Evidential Equivalence

1. Introduction

The thesis from the underdetermination of theory by evidence (UTE) takes many forms.[1] What they all have in common is the insufficiency of any given body of evidence to uniquely determine the truth of a theory. Unsurprisingly then, UTE supports anti-realist accounts, for it says that no matter how much evidence we amass we can never uphold any theory as true. Empiricist accounts, in particular, embrace UTE since they urge belief only in a theory's empirical, or as it is otherwise called 'observational', content. Typically, they urge belief only in the truth of the observational consequences of a theory.[2] We can now formulate an empiricist version of UTE:

(EUTE): Any body of observational evidence is insufficient to uniquely determine the truth of the theoretical (i.e. non-observational) consequences of a theory.

Along with EUTE, empiricists embrace the thesis of empirical equivalence. We say that two or more theories are *empirically equivalent* if and only if they entail the same observational sentences.[3] The empirical equivalence thesis (EE) holds that there are empirically equivalent theories to any given theory. A stronger, and more popular, formulation of the thesis holds that there are empirically equivalent *rival* theories to any given theory. The underdetermination and the empirical equivalence theses have also been formulated in terms of systems (see (Quine 1975)). A system can be thought of as a unit broader than a theory that includes one or more of the following: other theories, auxiliary hypotheses, models, methodological principles, beliefs, etc.

Though rarely explicitly stated, the theses of underdetermination and empirical equivalence are seen as either identical or at least logically equivalent.[4] Provided that the notions of observational consequence and observational evidence coincide –certainly a controversial assumption – a proof of EUTE's and EE's logical equivalence can be given:[5] Suppose that a given body of observational evidence underdetermines a theory *T*. This in effect means that the observational evidence cannot discriminate between *T* and at least one other theory, let us call it '*T**'. In other words, *T* and *T** are empirically equivalent. Now for the converse relation. Suppose that *T* and *T** are empirically equivalent. This just means that the observational consequences of T (and hence T*) cannot discriminate between the two theories, i.e. the observational evidence underdetermines the two theories.

UTE in its various guises has come under heavy fire. In an influential article, Larry Laudan and Jarrett Leplin (1991) present a two-pronged critique of those underdetermination arguments that rely on the notion of empirical equivalence.[6] On the one hand, they question the view that *all* theories have genuine empirically equivalent rivals. On the other, they argue that, even when theories have such rivals, there are still justifiable ways to choose between them, for: (1) a theory is not necessarily supported by the empirical consequences it entails and (2) a theory can be supported by evidence that it does not entail.

In this article, I probe the consequences and limits of the underdetermination and the empirical equivalence theses, using Laudan and Leplin’s fecund article as a springboard. Although a realist at heart, my primary intention is not to undermine the anti-realist arguments but rather to try to precisify the challenge the realist, and more generally the participant in the scientific realism debate, faces. Let us start by looking at the two prongs of Laudan and Leplin’s argument, one at a time.

2. Does Every Theory have Empirically Equivalent Rivals?

Laudan and Leplin cite three theses which, when taken together, allegedly raise doubts about empirical equivalence. These are:

(1) What is observable varies through time.

(2) The derivation of observable consequences typically requires auxiliaries.

(3) Auxiliaries vary through time.

On the basis of these, they argue that the observational consequences of a theory are not fixed but vary over time, and conclude that they are not clearly identifiable and that empirical equivalence is, therefore, defeasible. In their own words:

> … any determination of the empirical consequence class of a theory must be relativized to a particular state of science. We infer that empirical equivalence itself must be so relativized, and, accordingly, that any finding of empirical equivalence is both contextual and defeasible (p. 454).

One immediate reply to this argument takes the following form: Whether the observational consequences of a theory are fixed or vary over time is a matter independent of the empirical equivalence thesis. We can capture the variability of a theory's observational consequences over time by saying that out of a class *Q* which contains *n* sets of observation sentences as members, i.e. O_1, O_2, …, O_n, the theory (plus any auxiliaries) at any time *t* entails just one set, though we may not know which one. If the empirical equivalence thesis holds, then *whatever* the set of observational consequences of a given theory (plus auxiliaries), other theories (plus auxiliaries) will also entail that, and only that, set of observational consequences. That is, if the empirical equivalence thesis holds, then any member O_i of class *Q* will be entailed by more than one theory or system.[7] Thus, raising doubt about the stability of a theory's observational consequences is neither here nor there when it comes to the truth of the underdetermination or empirical equivalence theses.

A similar point is made by André Kukla, who argues that Laudan and Leplin only manage to force their opponents to relativise the notion of empirical equivalence so that it is construed as a relation between indexed theories, "i.e. triplets consisting of a theory, a partitioning of phenomena into observables and non-observables, and a specification of the permissible auxiliaries" (Kukla 1993, p. 2). In their reply to Kukla, Laudan and Leplin charge that recourse to changing rivals alone spells doom for the thesis of empirical equivalence. Their reasoning seems to be that when a theory *T* is successively compared with different rivals, each of these rivals is not strictly speaking empirically equivalent to *T*, for if it were we would stop looking for other rivals to compare. In short, *T* does not really have empirically equivalent rivals.

