that $\sim\mathcal{E}$-sep is an equivalence relation on $\mathcal{W}$. Hence $\mathcal{W}/(\sim\mathcal{E}\text{-}sep) = \{G_i : G_i$ is a $\sim\mathcal{E}$-sep equivalence class of worlds $w_i \in \mathcal{W}\}$ is a partition of $\mathcal{W}$. To facilitate the proof of the necessary and sufficient condition, the following definition and lemmas will be helpful:

Def. 4. Two worlds $w_i, w_j \in \mathcal{W}$ are *G-separated* ($G\text{-}sep(w_i, w_j)$) just when
$$\exists G_i, G_j \in \mathcal{W}/(\sim\mathcal{E}\text{-}sep), (i \neq j), \text{ such that } w_i \in G_i, w_j \in G_j.$$

Lemma 1. $\forall w_i, w_j \in \mathcal{W}, G\text{-}sep(w_i, w_j) \Leftrightarrow \mathcal{E}\text{-}sep(w_i, w_j)$
Proof. By definition of the G equivalence classes.

Lemma 2. $(\forall w_i, w_j \in \mathcal{W}, \mathcal{H}\text{-}sep(w_i, w_j) \Rightarrow G\text{-}sep(w_i, w_j)) \Leftrightarrow$
$\forall H_i \in \mathcal{H}, H_i = \bigcup_\ell G_\ell$, where $G_\ell \in \mathcal{W}/(\sim\mathcal{E}\text{-}sep)$ and the union is over countably many G_ℓ's.
Proof: ($\Leftarrow$) Suppose $\forall H_i \in \mathcal{H}, H_i = \bigcup_\ell G_\ell$. Then $\forall G \in \mathcal{W}/(\sim\mathcal{E}\text{-}sep), \exists H \in \mathcal{H}$ such that $G \subseteq H$. Now choose w_1, w_2 such that $\mathcal{H}\text{-}sep(w_1, w_2)$. Suppose $\sim G\text{-}sep(w_1, w_2)$; *i.e.*, $\exists G' \in \mathcal{W}/(\sim\mathcal{E}\text{-}sep)$ such that $w_1, w_2 \in G'$. Then $\exists H' \in \mathcal{H}$ such that $G' \subseteq H'$. Hence $w_1, w_2 \in H'$, and $\sim\mathcal{H}\text{-}sep(w_1, w_2)$. ($\Rightarrow$) Suppose $\forall w_i, w_j \in \mathcal{W}, \mathcal{H}\text{-}sep(w_i, w_j) \Rightarrow G\text{-}sep(w_i, w_j)$. Now show that for any $H \in \mathcal{H}$, there is a $G \in \mathcal{W}/(\sim\mathcal{E}\text{-}sep)$ such that $\forall w \in \mathcal{W}, (w \in G) \Rightarrow (w \in H)$. Suppose otherwise: $\exists H$ such that $\forall G, \exists w$ such that $w \in G$ and $w \notin H$. Now choose w', H' and G' such that $w' \in H'$ and $w' \in G'$ (this can always be done since $\mathcal{H}$ and $\mathcal{W}/(\sim\mathcal{E}\text{-}sep)$ are partitions of $\mathcal{W}$). Then $\mathcal{H}\text{-}sep(w, w')$ and $\sim G\text{-}sep(w, w')$.

The necessary and sufficient condition for ED can now be stated: $\mathcal{H}$ is ED *iff* all hypotheses $H_i \in \mathcal{H}$ are countable unions of partition elements G. Formally,

Theorem 1. ($\mathcal{H}$ is ED) $\Leftrightarrow$ $[\forall H_i \in \mathcal{H}, H_i = \bigcup_\ell G_\ell$, where $G_\ell \in \mathcal{W}/(\sim\mathcal{E}\text{-}sep)]$
Proof: ($\Leftarrow$) Suppose $\forall H_i \in \mathcal{H}, H_i = \bigcup_\ell G_\ell$. Then, by Lemma 2, $\forall w_i, w_j \in \mathcal{W}, \mathcal{H}\text{-}sep(w_i, w_j) \Rightarrow G\text{-}sep(w_i, w_j)$. Hence by Lemma 1, $\forall w_i, w_j \in \mathcal{W}, \mathcal{H}\text{-}sep(w_i, w_j) \Rightarrow \mathcal{E}\text{-}sep(w_i, w_j)$, and $\mathcal{H}$ is ED. ($\Rightarrow$) Suppose $\mathcal{H}$ is ED. Then, by Def. 3 and Lemma 1, $\forall w_i, w_j \in \mathcal{W}, \mathcal{H}\text{-}sep(w_i, w_j) \Rightarrow G\text{-}sep(w_i, w_j)$. Hence, by Lemma 2, $\forall H_i \in \mathcal{H}, H_i = \bigcup_\ell G_\ell$.

