

Another Look at Empirical Equivalence and Underdetermination of Theory Choice

Pablo Acuña and Dennis Dieks

Institute for History and Foundations of Science, Utrecht University

Budapestlaan 6, 3584 CD, Utrecht

Abstract. In 1991 Larry Laudan and Jarret Leplin proposed a solution for the problem of predictive equivalence and empirical underdetermination. In this paper we claim that, even though Laudan and Leplin's reasoning is essentially correct, their solution of the underdetermination problem must be accurately assessed in order to appreciate its nature and scope. We argue that Laudan and Leplin's analysis does not *refute* the problem (as they claim). Instead, what they show is merely that science possesses tools that *may eventually* lead out of an underdetermination impasse. We apply their argument to a real case of two empirically equivalent theories: Lorentz's ether theory and Einstein's special relativity. We argue that this example shows that the core of Laudan and Leplin's proposal works, but also that the reassessment we argue for is correct and necessary. We conclude that empirical equivalence and underdetermination are ordinary scientific problems rather than problems that should be solved by epistemology.

Keywords: Empirical equivalence, underdetermination, theory choice, non-empirical virtues, empirical evidence, confirmation, special relativity, Hendrik Lorentz.

1 Introduction

Empirical equivalence of theories is often regarded as a philosophically deeply problematic situation for science, because it would appear to make an evidence-based choice between theories impossible. The problem has received lots of attention in the philosophy of science literature. This is hardly surprising: if the problem were to prove intractable, the objectivity of theory choice and even its very rationality might be threatened.

Larry Laudan and Jarret Leplin (1991) have offered a very influential argument designed to show that the problem is a philosophers' fabrication and does not really exist. In this paper we argue that, even though Laudan and Leplin's analysis of the situation is essentially correct, the question of what conclusion can be drawn from it requires a reassessment. In particular, their reasoning is not enough to *refute* the problem, in spite of what they claim. It does show that the regular practice of science may eventually be able to break the deadlock of underdetermination. In other words, Laudan and Leplin's argument, in its reassessed form, indicates that empirical equivalence and underdetermination are problems for which science can provide a way out. That is, the regular practice of science can lead to the breakdown of the empirical equivalence between theories and the termination of the underdetermination of the choice to be made. However—just as

in any other problem of science—that the problem will be dissolved by scientific research is not guaranteed from the outset.

This paper is divided in five sections. In the first one we provide a simple and schematic presentation of the problem and its logical structure. In the second section we briefly consider the type of solutions that have been typically proposed, and we show their limitations or failure. In the third one we present and analyze Laudan and Leplin’s proposal along with the most relevant objections that have been leveled against it. A close consideration of such objections is helpful in order to understand what really has been achieved by Laudan and Leplin. The resulting reassessment, which constitutes the main subject of this paper, is presented in the fourth section. In the fifth and final part we use an actual example of empirical equivalence between two theories as a test for the reassessed solution: the case of Lorentz’s ether theory vs. Einstein’s special relativity.

2 The problem

The problem of empirical equivalence (EE) as leading to underdetermination (UD) of theory choice comes from an argument with two premises. The first premise is that *for any theory T and any body of observational evidence E , there is another theory T' such that T and T' are empirically equivalent with respect to E* . The support for this statement is twofold. First, some authors claim that there exist algorithms that are able to generate an empirically equivalent theory T' given any theory T . Second, there is the Duhem-Quine thesis, which states that a hypothesis can entail observable consequences only with the help of auxiliary assumptions. Thus, one can expect that conflicting evidence can be accommodated by any hypothesis through the introduction of suitable arrangements in the auxiliary assumptions. Therefore, one should generally expect that if a hypothesis H —along with the class of auxiliary assumptions A —entails the observational consequence e , there exists another hypothesis H' that can also entail e by introducing a suitable class of assumptions A' ¹.

The second premise of the argument says that *only entailment of the empirical evidence counts as epistemically justified confirmation of a theory*. This seems a standard principle of confirmation, according to which an observational report counts as evidence for a certain (deterministic) hypothesis only if it can be derived, explained or predicted from that hypothesis—with the help of auxiliary assumptions, of course².

These two premises entail the problematic conclusion. If there is an EE rival to any theory, and if a theory gets confirmed only by means of the observational consequences it entails, then *the choice to be made between two EE theories is empirically underdetermined*—and the universal scope of the first premise entails

¹ Suppose that the hypotheses H and H' are rivals, that $(H \wedge A) \rightarrow e$, that $(H' \wedge A) \rightarrow \neg e$, and that e is observed—so that H is confirmed and H' disconfirmed. The Duhem-Quine thesis implies that it is always logically possible to change A in a way such that $(H' \wedge A') \rightarrow e$. Therefore, it is always logically possible to create EE between H and H' .

² In *probabilistic* theories we have to refine the criteria: we should require that the *probability* of the evidence is the same according to the theories in question.

that theory choice is underdetermined for *all* theories. If this problem were intractable, then the objectivity and even rationality of theory choice would come under threat³.

3 A partial solution and some false exits

A straightforward way out of the problem would be to weaken premise 2 by having recourse to *non-empirical* virtues of the theories involved in order to make a choice. If one of the theories is simpler or proves to have more explanatory power than its rival, for example, one might have a reason to prefer this theory after all. However, both simplicity and explanatory power are features that cannot be assessed unambiguously. In the case of simplicity, the very definition of the concept is far from clear—it looks like a feature that depends on subjective considerations: *one person's simplicity is another person's complexity*. Moreover, as Mario Bunge states (1961), there are multiple senses in which a theory can be regarded as simple—syntactical, semantic, epistemological and pragmatic—and these different forms of simplicity are not necessarily compatible with each other. Therefore, it is very problematic, at best, whether a theory can be simpler than another in unambiguous terms.

Something similar holds in the case of explanatory power. What a scientific explanation is, or must be, is an open philosophical question. Moreover, there are compelling arguments that explanation is a context-dependent concept. Bas van Fraassen (1980, 134-157), for example, argues that an explanation is an answer to a why-question, so that the degree of explanatory power of a theory depends on the specific why-question that is being asked, plus the context of that question. Consequently, different why-questions and different contexts can yield different degrees of explanatory power for the same scientific hypothesis or theory⁴.

These remarks indicate that non-empirical features will often be pragmatic, so that they cannot be invoked in order to make an entirely objective, epistemically compelling choice between EE theories. It is true that pragmatic considerations regarding simplicity and/or explanatory power can provide plausible reasons to prefer one of the theories, and in this sense they may lead to dissolution of the problem with respect to the *rationality* of theory choice. Even if empirical evidence cannot be invoked to determine a decision, pragmatic aspects could be used to make a rationally grounded choice after all.

³ We will tackle the problem of EE and UD in terms of the challenge that it poses with respect to the rationality and objectivity of theory choice, not from the standpoint of the challenge that it poses with respect to the *realist* conception of scientific theories. For an overview of the problem as a problem for the realist, see (Psillos 1999, chapter 8).

⁴ See also (De Regt and Dieks 2005). There it is argued that ‘scientific understanding’, and *a fortiori* ‘explanation’, are pragmatic, context-dependent features. A phenomenon *P* is *understood* if there is an *intelligible* theory about *P*; and a theory *T* is intelligible if scientists are able to recognize qualitatively characteristic consequences of *T* without performing exact calculations. Different ‘conceptual toolkits’ can work as sources of intelligibility for a theory—visualization, causal explanations and unifications. The crucial point is that none of these explanatory virtues can be asserted as necessary or sufficient in order to obtain intelligibility for a theory; rather, which tools can provide intelligibility depends on contextual features: ‘there is no universal tool for understanding, but a variety of ‘toolkits’, containing particular tools for particular situations. Which tools scientists have at their disposal, depends on the (historical, social, and/or disciplinary) context in which they find themselves. This context-dependence is typical of a meso-level nature, i.e., it is the scientific community that determines what tools are available and which skills are required to achieve understanding’ (De Regt and Dieks 2005, 158).

However, the *objectivity* of theory choice cannot be rescued in this way. Because of their pragmatic nature, non-empirical virtues like simplicity and explanatory power are context dependent. Even if it is possible to state unambiguously that a theory possesses more explanatory power than its rival in a certain context, there may well be other contexts in which the rival is simpler or explains better. In this case we would have a situation in which there are good pragmatic reasons supporting *both* theories, but since they are rivals we cannot accept them at the same time⁵. The limitation of this kind of pragmatic solution is therefore that, even though pragmatic non-empirical features provide us with plausible reasons to favor one of the theories, they are not enough to provide a fully objective and uniquely determined choice—the opposite choice could be rational as well. But we can do better than this. As we will see below, there are arguments showing that, in spite of EE, *empirical evidence* can be invoked in order to find a way out of the problem—so that a fully objective and uniquely determined choice can be made.

Another possible solution of the problem relies on the view that EE simply means that we are presented with two different formulations of the *same* theory. This was the position that many logical positivists held. Their original stance depended on the verificationist criterion of meaning—if the meaning of all scientific terms reduces to the method to verify them (their empirical consequences), then the meaning of two EE theories must be the same. But there is no need to rehearse here the arguments for the untenability of the verificationist criterion of meaning: the meaning of scientific terms simply cannot be reduced to their empirical means of verification. Therefore, the meanings of two theories that predict exactly the same observational consequences can be different and even incompatible.