Laudan and Leplin's defence fails. True, *T*'s comparison to new rivals results from the collapse of its presumed empirical equivalence to old rivals. But does empirical equivalence collapse altogether? Perhaps Laudan and Leplin implicitly assume that at some point the rivals run out. This assumption is question begging, for that is the issue under debate, i.e. whether such rivals always exist. More damningly, their defence misleads the reader into thinking that the comparison is between a stable *T* and changing rivals whereas in actuality it is between *T* plus varying auxiliaries and changing rivals. A theory *T* augmented with different auxiliaries yields different systems S_1, S_2, S_3, … S_n. Naturally, a rival S_1', if empirically equivalent to S_1, will not be empirically equivalent to S_2, unless of course S_1 and S_2 are empirically equivalent. For Laudan and Leplin's defence to be successful, they need to tell us why they think some of these systems do not have rivals.

Back to their original article, Laudan and Leplin attempt to pre-empt similar concerns.[8] Their argument consists of two strategies. *First*, to reject the view that there exists an algorithm that can generate empirically equivalent rivals for any given theory. *Second*, to deny that the cases offered as examples of empirically equivalent theories are genuine.

I will begin with the second claim since it can be more easily dismissed. In pursuit of the second strategy, Laudan and Leplin modify one of van Fraassen's examples in *The Scientific Image*, and show how it fails to be a case of empirical equivalence. We need not delve into the details. Showing that the two theories under consideration are not empirically equivalent is, of course, a correct step on the path to showing that examples of empirically equivalent theories are not genuine. It is, however, a far cry from showing that neither theory has *any* empirically equivalent rivals. It is this latter claim that needs to be established in order for Laudan and Leplin's conclusion to go through, namely if there is at least one theory that has no empirically equivalent rivals, then, obviously, not all theories have empirically equivalent rivals.

In what seems to me a change of heart, Laudan and Leplin return to this point and claim that their original argument merely "gave reasons for regarding *apparent* cases of empirical equivalence as potentially adjudicable, so that such observational indiscriminability as we do encounter should not be parlayed hastily into EE [empirical equivalence]" (1993, p. 10) [original emphasis]. I wholeheartedly agree that we should be cautious when branding pairs of theories 'empirically equivalent'. The real issue, however, is not whether we can find convincing historical examples but rather whether empirically equivalent rivals exist for every theory. If they do exist, finding them may be, for any number of reasons, near impossible but that does not make them any less threatening to realism.

Algorithmically Produced Rivals

In pursuit of the first strategy, Laudan and Leplin examine the prospects of algorithms that reduce theories to their instrumentalist counterparts, and claim that these would invariably fail.[9] There are at least two problems with Laudan and Leplin's approach. Firstly, such algorithms do not suffice to produce rival theories, for they need to be augmented with a mechanism that expands the instrumentalist counterpart of the original theory into the theoretical vocabulary. Without this expansion, what we have is just an instrumentalist counterpart but *no theories* to make empirical equivalence appraisals. The expansion must be conducted in different ways so as to yield inconsistent, i.e. presumably rival, theories.[10] Unsurprisingly, after a short discussion of the limitations of instrumentalist algorithms, Laudan and Leplin state the obvious, i.e. "what application of an instrumentalist algorithm to a theory produces is manifestly not an alternative *theory*" (1991, p. 456) [original emphasis].

Secondly, after their rejection of instrumentalist algorithms, they hastily conclude: "We know of no algorithm for generating genuine theoretical competitors to a given theory" (1991, p. 457). Laudan and Leplin start this section of their paper claiming they can defeat the idea that there exist empirically equivalent theories for any given theory. All they end up showing, however, is that they do not know (and cannot themselves devise) an algorithm that produces such theories. As we have seen, the only type of algorithm they consider, viz. instrumentalist algorithm, is not even a suitable candidate, for it produces instrumentalist versions of theories and not genuine rival theories. It should be quite obvious that our inability to devise an algorithm that produces empirically equivalent theories does not entail that such theories do not exist.[11]

This point can be supported further by a lesson learned in the theory of recursion. Two definitions will help us here: (1) We say that a set *A* is *recursive* (or *decidable*) if there exists a Turing machine, i.e. an algorithm, which after a finite number of steps can decide whether or not any given object is a member of *A*. (2) We say that a set *A* is *recursively enumerable* if it can be stated in sequence form, i.e. $A = \{a_1, a_2, a_3, \ldots\}$, by a Turing machine. All recursive sets are also recursively enumerable but not vice-versa. A halting set, for example, is recursively enumerable but not recursive.[12] Some sets are not even recursively enumerable, e.g. the set of all Gödel numbers of non-theorems of first-order logic. The moral of the story is that these are non-empty sets that cannot be specified algorithmically. Likewise, in our context, should it turn out that there is no algorithm for producing empirically equivalent rivals, the conclusion need not be that sets containing such rivals are empty.