Note that each equivalence class G_ℓ is the intersection of countably many E_k (intuitively, by definition, any G_ℓ can't be "cut" by any element of $\mathcal{E}$). Hence, $\mathcal{H}$ is ED *iff* each $H_i \in \mathcal{H}$ is the countable union of the intersection of countably many E_k; *i.e.*, $H_i = \bigcup_\ell(\bigcap_k \mathrm{E_k})_\ell$. So Theorem 1 can be restated as ($\mathcal{H}$ is ED) $\Leftrightarrow$ ($\forall H_i \in \mathcal{H}, H_i \in \Sigma^{\boldsymbol{B}}{}_2$), where the Borel set $\Sigma^{\boldsymbol{B}}{}_2$ consists of countable unions of countable intersections of closed sets of worlds (see, *e.g.*, Kelly 1996, pg. 89).

Reliability in the limit for hypothesis identification
I now define what it means for a method of hypothesis identification F to be reliable.

Def. 4. The method F is *reliable in the limit for $\mathcal{H}$ with respect to $\mathcal{E}$ and $\mathcal{W}$* just when

$$\forall H_j \in \mathcal{H}, \forall w \in \mathcal{W}, \forall e^w \in \mathcal{DT}_{\mathcal{E}}(w), \exists N \text{ such that } \forall n \geq N,$$
$$(F(e_n{}^w) = H_j) \Leftrightarrow C(e^w, H_j),$$
where $C(e^w, H_j)$ just when $w \vDash H_j$.

To say that H_j is identified by F means just that if H_j is the C-correct hypothesis, then there will be a point in the data stream after which F posits H_j forever. Theorem 3 below establishes the type of hypothesis that can be reliably identified. First, note that a necessary condition for reliability in the limit is ED:

Theorem 2. (F is reliable in the limit for $\mathcal{H}$) $\Rightarrow$ ($\mathcal{H}$ is ED)
Proof: Suppose $\mathcal{H}$ is not ED. Then there exist $w_1, w_2 \in \mathcal{W}$ such that
(a) $\mathcal{H}$-sep(w_1, w_2), and
(b) $\sim\mathcal{E}$-sep(w_1, w_2).
Now suppose a reliable F exists. Then (a) implies $F(e_n{}^{w_1}) \neq F(e_n{}^{w_2})$, for all $e^{w_1} \in \mathcal{DT}_{\mathcal{E}}(w_1)$ and all $e^{w_2} \in \mathcal{DT}_{\mathcal{E}}(w_2)$. But (b) implies that there exists a sequence $e^w \in \mathcal{DT}_{\mathcal{E}}(w_1) \bigcap \mathcal{DT}_{\mathcal{E}}(w_2)$, and for this sequence, $F(e_n{}^w) = \mathcal{H}$ with $w_1, w_2 \in \mathcal{H}$, contradicting (a). Thus no reliable F exists.

A necessary and sufficient condition for reliability in the limit can now be stated:

Theorem 3. (F is reliable in the limit for $\mathcal{H}$) $\Leftrightarrow$ ($\mathcal{H} \subseteq \Sigma^B{}_2$)
Proof: ($\Rightarrow$) Theorems 1 and 2. ($\Leftarrow$) Kelly (1996), pg. 92.[34]

Now let EI ("empirical indistinguishability") denote the negation of ED. Theorem 2 indicates that when EI holds, evidence cannot decide in the limit on the correct hypothesis. Now suppose that empirical claims are expressed in the language in which the evidence is expressed whereas theoretical claims are expressed in a language that outstrips the evidence language. Theorem 2 then indicates that, if the existence of a reliable method of inquiry is a necessary condition for belief, then we can have good reasons for believing (some) empirical claims, but we can never have good reasons for believing theoretical claims. If one prefers the distinction in kinds between unobserved observables and in principle unobservables, Theorem 2 indicates that inferences to the

[34]It should be noted that this result holds only for hypotheses stated in languages with monadic predicates. For languages with more complex vocabulary, membership in $\Sigma^B{}_2$ may fail to be sufficient for reliable identification in the limit. Kelly (1996, pg. 300) and Kelly and Glymour (1989, pg. 198) demonstrate, for example, that for data restricted to 2-place predicates only, no reliable F exists for $\mathcal{H}$'s in $\Sigma^B{}_2$. (Such hypotheses are of the general form $\exists x \forall y Rxy$, for some 2-place predicate R.)

former (type A) are warranted whereas inferences to the latter (type B) are not. This nicely encapsulates the claim of the anti-realist who employs the EUA in particular, and the PMI in general.