John Norton (2008) has recently offered a sort of revival of the theory identity view, but now based upon the structural conception of scientific theories. He claims that if two theories are EE, their theoretical structures are most likely very similar—otherwise we could not demonstrate that they are EE to begin with. The two sets of theoretical structure may be inter-convertible with or without loss of structure. In the latter case it is quite likely that the structures are two formulations of the same theory, as in the case of matrix and wave mechanics. In the other case—if in the inter-conversion some structure is lost—Norton argues that since one of the theories contains extra-structure that is not really necessary for the derivation of the observational consequences—as the most economic theory shows—it is very likely that the extra-structure is superfluous and represents nothing physical. Consequently, Norton claims, this is a strong indication that we are dealing with two formulations of the same theory—the only difference being that in one of them there is superfluous, non-physical, extra-structure.

There is a significant difference between Norton's and the logical positivistic approach. For the latter, the verificationist criterion of meaning *a priori determines* that EE simply means different formulations of the

⁵ For constructive empiricists it is possible to accept both theories at the same time. Since they are not committed to the non-empirical content of the theories, they can accept both as empirically adequate and make a pragmatic preference if the context so requires. This stance only works if we are willing to accept that empirical adequacy is enough, that is, if empirical adequacy is the basic and sufficient feature that we should expect from a theory in order to accept it. The cost would be to quit to demands for *understanding* from scientific theories, for example. We think that a more general solution is available. There are arguments that show that a way out is possible regardless of whether one is a constructive empiricist, a realist, or what have you.

same theory. For Norton, on the other hand, EE is a strong *indication* that we might be dealing with two different presentations of the same theory, but it is still possible that the theories involved may be genuine rivals. Therefore, Norton's argument—even though it does point out that theory-identity is a possibility in cases of apparent EE—cannot count as a solution of the problem. EE as leading to UD of theory choice remains a possible scenario, and if this scenario is actually realized, there is nothing in Norton's view concerning theory identity to provide a criterion for choosing one of the theories. Furthermore, there is no general criterion of theory-identity available, and thus theory-identity cannot provide us with a general solution of the problem at issue. As P. D. Magnus formulates it,

I do not deny that a criterion of theory identity would be a nice thing to have. Problems of theory individuation, of which the problem of identical rivals is a special case, are interesting in their own right. Resolving them, however, can only come as the result of a careful examination of the history of science—an examination which must be left for some other time. I draw the modest conclusion that this open question need not turn us back from considering underdetermination. (Magnus 2003, 1263).

Jarret Leplin (1997a, 154-155; 1997b) has offered yet another attempt to block the problem. He claims that the EE thesis is inconsistent with the UD thesis. If two theories are shown to be EE, it is needed that the auxiliary hypotheses which permit the entailment of their empirical consequences are well determined and established. However, if the UD thesis is true, those auxiliary assumptions are underdetermined. Therefore, they cannot be well-established and cannot be used to derive empirical consequences from a given hypothesis:

Admissible auxiliaries are those independently warranted by empirical evidence. Unless auxiliaries are better supported than the theory they are used to obtain predictions from, those predictions cannot be used to test the theory. The significance of their success or failure would be indeterminate as between the theory and the auxiliaries. The result would be a holism that enlarges the possible units of empirical evaluation, and prevents epistemic support from accruing to theories directly. Such is the upshot of the classic theses of Duhem, who stressed the ineliminability of auxiliaries from prediction. (Leplin 1997a, 155)

Leplin claims that the empirical consequences of a theory cannot be determined if the required auxiliary hypotheses are underdetermined. If UD is the case, then EE cannot be decided; and if EE is true, then UD cannot be true. Leplin is thus arguing that the problematic argument is logically flawed:

The truth of underdetermination would prevent the determination that theories are empirically equivalent in the first place. Because theories characteristically issue in observationally attestable predictions only in conjunction with further, presupposed background theory, what observational consequences a theory has is relative to what other theories we are willing to presuppose. As different presuppositions may yield different consequences, the judgment that they have the same observational consequences—that they are empirically equivalent—depends on somehow fixing the range of further theory available for presupposition. And this underdetermination ultimately disallows (*ibid.*, 154-155).

But Leplin's argument does not work. His position presupposes a *radical* understanding of the Duhem-Quine thesis, but under a *moderate* view of it—a view which is much more compelling than a radical one

anyway—UD is consistent with EE⁶. Suppose that background scientific knowledge *B* is well-established and evidentially confirmed, so that it can be certainly used to provide auxiliary assumptions. Consider a new theory *T* given by hypothesis *H*, which along with the class of auxiliary statements (*A*₁, *A*₂, *A*₃) in *B* entails the class of observational consequences *E*. Consider yet another new theory *T'*, in which the hypothesis *H'* along with the class of auxiliary assumptions (*A*₄, *A*₅, *A*₆) in *B* also entails *E*. It is clear that *T* and *T'* are EE, and that the choice to be made between them is empirically underdetermined—EE and UD are thus consistent theses. What Leplin's argument shows is only that any further theory that in order to entail its observational consequences requires either *H* or *H'* will face the problem of justifiably determining such consequences, for *H* and *H'* are empirically underdetermined. That is, if we consider a moderate holist thesis of UD then Leplin's inconsistency problem does not arise, but the EE and UD problem does. The problematic argument does not require Leplin's radical version of the Duhem-Quine thesis⁷.

So the premises of the EE-UD problem are consistent and our problem still stands. As we have seen, Norton's strategy only *suggests* that in a case of EE we might be facing a case of theory-identity, and this is not enough to provide a general solution. Further, recourse to non-empirical features as a ground for making a choice does not work as a complete solution either, for such features will not be objective and will not determine unique choices—although they may be used to block the threat of *irrationality* of theory choice. Apparently, a (dis)solution of our problem must involve denying, or at least weakening, the premises of the problematic argument in a way different from what we have seen until now. This is exactly what Laudan and Leplin have done.

4 Laudan and Leplin's solution

Larry Laudan and Jarret Leplin's influential paper (1991) consists in a careful analysis of the validity of the two premises explained in section 2, with the conclusion that the problem evaporates. We will present their criticisms of the premises in turn.

4.1 The first premise

4.1.1 EE, observability and auxiliary assumptions

Laudan and Leplin affirm that three non-controversial theses regarding the nature of evidential confirmation imply that EE is not a universal feature of theories in the sense of the first premise. The first

⁶ See (Laudan 1990). Laudan compellingly argues that under a moderate conception of the Duhem-Quine thesis firm confirmation of hypotheses by empirical data is indeed possible. 'Firm confirmation' does not mean 'absolute' or 'infallible confirmation', though. A moderate version of the D-Q thesis states that the confirmation of hypotheses by means of empirical evidence faces inevitable holistic aspects, and is therefore fallible. According to this moderate viewpoint it is a mistake to think that all theories are equally confirmed, or that firm confirmation is not possible—this would be part of a 'radical' interpretation of the thesis.

⁷ If we pay attention to Duhem's concept of *good sense*—which indicates the good scientist when a falsification or confirmation of a theory has been obtained in spite of the *logical possibility* to introduce ad-hoc hypotheses or suitable logical tricks—it is quite clear that this *moderate* holism is Duhem's UD thesis, not the radical one that Leplin refers to. See (Duhem 1954, 37-39).

thesis focuses on *the variability of the range of the observable*: ‘any circumscription of the range of the observable phenomena is relative to the state of scientific knowledge and the technological resources available for observation and detection’ (Laudan and Leplin 1991, 451). Whether an entity or process described by a theory qualifies as observable or not depends not only on the meaning of the corresponding term. Observability also crucially depends on the available experimental methods and instruments at a certain stage of scientific development⁸.

The second thesis is *the need for auxiliaries in prediction*: ‘theoretical hypotheses typically require supplementation by auxiliary or collateral information for the derivation of observable consequences’ (ibid., 452). This *Duhemian* statement is so widely known and accepted that it does not require further comments.

The third thesis consists in *the instability of auxiliary assumptions*: ‘auxiliary information providing premises for the derivation of observational consequences from theory is unstable in two respects: it is defeasible and it is augmentable’ (ibid.). As a consequence of scientific progress the class of auxiliary assumptions which are suitable for the derivation of observational consequences from theoretical hypotheses may get enlarged by the introduction of new well-confirmed theoretical hypotheses or newly discovered facts, or it may get reduced through the rejection of theoretical hypotheses which were previously accepted.

The effect of these three non-controversial theses on our problem is clear. If what is observable is variable and depends on current background knowledge, and if the class of auxiliary assumptions that are available for the derivation of observational consequences is also variable and background knowledge-dependent, then the class of observable consequences of any theory is relative to a particular state of scientific knowledge. Therefore, EE between two theories is a feature that is relative to a certain state of scientific knowledge as well:

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. This contextuality shows that determinations of empirical equivalence are not a purely formal, a priori matter, but must defer, in part, to scientific practice. It undercuts any formalistic program to delimit the scope of scientific knowledge by reason of empirical equivalence, thereby defeating the epistemically otiose morals that empirical equivalence has been made to serve (ibid., 454).

The upshot is that if two theories make the same predictions now, it does not follow that they are EE, for further development of science could break the equivalence and, *a fortiori*, the empirical UD of the choice to be made. However, Andre Kukla (1993, 1996) has offered the following natural criticism. We can accept that two theories (T_1, A_1) and (T_2, A_2) —where A stands for the auxiliary assumptions—can be considered as

⁸ Laudan and Leplin acknowledge that van Fraassen would not accept this thesis. However, they claim that ‘we reject [van Fraassen’s] implicit assumption that conditions of observability are fixed by physiology. Once it is decided what is to count as observing, physiology may determine what is observable. But physiology does not impose or delimit our concept of observation. We could possess the relevant physiological apparatus without possessing a concept of observation at all. The concept we do possess could perfectly well incorporate technological means of detection. In fact, the concept of observation has changed with science, and even to state that the (theory-independent) facts determine what is observable, van Fraassen must use a concept of observation that implicitly appeals to a state of science and technology’ (Laudan and Leplin 1991, 452).