A Recipe for Algorithms

What makes a good recipe for algorithms? Let's start with a simple algorithm. We take an existing theory and add to it a hitherto un-included theoretical claim. Suppose we have a theory X. We add to it theoretical claim T_1, making sure that X does not contain or entail T_1. The result X&T_1 is a new theory 'rivalling' the old one. We can repeat this process indefinitely, each time adding a different theoretical claim that is not included in the original theory.

Obviously, this method is grossly inadequate. One major problem is that it does not guarantee empirical equivalence. To rectify this, we must require that the new theoretical claim does not, when taken together with the original theory, affect its observational consequences. I can think of three different types of theoretical claims that might conceivably satisfy this condition:

(1) Theoretical claims that have no observational consequences whatsoever.
(2) Theoretical claims whose observational consequences are already amongst the consequences of the original theory.

(3) Theoretical claims that have observational consequences only when conjoined to other statements, none of which are entailed by the original theory.

The first type of theoretical claim seems straightforward enough. Are there any examples of such claims? Take Newton's notion of absolute space. According to Newton, "[a]bsolute space, of its own nature without reference to anything external, always remains homogeneous and immovable" ([1726]1999, p. 410). Many doubt whether the sentence asserting the existence of absolute space has any observational consequences.[13] It is, however, always possible to construct conditional sentences that endow observational consequences to theoretical claims of type one. In the example just mentioned, such a conditional sentence would take the form 'If there is absolute space, then P', where P must be an observation sentence. This just means that no theoretical claims of the first type need be considered since these collapse to claims of type three. That is, we can always find additional sentences that when conjoined with any theoretical claim produce observational consequences.

The second type of theoretical claim comprises those that have observational consequences already entailed by the theory in question. This preserves the original set of observational consequences, and thus the empirical equivalence is not violated. It may be objected that if the extra theoretical claim does not contribute any new observational consequences then why append it to the original theory in the first place. In particular, it may be argued that these claims can be rejected on account of parsimony. Though intuitively sound, the objection does not eliminate all theoretical claims of this type, since such claims can be desirable for reasons independent of their observational consequences.[14] For example, a theoretical claim of this type may be explanatory in some way that the original theory is not. Given that realists value the explanatory power of theoretical claims, they would find it difficult to dismiss this possibility.

That some theoretical claims have observational consequences only when conjoined to other statements is a point that bodes well with a weaker version of Duhem's thesis, according to which *some* theories/theoretical claims cannot be tested in isolation for they do not have observational consequences on their own. The third type of theoretical claim noted above requires that none of the observational consequence-inducing statements be entailed by the original theory. More formally: Let T be the original theory and T_i' the theoretical claims to be added. T_i' are theoretical claims of the third type when: 1) they have no observational consequences unless accompanied by theoretical claims T_k'' and 2) $T \nvDash T_k''$. In such cases, the conjunction $T\&T_i'$ is empirically equivalent to T.[15]

In their original article, Laudan and Leplin do not really elaborate what characteristics a rival theory should have to be considered genuine. They mention the obvious, namely that rivals must be incompatible.[16] Their only other remarks on this issue are obscure: "As we do not question the empirical equivalence of logically equivalent theories, we ignore this suggestion and assume henceforth that theories whose empirical equivalence is at issue are logically and conceptually distinct" (1991, p. 455). The claim about conceptual distinctness also pops up in their reply to Kukla, but alas with no explanation affixed.

So far, the focal point of the discussion was producing empirically equivalent rivals by adding theoretical claims to existing theories. This allows the production of logically inequivalent rivals but not the production of incompatible rivals.[17] We thus come to the question: Does every theory have *logically inconsistent* rivals that are nonetheless empirically equivalent?

One way to produce such rivals involves replacing – instead of adding – theoretical claims with claims incompatible with them. Like before, eligible theoretical claims can take one of the following forms:

(1) Theoretical claims that, (a) have observational consequences but only when conjoined to other statements none of which are included in the original theory, or (b) have observational consequences already contained in the main theory. These must be replaced with theoretical claims that are either of form (a) or (b), and that result in a theory that is logically incompatible to the original.[18]

(2) Theoretical claims with observational consequences that, if removed, will alter the observational consequence set of a theory must be replaced with theoretical claims that will return the set to its original state and will result in a theory logically incompatible to the original.