Reliability in the limit for structure identification

Can evidence pick out the correct EI-equivalence class of hypotheses in the limit? This would be enough to ground inferences of type (C). A definition of reliability in the limit for structures $\mathfrak{H} = \{\mathcal{H}_1, \mathcal{H}_2, ...\}$, where the elements are EI $\mathcal{H}$'s is needed. To this end, let $e^G = (E_1^{\ G}, E_2^{\ G}, ...)$ be a data sequence where $E_k^{\ G}$ stands for E_k or $\sim E_k$ depending on whether $\forall w_i \in$ G, $w_i \vDash E_k$ or $w_i \vDash \sim E_k$, for G a $\sim\mathcal{E}$-*sep* equivalence class of worlds. Let $e_n^{\ G}$ denote the finite sequence that ends at $E_k^{\ G}$.

Def. 5. The method $\mathfrak{F} : \{e_n^{\ G}\} \rightarrow \mathfrak{H}$ is *reliable in the limit for* $\mathfrak{H}$ *with respect to* $\mathcal{E}$ *and* $\mathcal{W}$ just when

$$\forall \mathcal{H}_i \in \mathfrak{H}, \forall w \in \mathcal{W}, \forall e^G \in \mathcal{DT}_{\mathcal{E}}(G), \exists N \text{ such that } \forall n \geq N,$$
$$(\mathfrak{F}(e_n^{\ G}) = \mathcal{H}_i) \Leftrightarrow \mathfrak{C}(e^G, \mathcal{H}_i),$$

where $\mathfrak{C}(e^G, \mathcal{H}_i)$ just when $\exists \mathcal{H}_j \in \mathcal{H}_i$ such that $C(\mathrm{e}^{\mathrm{w}}, \mathcal{H}_j)$, $\forall w \in G$.

A few comments concerning Def. 5 are in order. First, it does not require that we identify which H_j is C-correct. This is crucial insofar as it leaves open the possibility for the existence of a reliable $\mathfrak{F}$ in cases in which there is no reliable F. Second, to say that $\mathfrak{H}$ is identified by $\mathfrak{F}$ means just that, if $\mathcal{H}$ is $\mathfrak{C}$-correct (*i.e.*, if it contains a C-correct H), then there will be a point in the data stream after which $\mathfrak{F}$ posits $\mathcal{H}$ forever.

It should now be possible to characterize the type of $\mathcal{H}$ that can be reliably identified. To this end, call two $\sim\mathcal{E}$-*sep* equivalence classes of worlds G_i and G_j $\mathfrak{H}$-*sep* just when they belong to different EI partitions $\mathcal{H}_i$, $\mathcal{H}_j$:

Def. 6. $\forall G_i, G_j \in \mathrm{Po}(\mathcal{W})$, $\mathfrak{H}$-*sep*(G_i, G_j) just when

$$\exists \mathcal{H}_i, \mathcal{H}_j \in \mathfrak{H}, (i \neq j), \text{ such that } G_i \subset \mathcal{H}_i, G_j \subset \mathcal{H}_j .$$

Then we have the natural extension of Def. 4:

Def. 7. $\mathfrak{H}$ is *empirically distinguishable (ED*) with respect to* $\mathcal{E}$ *and* $\mathcal{W}$ just when

$$\forall G_i, G_j \in \mathrm{Po}(\mathcal{W}), \mathfrak{H}\text{-}sep(G_i, G_j) \Rightarrow \mathcal{E}^*\text{-}sep(G_i, G_j).$$[35]

In words, $\mathfrak{H}$ is ED* if there is a piece of evidence that separates $\sim\mathcal{E}$-*sep* equivalence classes of worlds G_i, G_j belonging to different EI partitions $\mathcal{H}_i$, $\mathcal{H}_j$ of $\mathcal{W}$. The extension of Theorem 2 then is:

Theorem 4. ($\mathfrak{F}$ is reliable in the limit for $\mathfrak{H}$) $\Rightarrow$ ($\mathfrak{H}$ is ED*)

Proof: Similar to Theorem 2.

[35] Where $\mathcal{E}^*$-*sep*(G_i, G_j) just when $\mathcal{E}$-*sep*(w_i, w_j), for $\forall w_i \in G_i, \forall w_j \in G_j$. Note that $\mathcal{E}^*$-*sep* holds for all G_i, G_j of the same EI partition $\mathcal{H}$. It may fail to hold for G's belonging to distinct EI partitions.