EE only relative to a time t . However, ‘there is nothing in the argument that would force me to give up the view that every indexed theory has empirically equivalent rivals with the same index’ (Kukla 1996, 142). Although the first premise of the problematic argument has been relativized with respect to time, it remains universal in scope, for even if the EE between T_1 and T_2 were broken in T_2 ’s favor at time t' —by means of the new set of auxiliary assumptions $A_{t'}$ —, at t' there will be a theory $(T_3, A_{t'})$ which is EE with $(T_2, A_{t'})$ —and so on for any future t . As Kukla puts it, ‘the point is that we know that, whatever our future opinion about auxiliaries will be, there will be timeless rivals to any theory under those auxiliaries’ (ibid.).

This implies that even though EE is a time-indexed relation between two given theories, theory choice will be empirically underdetermined for any value of t . The crucial point in Kukla’s revival of UD is the universal scope of the EE premise—it is supposed to hold *for any* theory. Kukla argues for this universal scope on the basis of the existence of algorithms that provide an alternative EE theory for any input theory. If algorithms like this indeed exist and are effective, it certainly follows that any theory has a time-indexed EE rival. Therefore, Laudan and Leplin have to show that such algorithms are ineffective.

4.1.2 Algorithms

In their original 1991 paper Laudan and Leplin only offer a brief and general disposal of one kind of algorithm, namely, algorithms that excise the theoretical vocabulary from a theory without empirical loss. The *Ramsey sentence* and *Craig’s theorem*—logical tools which were originally considered by the logical positivists to show that theoretical terms are unnecessary in scientific theories⁹—are well-known examples. Laudan and Leplin argue that algorithms of this kind do not work. The reason is that the output theory T' they produce cannot be considered as a genuine *rival* to the input theory T . T' —the result of the application of Ramsey’s sentence or Craig’s theorem—is just T without its theoretical vocabulary. Thus, Laudan and Leplin maintain, T' is nothing but the *instrumentalized* version of T . The algorithmic theory T' does not include the theoretical terms of T , but since it is a logical consequence of T , and since the theoretical terms are crucial for the derivation of the observational consequences of T , T' is a *parasite* on T :

The algorithm does not produce a rival representation of the world from which the same empirical phenomena may be explained and predicted. On the contrary, a theory’s instrumentalized version posits nothing not posited by the theory, and its explanations, if any, of empirical phenomena deducible from it are wholly parasitic on the theory’s own explanations. A theory’s instrumentalized version cannot be a rival to it, because it is a logical consequence of the theory and it is bound to be endorsed by anyone endorsing the theory (Laudan and Leplin 1991, 456-457).

Kukla replies that these remarks are not enough to dismiss algorithms. In his (1993) he offers two other candidates which, according to him, are able to create an EE rival. First, he considers the following: ‘for any theory T , construct the rival T^* that asserts the world to be observationally exactly as if T were true, but denies the existence of the theoretical entities posed by T ’ (Kukla 1993, 4). If instead of calling it the

⁹ For a detailed explanation of the Ramsey sentence and Craig’s theorem, and of why both failed to accomplish the logical positivist goal, see (Suppe 1974, 27-35).

instrumentalized version of T , we call T^* the *antirealist* version of T , Kukla's candidate falls prey to the same criticisms that Laudan and Leplin level against T' —the output of Ramsey's sentence and Craig's theorem—namely, that T^* is parasitic with respect to T .

John Norton provides a similar reason to dismiss Kukla's algorithm. Even if we accept that T and T^* have the same empirical consequences, it is clear that the theoretical terms and entities in T are necessary for the derivation of such consequences for *both* theories¹⁰. Therefore, by negating those terms and entities T^* gets gratuitously impoverished:

If we assume that the algorithm is applied to a well-formulated theory T whose theoretical structure is essential to T 's generation of observational consequences, then the construction of T' [Kukla's T^*] amounts to a gratuitous impoverishment of theory T , the denial of structures that are essential to the derivation of observational consequences that are well confirmed by them (Norton 2008, 39-40).

Kukla's second candidate for an algorithm is the following: take a theory T with class O of observational consequences, and construct from it the theory T' , which states that T is true for the world under initial conditions in which it is being observed, but that also says that when nobody is observing the universe behaves according to the laws of T^* —where T^* is any theory which is incompatible with T . It is clear, Kukla asserts, that T and T' are EE rivals (1993, 4-5).

Laudan and Leplin's response to this algorithm is twofold (1993, 10-14). First, incompatibility with T cannot be the only constraint on T^* , if T' is to be EE with T , the laws in T^* must be such that T' mirrors the observational predictions of T — T^* cannot be *any* incompatible theory with T —, but the algorithm does not include any procedure to achieve that. Second, T^* must provide a plausible explanation of why the world changes its behavior when it is not observed—otherwise, T' will be quite a bizarre theory—, but, again, the algorithm does not provide any mechanism to obtain such an explanation. Therefore, Kukla's example is nothing but a *promissory note* for an algorithm¹¹.

Laudan and Leplin's and Norton's criticisms of the effectiveness of algorithms are quite natural and appealing. However, there is a deeper challenge to be met. Laudan and Leplin base their criticisms on certain features of the output of algorithms that purportedly disqualify them as *pseudo*-theories. Kukla therefore demanded a general criterion that grounds this disqualification, he asks for a criterion that determines which hypotheses count as genuinely scientific:

It seems to me that the whole philosophical dispute between the received-viewers and Laudan and Leplin comes down to the issue of distinguishing genuine theoretical competitors from logico-semantic tricks. Laudan and Leplin represent the issue as being concerned with the existence or nonexistence of empirical equivalents.

¹⁰ The theoretical terms are required to derive the empirical consequences of T^* , but they are denied in this theory. See the example of intentional psychology below.

¹¹ It is still possible to weaken the algorithm and take it just as stating that T' asserts that T holds when we are observing, but it does not hold when nobody is looking. As a *theory*, this would be way too bizarre to be considered as genuinely scientific. However, the weakened algorithm can still be taken as an instance of the evil-genius argument—as an instance of the fact that, from a logical point of view, there are many hypotheses consistent with the information of our senses but that deny them as providing reliable information about reality. But in this case the algorithm is no longer a problem of the philosophy of science, but of metaphysics.

But it is evident, both from my example as well from the example they reject in a footnote, that there *do* exist empirically equivalent propositions to any theory. The only question is whether these structures fail to satisfy some *additional* criteria for genuine theoreticity. The received-viewers are satisfied with their examples of empirical equivalence. The burden is on Laudan and Leplin to explain why empirical equivalence isn't enough (Kukla 1993, 5)

4.1.3 Theoreticity

Kukla (1996, 2001) continues that attacks on algorithms are not justified because philosophers of science have not been able to provide a satisfactory account of the constraints for theoreticity. He first considers 'parasitism', and claims that Laudan and Leplin have not shown that a theory T' —the *instrumentalized* or *antirealist* version of theory T —cannot be formulated as to circumvent any reference to T (1996, 148-149). This is a rather strange argument, for the burden of proof is clearly to provide an algorithm capable of producing a theory T' which does not make essential reference to T —or at least to offer an alternative presentation of T' which does not make reference to T . The only two serious attempts at such an algorithm, Craig's theorem and the Ramsey sentence, proved ineffective.

Kukla offers a second argument to show that 'parasitism' cannot be a theoreticity requirement (ibid., 149-150). He claims that there exists an example of a structure such as T' —which affirms that the observable consequences of T are true, but denies T —that actually forms a part of scientific practice. Following Daniel Dennett, he states that the instrumentalist view of intentional psychology is accepted by the relevant community because of its predictive power, but the ontology of the theory is not believed to be true because it is incompatible with physicalism. Therefore, if Laudan and Leplin reject the use of such a structure in terms of the parasitic nature of T' , they would be denying an accepted practice in real science.

However, Kukla here seems to misunderstand the parasitism attack on algorithms. Laudan and Leplin are not saying that the outcome T' of the algorithm must be dismissed from the outset—as a pseudo-theory—because it is a parasitic theory with respect to T . Their point is that the parasitic reference of T' to T means that T' is not a genuine *rival* to T , T' is simply the *instrumentalized* or *antirealist* version of T . The difference between T and T' boils down only to the epistemic stance one takes towards the very same theory. What Kukla shows is only that, according to Dennett, in the case of psychology the *instrumentalist* attitude with respect to intentional psychology is more appropriate than the realist one.

Yet another argument adduced by Kukla is that there exists a specific algorithm that produces a non-parasitic output-theory: given theory T , there is a T' that says that the world behaves exactly as T states when we are observing, but when we are not it predicts behavior according to T^* —and it is clear that T' does not necessarily make reference to T . This algorithm could be attacked, Kukla notes, on the basis of the *superfluity* of the hypothesis of 'the intermittency of natural law'. A hypothesis is superfluous if it could be dispensed with in the theory it belongs to without any loss of empirical content, i.e., if it does not contribute to deriving any observational consequences. However, Kukla argues that if this concept of 'superfluity' is to be invoked, then the theory T should also be dismissed because it contains the superfluous hypothesis that 'the laws of nature continue to hold when nobody is looking' (see Kukla 1996, 151-153).

This view also seems to rest on a misunderstanding. If T is a theory in which the laws of nature are the same regardless of whether anyone is looking, then the ‘continuity’ of the laws of nature is not an *extra* hypothesis in T , it is just a feature of its laws. In the case of Kukla’s T' two situations are possible: its laws establish a connection between observation and the behavior of the world, or the ‘intermittency’ of the laws of nature is an extra, unexplained hypothesis. In the second case, it is indeed the case that the ‘intermittency hypothesis’ is superfluous—also untestable—and therefore the theory has problems of theoreticity.