History might *never* produce such rivals, but, as we saw earlier, the point about underdetermination can be made independently of historical examples. The above method of constructing empirically equivalent theories that are logically incompatible is quite straightforward: Replace theoretical claims so that 1) the resulting theory is incompatible with the original theory, and 2) the observational consequence set remains untouched.

Quine (1975) has pointed out that to be a genuine rival it is not sufficient to be logically incompatible. This is a consequence of the fact that one may have logically incompatible theory formulations of the same theory. By way of example, he offers the following algorithm: If we take a theory formulation of any scientific theory and consistently switch any two of its theoretical terms, we end up with an incompatible (yet empirically equivalent) theory formulation of the same scientific theory.[19] We can take this qualification on board requiring from our algorithms, in addition to the above two conditions, that any potential rival is not a merely terminological variant of (or as Quine would say 'a theory formulation of') the theory in question.

What I have just sketched is a recipe for one type of potentially successful algorithms. The recipe consists of two conditions that algorithms must meet, namely 1) replace theoretical claims so that a) the resulting theory is incompatible with the original theory, and b) the observational consequence set remains untouched and 2) make sure that the theoretical incompatibility is not mere terminological variance. Whether there are successful algorithms of this type is a question that I cannot pursue here. My aim was simply to probe deeper into what is required for an algorithm to succeed. Similarly, I do not have an answer to the question upon which this whole section was framed. It is far from clear whether every theory has empirically equivalent rivals. What is clear is that Laudan and Leplin's arguments do not help the realist. The first prong of their two-pronged critique of underdetermination arguments that rely on the notion of empirical equivalence fails. It is now time to turn to the second prong of their critique.

3. Can we Justifiably Choose between Empirically Equivalent Theories?

Laudan and Leplin claim that even if two theories are empirically equivalent, we can still justifiably choose between them. Specifically, they hold that (1) the observational consequences of a theory need not provide support for it, and (2) a theory may be supported by evidence that does not form part of its observational consequences.

In support of their first claim, Laudan and Leplin simply point out that the fact that *H* entails *e* does not necessarily mean that *e* confirms *H*. This seems correct but, as it stands, does little or nothing to overcome the problem at hand. One need only restrict a theory's observational consequence set to a subset containing as members all and only those observational statements that have the power to confirm. Let us call this the 'confirmational consequence set'. Laudan and Leplin offer no reason to suppose that a theory has no co-confirmational rivals, i.e. theories whose confirmational consequence set is the same.[20] Presumably, the fact that not all empirically equivalent theories will be co-confirmational downsizes (and in ideal cases eliminates) the competition. The snag is that any gains made here might be lost, or, even worse, losses might exceed gains, since under the current suggestion one also needs to worry about empirically inequivalent yet co-confirmational theories. That is, there might be even more empirically inequivalent but co-confirmational theories than there are empirically equivalent but not co-confirmational ones. All that Laudan and Leplin's point achieves is to trade one empirical equivalence class for another, or, more precisely, to trade an empirical equivalence class for a co-confirmation class.[21]

In support of their second claim they make use of the following argument. Suppose that hypothesis *H* entails evidence *e* and that *e* confirms *H*. Evidence that confirms a hypothesis *H* will also confirm (a) any theory *T* that entails *H* and (b) any other hypothesis H_k entailed by *T*. The point is that while *H* entails *e*, H_k need not. In other words, a hypothesis may be supported by evidence that it does not entail. Here's a reconstruction of the form of argument that Laudan and Leplin sanction (see (1991, p. 464)):

For any *i*, *j*, and *k*:

$H_i \vdash e$ (1) - premise

e confirms H_i (2) - premise

$T_j \vdash H_i$ (3) - premise

$\therefore e$ confirms T_j (4) - by CCC (see below)

$T_j \vdash H_k$ (where $k \neq i$) (5) - premise

$H_k \nvdash e$ (6) - premise

$\therefore e$ confirms H_k (7) - by SCC (see below)[22]

Samir Okasha (1997) has correctly criticised this form of argument by saying that it relies on two principles that Hempel showed to be incompatible, namely the 'converse consequence condition' (CCC) and the 'special consequence condition' (SCC).[23] According to CCC, if some evidence confirms a statement *S* it also confirms any statement *S′* that entails *S*. According to SCC, if some evidence confirms a statement *S*, it also confirms any statement *S′* that *S* entails. Hempel demonstrated that SCC and CCC *can*, when used together, lead to absurdity. The following argument is an example of how the principles can be used to derive confirmation for *any* theory:

H_i confirms H_i	(a) - self-evident[24]
$(T_j \& H_i) \vdash H_i$	(b) - self-evident
H_i confirms $T_j \& H_i$	(c) - by CCC
$(T_j \& H_i) \vdash T_j$	(d) - self-evident
$\therefore H_i$ confirms T_j	(e) - by SCC

On the basis of CCC and SCC, this argument shows that anything can confirm anything, an obviously absurd result.