Insofar as it is not the case that (ED*) ⇒ (ED), it is possible for there to be a reliable $\mathfrak{F}$ in instances in which no reliable F exists. I submit that this possibility is all that is needed to establish the structural realist's epistemic claim that inferences of type (C),

(i) can be defined by means of reliable methods, and

(ii) can be reliable in contexts in which inferences of type (A) are and type (B) are not.

The idea above has simply been to show that one can define a reliable method of hypothesis identification F that identifies the correct hypothesis H_j among a group of competing hypotheses $\mathcal{H} = \{H_1, H_2, ...\}$ solely on the basis of empirical data. In addition, F has the property that it fails to identify a correct hypothesis if the members of $\mathcal{H}$ are empirically indistinguishable. What the above demonstrates is that, in this latter situation, an extension $\mathfrak{F}$ of F can still function to pick out the correct group $\mathcal{H}_i$ of empirically indistinguishable hypotheses from a collection of competing groups $\mathfrak{H} = \{\mathcal{H}_1, \mathcal{H}_2, ...\}$. Such a collection, for example, might be given by competing theories of gravity: $\{\{NG_N, NG_C\}, \{GR, A_1, A_2, ...\}, \{QG_1, QG_2, ...\}\}$. Here the groups consist of empirically indistinguishable classical non-relativistic theories, classical relativistic theories and quantum relativistic theories.[36] So, while there is no reliable method that can choose between the members of the sets $\{NG_N, NG_C\}$, $\{GR, A_1, A_2, ...\}$, $\{QG_1, QG_2, ...\}$, there is a reliable method that can tell us which of the sets has a member that is correct in the limit.

Admittedly, this demonstration is based on numerous simplifying assumptions. In particular, it would be beneficial to construct a definition of a reliable method of structure identification that is not restricted to hypotheses stated in languages with monadic predicates. Certainly a more realistic method would take account of the fact that hypotheses usually come in more complex forms. However, this should not detract from its use as a counterexample against the type of anti-realism that claims that there are no reliable inferences to theoretical claims in general, solely on the basis that there are no reliable inferences to in-principle unobservables in particular. Insofar as the EUA and the PMI are based on such a claim, they have been shown to be incorrect.

[36] In the latter two cases, of course, it is not clear if the groups have more than one member, or any member. A_1, A_2, ... denote EI alternatives to GR, while QG_1, QG_2, denote EI theories of quantum gravity.

References

Bain, J. (2006) 'Spacetime Structuralism', in Dieks, D. (ed.) *The Ontology of Spacetime, Vol. 1*, Elsevier Press, pp. 37-66.

Bain, J. (2004) 'Theories of Newtonian Gravity and Empirical Indistinguishability', *Studies in History and Philosophy of Modern Physics* **35**, pp. 345-376.

Bain, J. (1998) 'Weinberg on QFT: Demonstrative Induction and Underdetermination', *Synthese* **117**, pp. 1-30.

Bain, J. and J. D. Norton (2001), 'What Should Philosophers of Science Learn from the History of the Electron?', in Buchwald, J. and A. Warwick (eds.), *Histories of the Electron: The Birth of Microphysics*, Cambridge: MIT Press, pp. 451-465.

Belot, G. (1995) 'Whatever is Never and Nowhere is Not', Ph.D. dissertation, University of Pittsburgh.

Carrier, M. (1993) 'What is Right with the Miracle Argument: Establishing a Taxonomy of Natural Kinds', *Studies in History and Philosophy of Science* **24**, pp. 391-409.

Day, T. and H. Kincaid (1994) 'Putting Inference to the Best Explanation in its Place', *Synthese* **98**, pp. 271-295.

Earman, J. (1993) 'Underdeterminism, Realism and Reason', in P. French, T. Uehling, Jr. and H. Wettstein (eds.), *Midwest Studies in Philsophy, XVIII*, Notre Dame: University of Notre Dame Press, pp. 19-38.

Earman, J. (1992) *Bayes or Bust*, Cambridge: MIT Press.

Fine, A. (1986) 'Unnatural Attitudes: Realist and Instrumentalist Attachments to Science', *Mind* **95**, pp. 149-79.

Fine, A. (1984) 'The Natural Ontological Attitude', in J. Leplin (ed.), *Scientific Realism*, Berkeley: University of California Press, pp. 83-107.

French, S. (1999) 'Models and Mathematics in Physics: The Role of Group Theory', in J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, Cambridge: Cambridge University Press, pp. 187-207.