The other possibility for Kukla’s T' —that it includes laws which connect observation and the behavior of the world—is interesting because it allows us to clarify another criterion of theoreticity. If T^* in T' explains ‘the inconsistency of the behavior of the world’, then whether or not T' can be accepted as genuinely scientific will depend on the *kind* of explanation T^* provides:

Provisions that fly in the face of what we have good empirical reason to assume must claim some offsetting rationale if they are to be admitted as part of a theory. It would be different if the course of nature were known to exhibit such vast and mysterious ruptures or bifurcations as T' envisions, if natural law did not exhibit isometry, at least. One might then be willing to entertain wild, unexplained and unconfirmable scenarios as genuine possibilities. But the world is not known to be like that (Laudan and Leplin 1993, 14).

Though Laudan and Leplin do not dub this feature, it seems that ‘plausibility’ fits. In order to be considered as genuinely scientific a hypothesis must possess a minimum degree of plausibility, which is typically conferred by a background of empirically well-confirmed knowledge. This requirement must, of course, not be made so strict as to demand complete consistency between new hypotheses and background knowledge. Hypotheses that ‘fly in the face of what we have good empirical reason to believe’ have formed a part of successful science. But even those revolutionary hypotheses must be given a minimum of plausibility. In our context this means that Kukla’s T' will be genuinely scientific if the hypotheses in T^* that explain the ‘inconsistent behavior of the world’ by connecting observation with the course of nature possess a measure of plausibility, in the sense of some (perhaps indirect) empirical support. Since it is clear that the algorithm to produce T' does not contain any indication of how to obtain that minimum degree of plausibility, it follows that it is nothing but a promissory note for an algorithm.

One final requirement of theoreticity we would like to address is given by the ‘testability’ of hypotheses. This feature can also be used to disregard possible algorithms. If algorithms produce theories that contain superfluous additional hypotheses, in the sense that they do not participate in the entailment of observational consequences, these hypotheses will be untestable:

Because the purpose of theorizing is, at least in part, to gain predictive control over the subject matter under investigation, a theory must, at least in principle, be open to test. A ‘propositional structure’ that is not even in principle confirmable, that could not logically be an object of epistemic evaluation, is not a theory; for it could not in principle impart understanding nor advance practical interests (Laudan and Leplin 1993, 13).

In spite of Kukla, superfluity, implausibility and untestability are thus features that can be coherently defined and justifiably invoked in order to dismiss hypotheses as unscientific. The demand for testable, non-

superfluous and plausible hypotheses and/or theories is justified by basic goals of science. We will not deal with a detailed consideration of what the goals of science are, but both testability and non-superfluity are requirements which are grounded in the aim of achieving *empirical* knowledge and of excluding metaphysical-unfalsifiable elements from scientific theorizing. On the other hand, the demand for plausibility relies on the aim of achieving explanations of natural phenomena that make them intelligible¹² in the sense of making them fit in with general empirically based background knowledge.

Kukla's persistence that no theoreticity constraints can be coherently put forward seems rooted in a general misconception of the term:

Even though a hypothesis may possess the traditional empirical virtues of having a truth-value, being confirmable and disconfirmable, and generating indefinitely many testable predictions¹³, it might nevertheless be excluded from serious scientific discourse for failing to satisfy an *a priori* constraint on the proper form of a scientific theory. Let us call this property by the name of *theoreticity* (Kukla 1996, 146).

But to assume that theoreticity requirements are *a priori* is not needed at all. Whether or not a given hypothesis or theory is testable, superfluous or plausible, is not something to be determined *a priori*. A hypothesis is testable if, along with other assumptions, observable consequences can be derived from it, and a hypothesis is non-superfluous if it is required for the derivation of observational consequences of a theory. But the class of auxiliary assumptions available for the derivation of observational consequences changes with time. Therefore, it is possible that a hypothesis which is non-testable and superfluous in a given state of science may become testable and non-superfluous with the introduction of new auxiliary assumptions¹⁴.

In the case of 'plausibility', this property is typically grounded in background scientific knowledge. Consequently, a hypothesis that is completely implausible with respect to a certain stage of the development of scientific knowledge might become plausible enough with the acceptance of new theories¹⁵. Hypotheses are not superfluous, untestable and/or implausible in themselves, but with respect to a concrete state of scientific knowledge. Theoreticity is not an *a priori* matter.

It is important here to emphasize that theoreticity constraints serve as a tool for blocking algorithms that automatically yield EE theories; the main point in this subsection is to discuss the first premise of our problem, that given any theory *T* there is an EE rival *T'*. The universal scope of this premise crucially depended on the effectiveness of algorithms. But theoreticity requirements—which, in spite of Kukla, we

¹² We consider, unlike constructive empiricists, that explanation and understanding are essential aspects of science—see (De Regt and Dieks 2005).

¹³ This passage indicates that Kukla accepts testability as a requirement for a theory or hypothesis. But then his rejection of superfluity as a theoreticity constraint becomes problematic. If superfluous hypotheses can be admitted then untestable ones also can.

¹⁴ A good example is Hendrik Lorentz's *length contraction hypothesis*. In Lorentz's first version of his ether theory, the only 'testable' consequence of the hypothesis was given by the (null) result of the Michelson-Morley experiment. However, when the hypothesis was merged with the coordinate transformations in the final version of the theory, the hypothesis was logically relevant for the spectacular prediction of the velocity dependence of the electron's mass. See (Janssen 1995, chapter 3).

¹⁵ Einstein's *light quantum hypothesis* of 1905 was resisted and considered implausible by most of the scientific community of the time, including Bohr. Even though it explained puzzles like the photoelectric effect, it was accepted just as a *heuristic principle* (this is how Einstein shyly introduced it). However, in 1922 the explanation of the *Compton effect* it provided made it plausible and it was then more generally accepted in its full meaning.

have established to be well-defined and justified¹⁶—preclude that all outputs may be considered as genuinely scientific hypotheses or theories. When it comes to EE between *genuine* scientific theories these basic theoreticity requirements are fulfilled by the theories involved, by definition—otherwise the theories would not be genuinely scientific—, so they cannot function as criteria that provide a way out of the choice problem¹⁷.

Summarizing, Laudan and Leplin’s treatment of the first premise of the problematic argument shows that *i*) EE is an intrinsically time-indexed feature; and that *ii*) theoreticity constraints show that there are no automatic algorithms capable of producing an EE rival given any theory T ¹⁸. Therefore, EE and UD are not a guaranteed *universal* problem. It is not true that *for any* theory T there is *eo ipso* an EE rival T' , for the algorithms that were proposed to support this view are not effective¹⁹. On the other hand, EE and UD, if present, are not necessarily *everlasting* problems, for the development of science might be such that the EE between theories gets broken.

However, Laudan and Leplin have not disproved the possibility of time-indexed EE, so that a time-indexed UD of the choice between them is still possible. Although algorithms may not work, it is still possible that a genuinely scientific EE rival might be formulated after all. Moreover, that EE is essentially time-indexed does not logically *imply* that further development of science will *surely* break the equivalence. These remarks are crucial for the reassessment of Laudan and Leplin’s solution that we will argue for below.

4.2 The second premise

4.2.1 Evidence and entailment

The next part of Laudan and Leplin’s argument is directed against the second premise of the problem, namely, that *entailment of the evidence is the only thing that is epistemically relevant for the confirmation of a theory*. They claim that this statement is an overly simplified and inaccurate view of the dynamics of

¹⁶ As we mentioned above, theoreticity requirements are based on basic goals of science, such as the quest for empirical, non-metaphysical knowledge, and of intelligible explanations and consistency.

¹⁷ These remarks prevent a possible objection. The reader might complain that in section 3 non-empirical virtues were dismissed as a full solution of the problem because of their context-dependency, but now another context-dependent feature, theoreticity, is being used as a part of the defended solution. However, as we just mentioned, theoreticity constraints block algorithms and so undermine the first premise of the problem. We are not using theoreticity as a criterion to make a choice between EE theories.

¹⁸ More precisely, it has not been *demonstrated* that algorithms of this kind cannot exist. However, it is extremely unlikely—given the non-*a priori* character of the theoreticity requirements—that an algorithmic procedure could include a recipe for obtaining plausible hypotheses.

¹⁹ Notice that theoreticity constraints also block the holist Duhem-Quine thesis as providing support for the universal scope of the first premise of the problem. As Adolf Grünbaum showed, the Duhem-Quine thesis ‘nor other logical considerations can guarantee the deducibility of O' [the class of observational consequences] from an *explanans* constituted by the conjunction of H and some *non-trivial* revised set A' of the auxiliary assumptions which is logically compatible with A under the hypothesis H' (Grünbaum 1960, 77). Suppose rival hypotheses H and H' are given, and suppose that a *crucial experiment* to test them favors H' . The Duhem-Quine thesis implies that it is always *logically* possible to save H by arranging the set of auxiliary assumptions A and replacing it by A' , so that the outcome of the experiment could be accommodated. In that case, we could always have a case of EE between H and H' . Grünbaum shows that this *logical* feature is not enough to prove that there will be a suitable A' of *non-trivial* assumptions for H to accommodate the observations. In our context, we could simply replace *non-trivial assumptions* for *assumptions that accomplish theoreticity constraints*. Moreover, these remarks are quite coherent with Duhem’s concept of *good sense*. The relevance of this concept for what Duhem really meant with his holistic thesis has been commonly undervalued.

evidential confirmation. According to them, a correct assessment of the nature of evidence and confirmation shows that ‘significant evidential support may be provided a theory by results that are not empirical consequences of the theory’ (1991, 460). If theories can obtain evidential support from empirical facts which do not belong to the class of their observational consequences, then—in the context of our problem—‘the relative degree of evidential support for theories is not fixed by their empirical equivalence’ (ibid). Therefore, the fact that two theories are EE does not *imply* that the choice to be made between them is empirically underdetermined.