To Okasha's critique I want to add that, even if employed on their own, the two principles *can* lead to incorrect inferences. It is well known that employing CCC on its own still allows us to derive that evidence which confirms a hypothesis will confirm *any* theory, no matter how ridiculous, that entails the hypothesis. Similarly, employing SCC on its own allows us to derive that evidence which confirms a theory will confirm *any* hypothesis, no matter how ridiculous, that is entailed by the theory.

Although as inference rules CCC and SCC fail to help Laudan and Leplin's point, the question remains whether a theory can be supported by evidence that is not part of its observational consequences. There are two ways to answer this question in the affirmative. First, to show that observation statements not entailed by a given theory are evidentially relevant, i.e. have the power to confirm. Second, to show that the theory can be confirmed by non-observational evidence.

Take the first answer. The fact that more observational evidence (than that contained in the observational consequences of a theory) is on hand does not guarantee that the resulting equivalence class, now determined by the predicate 'being confirmed by the same observational evidence', will consist of only one member. Also, just because some such observation statements have the power to confirm does not mean that they can confirm the non-observational parts of the theory.[25] On its own, the claim that these observation statements are evidentially relevant precipitates a retreat from consequence-empiricism but does not support realism. The definition of empiricism at the beginning of the paper is what I here call 'consequence-empiricism', urging belief only in the truth of the observational consequences of a theory. Empiricists could always circumscribe their epistemic commitments in a different way so that, for example, they urge belief only in the truth of the evidentially relevant observation statements.[26] I am not going to evaluate the viability of this empiricism here. Suffice it to say that realists need additional arguments to the effect that evidentially relevant observation statements have the power to confirm the non-observational parts of the theory.

The second answer is more popular in the literature and presumably circumvents the difficulty just mentioned. Realists view theoretical virtues like unity, simplicity, explanatory power, and comprehensiveness as extra-observational evidence that can overcome claims of empirical equivalence (see, for example, Nelson (1996) and Psillos (1999)).[27] This is a trivial point once one accepts that the possession of the above virtues counts as (non-observational) evidence. In the Bayesian framework, for example, these theoretical virtues can be reflected in the choice of priors.[28]

Van Fraassen, a constructive empiricist, objects that the so-called 'theoretical virtues' are nothing but pragmatic features of theories with no epistemic significance (1980, pp. 87-89). In other words, he denies that these virtues have any evidential status, i.e. that they can confirm or disconfirm a theory. They can, of course, be used as pragmatic criteria for theory acceptance, but their role is restricted to just that. The lack of consensus on how to understand theoretical virtues certainly does not help the realists, since it motivates the suspicion that they might be merely conventional and/or pragmatic features of theory choice. Even more damaging is the insufficiency of evidence that nature is amenable to a unified, comprehensive, simple, and explanatorily powerful account. This latter view finds some proponents in the realist camp. Nancy Cartwright, for example, argues in favour of the disunity of science.[29]

4. Evidential Equivalence

There is thus a standoff over the epistemic import of theoretical virtues.[30] I will not enter into this dispute. Instead, I propose to look at some of the ramifications in case the dispute is resolved either way. Suppose theoretical virtues are evidentially irrelevant. In that case, we would say that a theory could not be confirmed or disconfirmed on the basis of theoretical virtues. Does that mean that only observational statements are evidentially relevant to a theory? The answer depends on whether anything other than theoretical virtues counts as evidentially relevant. One suggestion follows Timothy Williamson's thesis that knowledge equals evidence (see his (2000)). If the thesis holds, one might argue that non-empirical knowledge is evidence that is sometimes relevant in matters of theory choice. I will not pursue this line of thought here.

Suppose theoretical virtues are evidentially relevant. In that case, we would say that a theory could be confirmed or disconfirmed on a non-observational basis, i.e. we could decide between empirically equivalent theories on the basis of theoretical virtues. This effectively goes beyond the epistemic commitments of empiricist views like constructive empiricism. But is the issue of underdetermination settled? The answer to this question is not all that clear. Suppose T_1 and T_2 are empirically equivalent, yet evidentially inequivalent. What guarantees do we have that there are no other theories, e.g. T_3 and T_4, which are evidentially equivalent to T_1 and T_2 respectively? Perhaps there are evidentially equivalent rivals to every theory. We can formalise this worry in the following way:

(EVE) $\forall T\ \exists T'\ [((T' \vdash \neg T)\ \&\ T'$ is not a terminological variant of T) & T′ is evidentially equivalent to T].[31]

Even if empirical equivalence claims can be defeated by appeal to theoretical virtues, there remains the issue of evidential equivalence, or EVE for short. If EVE is false, there are at least some theories that the evidence can uniquely identify. It is noteworthy that the truth of a theory is not guaranteed simply because it has no evidential equivalents. If EVE is true, then underdetermination remains rife, albeit in a restricted form that no longer supports views like constructive empiricism. The difference between the two truth-values consists in the extent to which underdetermination can be mollified.