Friedman, M. (1983) F*oundations of Spacetime Theories*, Princeton: Princeton University Press.

Glymour, C. (1977) 'Indistinguishable Space-Times and the Fundamental Group', in J. Earman, C. Glymour and J. Stachel (eds.), *Foundations of Space-Time Theories*, Minneapolis: University of Minnesota Press, pp. 50-60.

Horwich, P. (1991) 'On the Nature and Norms of Theoretical Commitment', *Philosophy of Science* **58**, pp. 1-14.

Horwich, P. (1986) 'A Defense of Conventionalism', in G. MacDonald and C. Wright (eds.), *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*, New York: Blackwell, pp. 163-187.

Horwich, P. (1982a) 'How to Choose Between Empirically Indistinguishable Theories', *Journal of Philosophy* **2**, pp. 61-77.

Horwich, P. (1982b) 'Three Forms of Realism', *Synthese* **51**, pp. 181-201.

Kelly, K. (1996) *The Logic of Reliable Inquiry*, Oxford: Oxford University Press.

Kelly, K. and C. Glymour (1992) 'Inductive Inference from Theory Laden Data', *Journal of Philosophical Logic* **21**, 391-444.

Kelly, K. and C. Glymour (1989) 'Convergence to the Truth and Nothing But the Truth', *Philosophy of Science* **56**, 185-220.

Kelly, K., C. Juhl and C. Glymour (1994) 'Reliability, Realism, and Relativism', in P. Clark (ed), *Reading Putnam*, London: Blackwell.

Kitcher, P. (1993) *The Advancement of Science*, Oxford: Oxford University Press.

Kukla, A. (1996) 'Antirealist Explanations of the Success of Science', *Philosophy of Science* **63**, pp. S298-S305.

Kukla, A. (1994) 'Scientific Realism, Scientific Practice, and the Natural Ontological Attitude', *British Journal for Philosophy of Science* **45**, pp. 955-975.

Ladyman, J. (1998) 'What is Structural Realism?', *Studies in History and Philosophy of Science* **29**, pp. 409-424.

Laudan, L. and J. Leplin (1991) 'Empirical Equivalence and Underdetermination', *Journal of Philosophy* **88**, pp. 449-472.

Leplin, J. (1997) *A Novel Defense of Scientific Realism*, Oxford: Oxford University Press.

Malament, D. (1977) 'Observationally Indistinguishable Space-times', in J. Earman, C. Glymour and J. Stachel (eds.), *Foundations of Space-Time Theories*, Minneapolis: University of Minnesota Press, pp. 61-80.

Martin, C. (2002) 'Gauge Principles, Gauge Arguments, and the Logic of Nature', *Philosophy of Science* **69**, pp. S221-S234

Norton, J. (1994) 'Science and Certainty', *Synthese* **99**, pp. 3-22.

Norton, J. (1993) 'The Determination of Theory by Evidence: The Case for Quantum Discontinuity, 1900-1915', *Synthese* **97**, pp. 1-31.

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

Putnam, H. (1978) *Meaning and the Moral Sciences*, London: Routledge.

Redhead, M. (1999), 'Quantum Field Theory and the Philosopher', in T. Cao (ed.), *Conceptual Foundations of Quantum Field Theory*, Cambridge: Cambridge University Press, pp. 34-40.

Ruetsche, L. (2002), 'Interpreting Quantum Theories', in P. Machamer and M. Silberstein (eds.), *The Blackwell Guide to the Philosophy of Science*, Malden: Blackwell, pp. 199-226.

Salmon, W. (1966) *The Foundations of Scientific Inference*, Pittsburgh: University of Pittsburgh Press.

Stein, H. (1989) 'Yes, but... Some Skeptical Remarks on Realism and Anti-Realism', *Dialectica* **43**, pp. 47-65.

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

Ward, R. (1981), 'The Twistor Approach to Differential Equations', in C. Isham, R. Penrose, D. Schiama (eds.), *Quantum Gravity 2*, Oxford: Clarendon Press, pp. 593-610.

Ward, R. S. (1977) 'On Self-Dual Gauge Fields', *Physics Letters* **61A**, pp. 81-82.

Ward, R. S. and R. Wells (1990), Twistor Geometry and Field Theory, Cambridge: Cambridge University Press.

Worrall, J. (1994) 'How to Remain (Reasonably) Optimistic: Scientific Realism and the "Luminiferous Ether", in D. Hull, M. Forbes and R. M. Burian (eds.), *PSA 1994 Vol. 1*, East Lansing: Philosophy of Science Association, pp. 334-342.

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