Laudan and Leplin refer to two principles of the dynamics of empirical confirmation to support their view. First, if a theory T entails the statements e and S —where e is observational and S is observational or theoretical—, and if those two statements are logically independent, then if e is true it counts as evidential support for T and also for S . The non-consequential evidential relationship between e and S is based on a logical entailment after all, for the confirmation of S by e occurs *via* the general statement T that entails both of them. However, this is enough to show that the second premise of the problematic argument is not correct:

Allowing to a statement to accrue indirect evidential support in this fashion already undermines the claim that statements are confirmable only by their empirical consequences. This result alone suffices to establish that the class of empirical consequences of a statement and the class of its prospective confirming instances are distinct (ibid., 461).

The second principle of evidential confirmation that Laudan and Leplin invoke is that if a theory T entails two logically independent theoretical hypotheses H_1 and H_2 , and if in turn these hypotheses entail the classes of observational consequences E_1 and E_2 , respectively, then the truth of any member of E_1 will support H_1 and *also* H_2 —even though H_2 does not entail any statement in E_1 —(the same holds, *mutatis mutandis*, for the truth of the statements in E_2). (ibid., 461-462).

Once again, this principle implies that the class of the observational consequences a theory entails is not identical with the class of observational statements that can confirm that theory. In turn, this last remark implies that the EE of two theories is not a sufficient condition for the UD of the choice between them—and thus a way out of the problem becomes available:

Theoretical hypotheses H_1 and H_2 are empirically equivalent but conceptually distinct. H_1 , but not H_2 , is derivable from a more general theory T , which also entails another hypothesis H . An empirical consequence e of H is obtained. e supports H and thereby T . Thus, e provides evidential warrant for H_1 , of which it is not a consequence, without affecting the credentials of H_2 (ibid., 464).

It is important to underline that the relation between the subsumed H_1 and the more general theory T does not need to be a strict *logical consequence*. Laudan and Leplin state that H_1 can be reduced to a special case of T even if T is not (yet) a fully-developed or a very precisely-formulated theory. It is enough if T is a plausible and well-grounded *program*.

Laudan and Leplin’s attack on the two premises of our problem can be nicely complemented by remarks made by Richard Boyd (1973). As Boyd points out, given two EE theories, their different inter-theoretical

connections with background knowledge may be invoked to make a decision based on evidential grounds. Given EE theories, the background knowledge available might be such that it is—or becomes—at odds with essential hypotheses in one of the theories, but completely coherent with the other one. The friction between the rest of the well-confirmed theories that constitute the background knowledge and the core-structure of one of the EE theories can count as indirect empirical evidence to reject the latter.

Boyd explains his point by an example. He takes the famous Poincaré-Reichenbach argument for the conventionality of geometry as an instance of two EE theories. $F \& G$ is a theory which asserts that the world is governed by a class of forces F , and that its spatial features are described by a geometry G . $F' \& G'$ is a rival theory asserting that the world is governed by the class of forces F' —a class containing all the forces in F plus a universal force f' —and that its spatial features are described by the geometry G' . The theories are EE, however,

even though “ $F \& G$ ” and “ $F' \& G'$ ” have the same observational consequences (in the light of currently accepted theories), they are not equally supported or disconfirmed by any possible experimental evidence. Indeed, *nothing* could count as experimental evidence for “ $F' \& G'$ ” in the light of current knowledge. This is so because the force f' required by F' is dramatically unlike those forces about which we know—for instance, it fails to arise as the resultant of fields generating in matter or in the motions of matter. Therefore, it is, in the light of current knowledge, highly implausible that such a force f' exists. Furthermore, this estimate of the implausibility of “ $F' \& G'$ ” reflects *experimental* evidence against “ $F' \& G'$ ”, even though this theory has no falsified observational consequences (Boyd 1973, 7-8).

Boyd’s view is a good complement to Laudan and Leplin’s argument because it relies on similar grounds. First, that inter-theoretic relationships can count as indirect evidence to accept or reject theories shows that the class of statements which are confirmationally relevant for a theory does not reduce to the class of its observational consequences, just as Laudan and Leplin claim. Second, the development of background knowledge over time is crucially relevant for Laudan and Leplin and also for Boyd’s view. Suppose that T and T' are EE and at the time of their formulation equally coherent with the rest of background knowledge. However, new well-confirmed theories might be deeply at odds with T' , but coherent with T . This feature is an evidential reason to choose T . Even though T and T' remain EE, the UD of the choice has been broken by inter-theoretical connections. That is, just as EE, UD is a time-indexed feature²⁰.

Laudan and Leplin’s view on the second premise, as complemented by Boyd, seems quite compelling. However, it is useful and fruitful to consider some criticisms that have been put forward against it in order to appreciate what has been really achieved.

²⁰ Boyd’s own position is that, in a case of EE between T and T' , the compliance of T with the form of causal explanations present in empirically successful theories in background knowledge counts as an indicator for the truth of T that is lacking in T' —for the explanations in T' do not have the mentioned form. The principle of confirmation just defended weakens Boyd’s original position in the sense that it is detached from any realist commitments (Boyd considers the problem of EE and UD as a threat to the realist), and at the same time generalizes it in the sense that possible friction with background knowledge is not given only by divergence from the canonical form of *causal* explanations.

4.2.2 Okasha's objection: Hempel's problem

Samir Okasha (1997) objected that Laudan and Leplin's argument falls prey to a problem that Hempel had already noticed in 1945. Okasha claims that Laudan and Leplin's argument is committed to the following two principles: *i*) if evidence confirms a hypothesis, then it also confirms any statement that entails the hypothesis; and *ii*) if evidence confirms a hypothesis, then it also confirms any statement that is entailed by the hypothesis. Hempel (1945, 103-104) labeled these two principles as the *converse consequence condition* and the *special consequence condition*, respectively²¹, and Okasha reminds us that Hempel had also noticed that a simultaneous commitment to these principles leads to a problem:

The absurdity that results is this: every statement confirms any other one. For consider any statement *S*. Every statement confirms itself, so *S* confirms *S*. By converse consequence, *S* confirms (*S* & *T*), since (*S* & *T*) → *S*. By special consequence, *S* confirms *T*, since (*S* & *T*) → *T*. This result holds for arbitrary *T*, and must therefore be regarded as a *reduction ad absurdum* of the simultaneous use of the special and converse consequence conditions (Okasha 1997, 253).

Okasha is certainly right in that Laudan and Leplin endorse both the special and converse condition. However, his criticism is not enough to threaten their argument. Hempel's problem comes up if we understand the dynamics of confirmation only as a matter of logic, but if we consider *theoreticity* conditions—more specifically, testability and non-superfluity—the problem does not automatically arise. In Okasha's reconstruction, *T* cannot be *any* arbitrary statement: it has to be testable and non-superfluous, i.e., it must be relevant for the derivation of at least some of the statements in the class of observational consequences derived from (*S* & *T*). If the only extra statement that can be derived from (*S* & *T*)—with respect to the ones derivable from *S* alone—is *T*, then (*S* & *T*) will not be considered a genuine theory.

One could still argue that the example of the problem that Hempel himself offered cannot be dismissed in this way (1945, 104-105). He took the theory *T* to be (*H*₁ & *H*₂)—where *H*₁ is 'all the ravens are black' and *H*₂ is 'Hooke's law'. The class of observational consequences of *T* is *O*_{*T*}, which is defined as (*O*₁ & *O*₂), where *O*₁ is the class of observational consequences of *H*₁ and *O*₂ is the corresponding class of *H*₂. Since *H*₁ is relevant for the derivation of *O*₁ and *H*₂ is relevant for the entailment of *O*₂, both hypotheses are testable and non-superfluous. In other words, the problem is now that any evidence confirming a theory or hypothesis could be used to confirm any other *scientific* hypothesis.

²¹ As Okasha asserts (1997, 254), Laudan and Leplin's argument can be schematized this way:

- i) H_1 and H_2 are EE
- ii) $T \Rightarrow H_1$
- iii) $T \not\Rightarrow H_2$
- iv) $T \Rightarrow H$
- v) $H \Rightarrow e$
- vi) $H_1 \not\Rightarrow e$,
- vii) $H_2 \not\Rightarrow e$,
- viii) e ,

therefore; ix) *e* confirms *T* (this requires the *converse consequence condition*), and then x) *e* confirms *H*₁ (this requires the *special consequence condition*); but *e* does not confirm *H*₂.

It is rather clear that such a maneuver is against good sense, and it would be certainly dismissed in scientific practice. The reason is, again, theoreticity. Two different hypotheses or theories can be fruitfully conjoined in order to form one single theory only if by so doing new observational consequences can be derived, or if by so doing unexplained phenomena become explained by the new theory—consequences and phenomena which could not be predicted or explained by means of any of the conjoined theories alone. Simply put, the resulting theory must be more than the mere sum of its parts. This should be adopted as a principle, otherwise cosmologists could simply conjoin string theory with genetics and then claim that the discovery of a new gen confirms that space-time has eleven dimensions.

Once again, the second premise of the problem relies on an oversimplified conception of the dynamics of empirical evidence. Logical entailment of an observational statement from a hypothesis is not a necessary condition for that statement to confirm the hypothesis. Inter-theoretical relations are also crucial features to be considered.