Breaking through empirical equivalence means making some headway towards narrowing down theory choice but not necessarily defeating underdetermination altogether. To defeat empiricist views such as constructive empiricism, it suffices to provide non-observational constraints that mitigate the effects of underdetermination on theory choice. To defeat the most radical form of underdetermination, the following thesis must be shown to be true:

(EVE2) $\neg\ \exists T\ \exists T'\ [((T' \vdash \neg T)\ \&\ T'$ is not a terminological variant of T) & T′ is evidentially equivalent to T].

That is, if we want to defeat underdetermination entirely, we need to show that *no* theory has evidentially equivalent rivals, i.e. all theories are uniquely identified by some set of (observational and non-observational) evidence.

EVE2 is too strong a claim, since it requires all theories to lack evidentially equivalent rivals. Realists, it might be argued, can achieve their aim of vanquishing underdetermination with the weaker claim that all *empirically adequate* theories lack evidentially equivalent rivals.[32] That is, they can achieve their aim if they can show the following sentence to be true:

(EVE3) $\neg \exists T \exists T'$ [((((T′⊢¬T) & T′ is not a terminological variant of T) & (T′ is evidentially equivalent to T)) & T is empirically adequate].

Suppose EVE3 is true. What follows? Once scientific inquiry arrives at an empirically adequate theory, this theory will have no evidentially equivalent rivals. Since EVE3 refers only to empirically adequate theories, theories that do not possess this trait may well have evidentially equivalent rivals. To the extent that all of our current theories are at best approximately empirically adequate, establishing EVE3 is of no immediate help to the issue of underdetermination. One solution is to take the bull by the horns, i.e. to accept that current theories might have such rivals, hoping that one day our theories will reach empirical adequacy. Another solution is to require that approximately empirically adequate theories lack evidentially equivalent rivals:

(EVE4) ¬ ∃T ∃T′ [((((T′⊢¬T) & T′ is not a terminological variant of T) & (T′ is evidentially equivalent to T)) & T is at least approximately empirically adequate].

Perhaps this version is also too strong since some approximately empirically adequate theories might have evidentially equivalent rivals. Here's another suggestion:

(EVE5) ∃T ¬ ∃T′ [((((T′⊢¬T) & T′ is not a terminological variant of T) & (T′ is evidentially equivalent to T)) & T is at least approximately empirically adequate].[33]

If true, this says that there are some at least approximately empirically adequate theories that have no evidentially equivalent rivals.

EVE5 presents a more accurate formulation of the underdetermination challenge. Further refinements are probably required. More specifically, the central concepts of evidential equivalence, approximate empirical adequacy and theoretical rivalry need more unpacking. I hope that the above discussion will stimulate research in this area and hopefully a better understanding of the consequences and limits of the underdetermination, the empirical equivalence and the evidential equivalence theses.

5. Conclusion

To make sense of the broader picture here's a recapitulation of the main points:

- The stability of a theory's observational consequence set is irrelevant to the truth of the underdetermination or empirical equivalence theses.
- The lack of convincing empirically equivalent rivals from the history of science does little to undermine the empirical equivalence thesis. For this objection to have any potency it must be shown that empirically equivalent rivals, where they exist, are not hard to find.
- Our failure to conceive an algorithm that produces empirically equivalent rivals cannot be taken as a reliable indicator that such rivals do not exist.
- A recipe for one type of potentially successful algorithms specifies the following conditions algorithms must meet: 1) replace theoretical claims so that a) the resulting theory is incompatible with the original theory, and b) the observational consequence set remains untouched and 2) make sure that the theoretical incompatibility is not mere terminological variance.
- Laudan and Leplin's arguments against the empirical equivalence thesis fail and, as a consequence, cannot help the realist cause. The question whether every theory has empirically equivalent rivals remains open.
- It does not necessarily help the realist to hold that some of a theory's observational consequences need not provide support for it. Belief can always be limited to those observational consequences that have the ability to confirm. This forces a retreat from consequence-empiricism but not necessarily towards the direction of realism.
- That a hypothesis may be supported by evidence that it does not entail cannot be upheld via principles CCC and SCC. It can be upheld in two other ways: 1) by showing that observation statements not entailed by a given theory are evidentially

relevant, i.e. have the power to confirm and 2) by showing that the theory can be confirmed by non-observational evidence. The first option produces a retreat from consequence-empiricism, but, as before, not necessarily towards realism. To be successful, realists need to compellingly argue that evidentially relevant observation statements have the power to confirm the non-observational parts of the theory. For the second option to be successful, realists need to show that theoretical virtues are evidentially relevant to the truth of the non-observational parts of the theory.