4.2.3 Bangu's objection: encompassing theories

Sorin Bangu (2006) introduced yet another objection against Laudan and Leplin's argument which is also illuminating. We saw above that a way out of the EE and UD problem can be found if there is a well-confirmed general theory T that encompasses only H_1 in the EE pair—the evidential support that e gives to T , although neither H_1 nor H_2 entails e , flows to H_1 but not to H_2 . Bangu claims that this does not work, for the possibility of yet another general theory capable to encompass H_2 has not been ruled out—and this alternative general theory may be also supported by the same evidence e :

The supporter of underdetermination can reply that nothing rules out the possibility that another theory T^* exists, such that $T^* \rightarrow H_2$ [H_2 being the other member in the EE pair]. Moreover, it is possible that T^* is supported by evidence e as well [...].

The only constraint imposed on the relation between T and T^* is that they behave differently with respect to H : T^* entails it, while T does not. What evidence supports each of these theories is another matter. So, can two different theories, each entailing different hypotheses, be supported by the same evidence? This is trivially true. (Bangu 2006, 273-274).

If a theory such as T^* were given, then the evidence e would also flow to H_2 , and the UD of the choice between H_1 and H_2 would come up once again. However, Bangu overlooks one further constraint on T^* for the UD of the choice to be restored: it is required that T and T^* are also EE—the evidence supporting each of the theories is relevant. Otherwise the case between H_1 and H_2 could be settled by means of the different evidential support between T and T^* . If T^* is a theory with more evidential support than T , then we should choose H_2 .

Bangu's objection is correct, but it is not enough to undermine Laudan and Leplin's argument. If there is a theory T which encompasses H_1 but not H_2 , and if there is no theory such as T^* , the evidence e does break the UD. It is important to emphasize that Bangu's argument is not based on algorithms, for he has not shown that given *any* theory such as T there is a theory such as T^* . Actually, this is a rather unlikely situation. EE

between theories is not a common feature in science—scientists look for better theories, not for equivalent ones. Moreover, most of the times it is a very difficult task to come up with *one* empirically successful theory with respect to a certain domain of natural phenomena, and Bangu’s reply requires not only one pair of EE theories, but two pairs.

However, Bangu’s argument is clarifying regarding the nature and scope of Laudan and Leplin’s solution. Their argument, as complemented by Boyd’s, is that UD is a contingent feature—even if two theories are EE non-consequential evidence that could be available might provide an evidentially justified reason to make a choice. Bangu’s argument implies that this breaking of UD may be undone by contingent scientific developments. He has effectively established that even if UD is broken à la Laudan and Leplin, this UD breakdown need not be a definitive resolution of the choice problem. So although Laudan and Leplin have shown that there are ways in which the underdetermination problem can be overcome, they have not shown that the problem cannot happen at all or cannot return.

5 Reassessing Laudan and Leplin

Laudan and Leplin’s attack on the first premise of our problem shows that EE is a time-indexed feature and that there is no guarantee that it is universal in scope. Theoreticity conditions block the effectiveness of algorithms to produce EE rivals, and the development of science can be such that variation in the class of available auxiliary hypotheses may break the EE between two theories. Their attack on the second premise shows that the class of observational statements that can count as evidence for a theory is not limited to the observational consequences of the theory. The UD between two EE rivals can be broken by subsuming one of the theories in the EE pair under a more general and well-confirmed theory, whose particular evidential support flows to the encompassed theory but not to the non-encompassed rival. In addition we have Boyd’s argument: the inter-theoretic connections of a theory can work as indirect evidential (dis)confirmation. A theory in an EE pair could be rejected if it contains essential parts which are (or become) at odds with respect to the background knowledge—whereas its rival is (or remains) coherent with it. Laudan and Leplin draw the following conclusion: ‘The thesis of underdetermination, at least in so far as it is founded on presumptions about the possibility of empirical equivalence for theories—or ‘systems of the world’²²—stands refuted’ (1991, 466; our italics).

But this conclusion goes too far. Laudan and Leplin have certainly shown that it may happen that there is a way out of EE and UD situations, but this is not enough to *refute* the problem. Even if EE is a time-indexed feature and theoreticity requirements can block algorithms, that two theories may be EE remains a possible scenario—a genuinely scientific EE rival to a given theory might be found after all. Further development of science and consequent variation of the available auxiliary hypotheses might break the EE, but it is also possible that this will not happen. It is also true that a more general theory could break the UD by means of

²² In this passage Laudan and Leplin cannot be using the term ‘systems of the world’ in its canonical meaning. See (Hofer and Rosenberg 1994).

‘transferring’ empirical evidence. However, there is nothing in science or in epistemology that assures that such a theory will be actually available. Moreover, Bangu has shown that there is nothing in science or in epistemology that precludes that an alternative general theory that restores the UD may be formulated. Finally, recourse to inter-theoretic connections is not a guaranteed way out either. It is possible that if a pair of EE is given both theories are equally coherent with respect to background knowledge—further development of science could be at odds with one of the theories in the pair, but nothing can assure that either.

These remarks demonstrate that EE leading to UD *can* happen in science, and that the success of the solution provided is not *guaranteed* from the outset. Therefore, Laudan and Leplin’s stance is not enough to *refute* the problem. What their argument has achieved is to show that there are typical goals and methods in science which *might be* effective in overcoming the problem if it comes up. In other words, though Laudan and Leplin’s argument does not refute the problem of EE and UD, it clarifies that we are dealing with a problem that science may solve. The tools of scientific practice that could solve it do not come with a guaranteed-success certificate—we do not know in advance if science will be always effective in finding its way out. But, after all, this is a feature that all the problems that science is to solve share—so the problem of UD, given EE, is a *scientific* problem, not different in principle from theory making in general.

5.1 Evidential support vs. theoretical virtues

The difference between the (contingent) solution based on non-consequential evidence and the partial solution given by non-empirical virtues is that the former can provide a uniquely determined evidence-based choice between EE theories. We mentioned in section 3 that even though non-empirical virtues of one of the theories may ground a rational preference, the alternative choice might be also rationally grounded. On the contrary, encompassing theories and inter-relations with background knowledge can result in evidential support distinctly favoring one of the rivals, thus providing an objective and uniquely determined choice.

Empirical evidence is more basic than non-empirical virtues when it comes to theory choice. Think of a case of two EE rivals T and T' such that both theories totally comply with theoreticity requirements. Suppose that, from the non-empirical virtues point of view, T' is clearly superior to T , in all features and in all contexts. Suppose also that T is encompassed by a more general theory H , which has been evidentially confirmed but that is incompatible with T' . Suppose finally that T' is also incompatible with another well-confirmed theory P —whereas T is consistent with P . The arguments by Laudan and Leplin and Boyd say that in this case—and in spite of its non-empirical inferiority— T has more evidential support than T' ; and it is quite plausible that the greater degree of evidential support for T would determine the choice made by scientists (at least in the longer run)²³.

²³ This presupposes that, in principle and regardless of cases of EE, empirical evidence can count as a fully objective criterion for theory acceptance. There are some stances in the philosophy and sociology of science that reject this presupposition. A famous example is given by certain interpretations of the work of Thomas Kuhn. A refutation of these positions is not possible here for reasons of space.

Empirical confirmation by means of non-consequential evidence, even though it is indirect, is empirical evidence after all. In the case of the encompassing-theories solution, we saw that if we hold on to two very plausible principles of confirmation—namely, that evidence for a scientific hypothesis h is also evidence for scientific hypotheses that entail h , and that evidence for a scientific hypothesis h is also evidence for the statements that h entails—it follows quite neatly that evidential confirmation for a scientific hypothesis can come from observational statements that are not derived from it. That is, non-consequential evidence must be included in our model of confirmation if we accept the two mentioned principles.

In the case of (dis)confirming evidence coming from inter-theoretical relations we have something similar. This time, though, a different kind of principle is presupposed. Suppose again that T and T' are EE, that T is consistent with another well-confirmed theory P , and that T' is at odds with it. The evidential support for P counts as empirical evidence against T' *granted that we agree that consistency between the theories we accept is a basic principle of science*. If we want that our theories are mutually consistent, then Boyd's argument can be taken as a sound principle in the dynamics of empirical confirmation.

In order to clearly recognize that this consistency principle is a basic tenet of science it is enough to imagine a case in which two well-confirmed theories are mutually inconsistent. Suppose that in a certain domain of physics theory T is introduced and that all of its predictions are confirmed, that in a different domain theory P is proposed and all its predictions are confirmed, and that P and T are incompatible. This situation, of course, would be taken as a serious problem for science, and it would be expected that endeavors in order to show that one of the theories must be given up would be undertaken by scientists²⁴.

These remarks should not be taken to imply that non-empirical virtues and empirical evidence are two completely unrelated concepts—we are not arguing that non-empirical virtues must necessarily be merely pragmatic, subjective and context-dependent, totally unconnected to empirical evidence. Our stance is compatible with the account of theoretical virtues proposed by James McAllister (1989). He argues that 'indicators of beauty' that can work as relevant criteria for theory choice are grounded on meta-inductions on the aesthetic aspects of empirically successful theories of the past²⁵. That is, the criteria that define non-empirical virtues may be indirectly based on the empirical success of theories.

Two features of McAllister's view illustrate its coherence with our argument. First, *because* so-called non-empirical virtues may be inductively linked to the empirical success of theories, they are subordinated to empirical evidence when it comes to theory choice. As McAllister states:

²⁴ The epistemological justification for the consistency principle is quite simple. If we aspire to obtain knowledge of reality by means of scientific theories, it is clear that if the set of scientific theories we accept is inconsistent, we would hardly call such set 'knowledge'. To say that 'we *know* that $2 + 2 = 4$ and that $2 + 2 = 5$ ' is clearly nonsense.