- Supposing that these two options are successfully pursued, there are still well founded doubts as to whether the threat from underdetermination expires. The prospect of evidential equivalence and its associated brand of underdetermination raise these doubts. Theory choice may well remain indeterminate even after a potential defeat of the empirical equivalence thesis. For this reason we need to ascertain the truth-value of theses like EVE5, the thesis that there are some approximately empirically adequate theories that have no evidentially equivalent rivals.

Notes

[1] Michael Devitt (2002) offers an extensive discussion of UTE's various manifestations.

[2] This is a syntactic formulation of empiricism. One semantic formulation is given by constructive empiricism which at most urges belief in a theory's empirical adequacy, i.e. commitment that the empirical substructures of the theory (construed as a family of models) are isomorphic to all the phenomena.

[3] This is the standard formulation. See, for example, Boyd (2002) and Laudan and Leplin (1991). Quine (1975) adopts a similar definition, viz. empirically equivalent systems imply the same *observation conditionals*, roughly speaking generalised observation statements.

[4] For example, Devitt seems to think they are identical when he says "The underdetermination theses that concern us claim that a theory (belief) has rivals that stand in some sort of equivalence relation to it with respect to certain evidence" (2002, p. 26). Boyd (2002) makes similar remarks.

[5] I examine the assumption in section 3.

[6] Laudan and Leplin adopt the standard (syntactic) formulation of empirical equivalence but also offer a semantic version: "empirically equivalent theories have the same class of empirical models" (1991, p. 451). Though they focus on defeating the inference from empirical equivalence to underdetermination, the tenor of their claims suggests that they want to defeat underdetermination altogether. For example, they say: "…we shall argue that underdetermination does not in general obtain, not even under conditions of empirical equivalence" (1991, p. 460). In a follow-up article, Laudan and Leplin categorically state that the failure of empirical equivalence does not refute UTE but still manage to convey their disbelief in UTE (1993, p. 16). Although their denial of the logical equivalence between the theses of UTE and empirical equivalence is asserted without explanation, it is probably motivated by their arguments against the assumption that the notions of observational consequence and observational evidence coincide.

[7] The real question is whether such theories or systems are always going to be genuine rivals. I return to this question below.

[8] They say: "[t]he response we anticipate to our argument is a challenge to its assumption that empirical consequence classes must be identified for their empirical equivalence to be established" (Laudan and Leplin 1991, p. 455).

[9] Though they do not give any examples of instrumentalist algorithms, they may have something like the Ramsey-sentence in mind. Some philosophers, for example structural realists like Grover Maxwell and John Worrall, oppose this identification. They believe that the Ramsey-sentence of a theory *T* is not an instrumental reduction of *T*, but rather is *T* when properly construed. Indeed, they argue that the Ramsey-sentence of a successful theory *T* reveals certain structural features of the external world; hence the label 'structural realism'.

[10] As we will shortly see, to be a genuine rival requires more than just empirical equivalence and theoretical inconsistency.

[11] In the formulation of underdetermination arguments, there is no requirement that empirically equivalent alternatives must be produced first, but rather that such alternatives exist.

[12] A halting set is a recursively enumerable set consisting of all inputs on which a computer program halts.

[13] Stephen Brush and Gerald Holton go as far as to brand such claims meaningless (see ([1952]2001, p. 163). Naturally, supporters of Newton's bucket experiment claim that the sentence asserting the existence of absolute space has observational consequences, namely the concavity observed in the water when it is spinning together with the bucket.

[14] I am not claiming here that the reasons are epistemologically significant. They could very well be merely pragmatic.

[15] It might be objected that T_k'' can eventually be added to T in which case the resulting conjunction, i.e. $T\&T_i'\&T_k''$, will not be empirically equivalent to T. In reply, I would say that the empirical equivalence claim holds only between T and $T\&T_i'$, not between T and $T\&T_i'\&T_k''$.

[16] To this they subsequently add the condition that there should be "antecedent reason to regard [the incompatible theoretical account] as subject to unified explanation" (Laudan and Leplin 1993, p. 12). I will return to the condition of unified explanations in the next section.

[17] Adding a theoretical claim T_1 to a theory T does not always mean that the new theory $T\&T_1$ will be logically inequivalent to the original theory T. This is a simple logical point. The result of conjoining some sentence Q to some sentence P, i.e. P&Q, is not always logically inequivalent to P. Indeed, it is equivalent if and only if P entails Q.

[18] I treat the two types of theoretical claims discussed above jointly, because the empirical equivalence between original and modified theory will still hold even if the theoretical claim replaced is not of the same type.

[19] Quine's example switches the terms 'electron' and 'molecule'.

[20] Who has the burden of proof here is beyond me.

[21] To be fair, if their point is correct, it would require a modification of empiricism, at least as it was presented in the introduction. See the discussion on 'consequence-empiricism' below.