²⁵ 'A community selects its aesthetic canon at a certain date from amongst the aesthetic features of all past theories by weighting each feature proportionally to the degree of empirical success scored to that date by all the theories which have appeared to embody it. The community's aesthetic canon is then composed of the set of such mutually consistent features which have gained the greatest weighting. This is a clearly inductive procedure: as a theory demonstrates empirical success its aesthetic features will gain proportionate weight within the canon which is to serve in the evaluation of current theories, while conversely the aesthetic features of a theory which suffers a streak of empirical failures will win a progressively lesser weighting in theory-reference' (McAllister 1989, 39).

T. H. Huxley's aphorism about 'the great tragedy of Science—the slaying of a beautiful hypothesis by an ugly fact—which is so constantly being enacted under the eyes of philosophers' aptly describes the lag of aesthetic appreciation behind empirical assessment. The perceived beauty of a hypothesis is a function of the observational success of antecedent theories aesthetically similar to it; the novel fact appears as yet ugly because unassimilated within a theory of which the aesthetic qualities have been sufficiently weighted by the community. In time the community's indicators of beauty will evolve to render the theory erected about the new fact a structure of sovereign beauty and the disproven hypothesis merely passé. (ibid., 39-40).

Second, McAllister states that 'indicators of truth', the evidential criteria which are relevant for theory choice—here we include the non-consequential aspects of confirmation explained above—are determined by the basic goals of science²⁶. Thus, indicators of truth are much more stable than indicators of beauty. Empirical success remains the ultimate criterion for theory choice even after scientific revolutions, whereas the canons of theoretical virtues are intrinsically related to a specific state of science, and to the specific epistemological credos of scientists. Therefore, 'indicators of beauty' are not able to provide an objective and uniquely determined choice in cases of EE and UD:

The hope that indicators of beauty will defeat the threat of underdetermination is incidentally revealed illusory: any decision on aesthetic grounds between empirically equivalent theories will in general be perceived as valid only within the paradigm then current and cannot hence be considered definitive. (ibid., 44).

Notwithstanding the epistemic respectability of non-empirical virtues, the main difference between a solution based on this kind of features and the solution based on non-consequential evidence is that only the latter can lead to a uniquely determined, fully objective choice that is justified according to the ultimate tribunal of theory choice—empirical evidence.

6 An example

Between 1886 and 1904 the Dutch physicist Hendrik Antoon Lorentz developed a theory with the aim of solving the conflict between experiments and Maxwellian electrodynamics that arose during the second half of the 19th century. In order to achieve that aim, Lorentz's theory introduced several radical modifications to the Newtonian framework of physics. The main elements of his theory were *i*) a set of coordinate transformations under which Maxwell's equations remain invariant—the famous *Lorentz-transformations*; *ii*) the concept of 'local time'—the time that is experimentally determined in a frame of reference in motion with respect to the ether as opposed to the 'real time' that is measured only in the ether-rest frame—; *iii*) the

²⁶ 'Metarationalism is clearly responsible for the genesis of indicators of truth because their inclusion among the desiderata of theories derives entirely from the a priori definition of the goal of science, the complete and true explanatory account of the universe. The requirements of internal consistency or predictive accuracy are prized not because they have previously been witnessed to accompany verisimilitude but because they are the elements of an explication of that very concept: indicators of truth appear in other terms to provide not a mere ampliative connotation but rather an analytic definition of truthlikeness. It remains of course possible for indicators of truth to be inductively learned by a scientific community but this is irrelevant to the a priori logical status of such criteria' (McAllister 1989, 38). In order to retain neutrality regarding the realism-antirealism schism, we can replace 'indicators of truth' for 'indicators of empirical success'.

Lorentz-Fitzgerald length-contraction hypothesis that affects bodies in motion through the ether; and *iv*) a model of the electron in which this elementary charged particle was subject to the Lorentz-Fitzgerald contraction and its mass was velocity-dependent²⁷.

Lorentz's theory was formulated against the backdrop of a basic ontology of an immobile ether pervading all of space, and of electrons, the ultimate (charged) particles of nature whose mass was of entirely electromagnetic origin—the mass of electrons, and *a fortiori* of all of matter, was due to the electron's electromagnetic field interaction with the ether²⁸.

Lorentz's eventual theory was empirically successful and was capable of providing an explanation for all of the main physical puzzles of the time: stellar aberration, Fresnel's coefficient, Fizeau's experiment, why massive solid objects can move through the ether without disturbing it, and (the negative results of) the experiments designed to find optical effects caused by the motion of the earth through the ether—the most famous of them being the Michelson-Morley experiment²⁹.

In 1905 Albert Einstein published his first paper on the special theory of relativity. Its two principles—that all inertial frames of reference are physically equivalent, and that the speed of light is constant and independent of the state of motion of its source—along with his revolutionary conception of simultaneity and the nature of time, allowed him to explain and predict everything that Lorentz's theory explained and predicted.

Therefore, Lorentz's theory is usually considered as EE to Einstein's special relativity. This is correct, although some cautious is needed. Lorentz's theory of 1904 is not yet strictly empirically equivalent to Einstein's theory. For example, as Michel Janssen (2003) has shown, the theories predict a different outcome for the Trouton experiment. The full equivalence of the theories is obtained only if several crucial amendments and contributions introduced by Henri Poincaré are taken on board³⁰.

On the other hand, the reason why the theories were *rivals* is that special relativity claims that we live in a space-time determined by a *Minkowskian metric*—so that the time coordinates of events and the spatial dimensions of objects are essentially relative to the frames in which they are measured—whereas Lorentz's theory says that we live in a *Newtonian* space-time described by a *Euclidean metric* plus *absolute time*, but in which certain dynamical effects—caused by the interaction between matter and the fields propagating in the ether—*deceive* observers who move with respect to the ether and make time and length measurements, so

²⁷ For an excellent presentation and analysis of Lorentz's theory see (Janssen 1995, chapter 3).

²⁸ For an analysis of the physical and epistemological framework of Lorentz's theory see (McCormmach 1970).

²⁹ For the scientific context and motivations of Lorentz's invention of his theory, see (Hirosige 1976; Janssen and Stachel 2004).

³⁰ Poincaré's contributions and corrections on Lorentz's work were the following: *i*) he showed that 'local time' was not a mere mathematical tool, as Lorentz originally claimed, for it was connected to observable effects in the behavior of moving clocks; *ii*) the introduction of a *fictitious fluid* in the ether that carried an amount of electromagnetic momentum; *iii*) the introduction of the *Poincare-pressure* which kept the moving electron stable and precluded its explosion due to Coulomb forces; *iv*) he corrected Lorentz's expressions for the transformation of velocities and charge density between moving frames; and *v*) he showed that the Lorentz's transformations form a *group*, and by so doing he showed that they are fully symmetric. Only with these amendments Lorentz's theory becomes completely predictively equivalent to special relativity. See (Cuvaj 1968; Darrigol 2006; Goldberg 1967; Miller 1973).

that the world looks Minkowskian to them. In other words, the theories assign different chrono-geometrical structures to the space-time we inhabit.

Therefore, in 1906 physics was facing an instance of our problem³¹: there were two predictively equivalent theories and the choice between them was underdetermined by empirical data. We know that in the end Einstein's theory won the competition. The historical course of events that led to this victory was rather complex, though. For example, between 1906 and 1909 the scientific community often spoke about the *Lorentz-Einstein theory*, as the difference between the two rival theories had not been generally recognized. Clarification concerning this rivalry had to wait until Minkowski's groundbreaking work (1908). A couple of years later, around 1911, the expression *Lorentz-Einstein theory* disappeared from the vocabulary of physics, and special relativity was adopted as the main-stream theory³².

Some non-empirical features of Einstein's theory have undoubtedly been historically relevant for its victory; especially its comparative simplicity as a theory based on only two principles (though from the beginning there have also been complaints about relativity's lack of understandability as judged from an intuitive viewpoint)³³. In the philosophy of science literature one can also find several arguments for explaining and justifying why special relativity defeated Lorentz's theory that invoke non-empirical features as the decisive factor. The accusation of ad-hocness leveled against the length-contraction hypothesis is a well-known older example. More recently, Michel Janssen (1995; 2002a; 2002b; Balashov and Janssen 2003; 2009) has proposed and defended an argument of this kind. In a nutshell, Janssen states that special relativity must be (and was) preferred because of its explanatory superiority. He argues that in Lorentz's theory the fact that all laws of physics are Lorentz-invariant remains an unexplained coincidence—in special relativity, however, this feature gets naturally explained by the structure of Minkowski space-time.

Notwithstanding the plausibility of the analysis by Janssen and of other arguments based on non-empirical features (we defer a detailed evaluation of Janssen's interesting views to another paper), in section 5.1 we have argued that even though features relating to theoretical virtues can provide good pragmatic reasons to prefer one theory over another in a case of EE, they are not able to solve the problem in a fully objective and uniquely determined way. We shall now show for the specific case of Lorentz vs. Einstein how it can be assessed from the empirical side, using non-consequential empirical evidence, and how in this way the decision in favor of special relativity can be objectively and uniquely determined.