[22] N.B.: Lines 1-4 count as additional premises to the second argument.

[23] Hempel (1945a, 1945b) opts for SCC over CCC.

[24] This assumes that a statement can confirm itself. Another justification for this line is Hempel's entailment condition which holds that if S_1 entails S_2 then S_1 confirms S_2. H_i entails itself, hence by the entailment condition H_i confirms itself.

[25] Indeed, Laudan and Leplin's example, namely "previous sightings of black crows support the hypothesis that the next crow to be sighted will be black", is a case where an observation statement not entailed by a given observational hypothesis supports the hypothesis (1991, p. 461). Note, however, that this is a purely observational hypothesis.

[26] A similar retreat from consequence-empiricism is required by my rejoinder to Laudan and Leplin's point at the beginning of this section. They argue that not all observational consequences of a theory need provide support for it. I reply that one can always restrict their belief from the set of observational consequences to those and only those observational statements that have the power to confirm.

[27] Some take these virtues to denote one and the same thing that is merely expressed in different ways. Kukla has called this the 'theoriticity' or 'charm' of a theory. It is worth noting that Laudan and Leplin (1993) appeal to this type of extra-observational consideration.

[28] Of course, being a Bayesian does not necessarily mean that one takes theoretical virtues as formal criteria for restricting one's priors. Thus, in considering the role of the notion of simplicity in Bayesianism, Howson and Urbach (1996) state that "the addition of *any* criterion [including simplicity] for determining prior distributions is unwarranted in a theory which purports to be a theory of consistent degrees of belief, and nothing more" (p. 418) [original emphasis].

[29] That does not mean that she rejects any kind of appeals to unification, let's say within specific domains, but rather that she rejects overall claims of the unification of the science.

[30] Kukla challenges Laudan and Leplin to explain wherein theoriticity exceeds 'logico-semantic trickery'. Their reply is unsatisfactory, deferring these issues to scientific judgment (Laudan and Leplin 1993, p. 13). The only other thing they add to their defence is that a theory needs to be confirmable at least in principle. Crucially, their reply does not address the empiricist complaint that only empirical statements are confirmable.

[31] The first two conjuncts of EVE (and of the amended versions of EVE below) specify that for a pair of theories to qualify they must be 'rivals' in the minimal sense elucidated at the end of section two.

[32] Another, even weaker, suggestion is to require that all *true* theories lack evidentially equivalent rivals. The problem with this suggestion is that it says something trivial. If *T* is a true theory, then surely *T′* must be false since it contradicts *T*. There is no point then to require that all true theories lack evidentially equivalent rivals. The notion of empirical adequacy adopted here is the one advocated by van Fraassen, i.e. a theory is empirically adequate if it saves all (and only) the observable phenomena.

[33] Note that EVE5 without the last conjunct is logically equivalent to the negation of EVE.

References

Boyd, R. (2002) 'Scientific Realism', *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/ archives/sum2002/entries/scientific-realism/

Brush, S. and Holton, G. ([1952]2001) *Physics*, *The Human Adventure*: *From Copernicus to Einstein and Beyond*, 3rd edition, New Brunswick, NJ: Rutgers University Press.

Devitt, M. (2002) 'Underdetermination and Realism', *Philosophical Issues*, vol. 12: 26-50.

Hempel, C.G. (1945a) 'Studies in the Logic of Confirmation: Part I', *Mind*, vol.54: 1-26.

Hempel, C.G. (1945b) ‘Studies in the Logic of Confirmation: Part II’, *Mind*, vol.54: 97-121.

Howson, C. and Urbach, P. (1996) *Scientific Reasoning*: *The Bayesian Approach*, 2nd edition, Chicago and La Salle (IL): Open Court.

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

Laudan, L. (1996) *Beyond Positivism and Relativism*, Boulder, CO: Westview Press.

Laudan, L. and Leplin, J. (1991) ‘Empirical Equivalence and Underdetermination’, *Journal of Philosophy*, vol. 88: 449–72.

Laudan, L. and Leplin, J. (1993) ‘Determination Underdeterred: Reply to Kukla’, *Analysis* 53: 8-16.

Nelson, D. (1996) ‘Confirmation, Explanation, and Logical Strength’, *British Journal for the Philosophy of Science*, vol. 47(3): 399-413.

Newton, I. ([1726]1999) *Mathematical Principles of Natural Philosophy*, trans. by I. B. Cohen and A. Whitman, Berkeley: University of California Press.

Okasha, S. (1997) 'Laudan and Leplin on Empirical Equivalence', *British Journal for the Philosophy of Science*, vol. 48(2): 251-256.

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

Quine, W.V.O. (1975) 'On Empirically Equivalent Systems of the World', *Erkenntnis*, vol. 9: 313-328.

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

Williamson, T. (2000) *Knowledge and its Limits*, Oxford: Oxford University Press.