One of the main reasons why Lorentz's theory actually got rejected and special relativity accepted is in fact an instance of the non-consequential confirmation solution of our problem. In 1899 Max Planck introduced the concept of the *quantum of energy* in order to derive the correct law for black-body radiation. Before Planck's trailblazing work, the radiation spectrum of black-bodies had been an intractable problem for classic electrodynamics and thermodynamics. In 1905, a few months before his paper on special relativity

³¹ Poincaré published his *On the Dynamics of the Electron*—the work where he introduced the amendments and developments of Lorentz's theory that make it predictively equivalent to special relativity—in 1906. For a detailed presentation and analysis of this work, see (Miller 1973).

³² For a historical treatment of the formulation and early reception of special relativity see (Miller 1981).

³³ See (Brush 1999).

appeared, Einstein himself published his famous work on the *light-quantum hypothesis*—a hypothesis that relied on Planck’s concept. The relevance of the rise of quantum physics for our subject is that the quantum hypothesis was soon acknowledged to be deeply at odds with the foundations of classic electrodynamics. In a nutshell, the problem was that classic electrodynamics and thermodynamics predict that an accelerated electron must emit radiation of all wavelengths in a continuous range of energy, whereas the quantum hypothesis postulated emission in determinate, specific wavelengths in a discrete spectrum. Since Lorentz’s ether theory and his model of the electron theory were essentially built upon the very core of classic electrodynamics, namely, Maxwell’s equations, the groundbreaking new physics of the quantum led physicists to gradually abandon it. The problem got even deeper with Bohr’s first contributions on the structure of the atom, for the quantum hypothesis and the abandonment of classical electrodynamics were central features in Bohr’s work.

As Russell McCormach reports (1970, 486-487), Hendrik Lorentz himself played a central role in the recognition of the conflict between quantum physics and the core of classic electrodynamics. Between 1900 and 1903 he tackled the task of finding the dependence of the energy density in a blackbody as a function of its temperature and the wavelength of the radiation. The formula he derived on the basis of electrodynamics applied to thermodynamics—essentially the Rayleigh-Jeans law—worked in the long-wavelength part of the spectrum of emission, but failed in the short wave-length part of the spectrum³⁴. Although he first took this result as very promising, he soon realized that there was a deep conflict between the quantum hypothesis and the results of experiments, on one side, and the core of electrodynamics on the other:

In 1908 Lorentz came out in support of Planck’s theory; it was then that he emphasized the profound antithesis between the quantum hypothesis and the electron theory. At a mathematical congress in Rome that year Lorentz spoke on Planck’s and James Jeans’ theories of blackbody radiation. His object was to prove that the union of the electron theory with Hamilton’s equations of motion and J. W. Gibbs’ statistics leads inescapably to Jeans’ radiation law, which, like his own of 1903, agrees with experience only in the case of long wavelengths. He said that the alternative, Planck’s theory, demands far-reaching changes in electron theory. He pointed out that this is easily seen, since an accelerated electron should emit rays of all wavelengths, a result incompatible with the hypothesis of energy elements whose magnitude depends on wavelength. At the time of his lecture he had not yet decided between the two theories. Wien, however, called his attention to experiments showing that for short wavelengths a body emits much less light in proportion to its absorbing power than that predicted by Jeans’ theory. This proves, Lorentz said in a note appended to the published version of his talk, that any theory that bases itself on the electron theory and the equipartition theorem has to be profoundly revised. Later in that year he elaborated that note: he had ‘long hoped’, he confessed, ‘that it would be possible to escape the universal applicability of that theorem [equipartition] by combining electron theory and kinetic theory’. He added, ‘this hope has not been fulfilled’. He was now ready to concede that the interaction of matter and ether takes place by means of vibrating charged particles to which Gibbs’ statistics, for unknown reasons, are inapplicable.

³⁴ ‘Lorentz pointed out that his black-body formula agrees with the long wavelength limit of the quantum formula that Planck had derived in 1900, a coincidence which struck him as highly remarkable considering the widely different assumptions in the two cases. It was characteristic of Lorentz to spell out what was incomplete in his work and what was still unknown; he stressed that his theory is valid only for long wavelengths and that Planck’s applies to the whole spectrum. So it was Lorentz, an originator of the electron theory, who first intimated the possible limits of the theory. Starting from the electron theory and from a mechanism appropriate to the theory, he arrived at the limiting case of the radiation law; and he did not see how to extend his theory to Planck’s general case.’ (McCormach 1970, 486-487).

Lorentz thus accepted the quantum theory as the only theory capable of explaining the complete spectrum of black-body radiation, while at the same time regarding it as very incompletely understood in its connection with the other branches of physics and in particular with electron theory. (ibid., 487).

The conflict between electrodynamics and quantum physics that Lorentz helped to clarify became more and more important for the scientific community in the following years, and it thus was one of the main reasons that led to the abandonment of the electromagnetic world-view program—of which Lorentz's ether theory was a fundamental part—leaving the path open for Einstein's special relativity to be accepted:

The sense of the first Solvay Congress in 1911 was that the electron theory was incompatible with quanta and that it could not be made compatible without far reaching reform. The Congress and especially its published proceedings went far to redefine the fundamental problems for fundamental physical theory. Niels Bohr's doctoral dissertation in 1911 was a reformulation of Lorentz's theory of metals on more general principles. In his dissertation Bohr pointed to persuasive evidence of the ultimate incompetence of mechanics and electrodynamics on the molecular level. His 1913 quantum theory of atoms and molecules, which gave sharp focus to the quantum problems and intimated their enormous fruitfulness, was based on the explicit denial of the validity of ordinary mechanics and the classical electron theory in the atomic domain. Lorentz's theory continued to be worked on, but its concepts were increasingly recognized as unsuited for the basic reconstruction of physical theory demanded by the quantum hypothesis. (McCormach 1970, 488)

Einstein's special relativity, though predictively equivalent to Lorentz's theory, did not rely on any electrodynamic assumptions about the ultimate nature of matter. Consequently, it was not at odds with quantum physics. As Minkowski's seminal contributions clarified, special relativity was a theory grounded on the kinematics of the four-dimensional continuum (Minkowski space-time), and this made superfluous any specific assumptions about dynamics. This difference between Einstein's and Lorentz's theory was crucial for the abandonment of the latter and the acceptance of the former.

Considering that quantum physics became more and more important during the first two decades of the 20th century, it was rather natural that, given two predictively equivalent theories, the one which was not at odds with quantum theory was to be accepted. Indeed, as already mentioned, by 1909 the expression *Lorentz-Einstein theory* started to disappear from the physicists' vocabulary. Minkowski's work had shown that they were two theories of a different nature, and by 1911 the rise of quantum mechanics turned the balance decidedly in Einstein's favor:

Why did the electromagnetic program run out of power? [...] More important was the competition from other theories that were either opposed to the electromagnetic view or threatened to make it superfluous. Although the theory of relativity was sometimes confused with Lorentz's electron theory or claimed to be compatible with the electromagnetic worldview, about 1912 it was evident that Einstein's theory was of a very different kind. It merely had nothing to say about the structure of electrons and with the increasing recognition of the relativistic point of view, this question—a few years earlier considered to be essential—greatly changed in status. To many physicists it became a pseudo-question. As the rise of relativity made life difficult for electromagnetic enthusiasts, so did the rise of quantum theory. Around 1908, Planck reached the conclusion that there was a fundamental conflict between quantum theory and the electron theory, and he was cautiously supported by Lorentz and other experts. It seemed that there was no way to derive the blackbody spectrum on a purely electromagnetic basis. As quantum theory became more and more important, electron theory became

less and less important. The worst thing that can happen to a proclaimed revolution is that it is not needed. (Kragh 1999, 115).

Clearly, this situation is an instance of one of the ways out of the UD problem explained in section 4. As mentioned there, inter-theoretical relations of theories with background knowledge can count as (non-consequential) empirical evidence for or against a theory, and this evidence may be able to break the UD between two EE theories. This is exactly what happened in the Einstein *vs.* Lorentz case. The rise of quantum physics implied that new background knowledge became at odds with the core of Lorentz's theory, whereas this friction did not occur in the case of Einstein's special relativity. Therefore, inter-theoretical connections with background knowledge resulted in indirect empirical evidence against Lorentz's theory, whereas Einstein's remained unharmed. In spite of the predictive equivalence, the empirical UD of the choice got broken.

There is another reason to prefer special relativity over Lorentz's theory that exemplifies one of the ways out of the problem that Laudan and Leplin propose. In 1916 Einstein finished his general theory of relativity. In this gravitational theory, the Minkowski space-time described by special relativity becomes a special case. General relativity tells us that Minkowski space-time depicts a flat spatiotemporal region without mass-energy, or an infinitesimal part of a curved space-time that does contain mass-energy—just as in Gauss geometry an infinitesimal region of a curved surface approximates to a Euclidean tangent plane. That is, special relativity becomes a special case of the general theory, both in the mathematical and the physical sense.

Lorentz's theory—in spite of its predictive equivalence with respect to special relativity—does not fit within general relativity. As mentioned above, it claims that the physical world, in its spatiotemporal features, has the structure of Newtonian space-time. Therefore, even if we take Lorentz's theory as holding for a local region of a global space-time, it is in conflict with the meaning of general relativity. Einstein's gravitational theory claims that empty space-times or local regions of space-times containing matter have a Minkowskian metric. Despite the mathematical and empirical equivalence between Lorentz's theory and special relativity, the former cannot be understood as a special case of general relativity; the theories are incompatible.

As a result, special relativity possesses greater empirical support than Lorentz's ether theory. General relativity entails predictions that special relativity cannot entail on its own—it allows a satisfactory description of the motion of the perihelion of Mercury, and it predicts that light gravitates and 'bends', for example. However, since the special theory is a special case of the general one, the empirical support for the latter *flows* to the former: the perihelion of Mercury and the light-bending effect are empirical evidence for the Minkowskian metric of a local region of space-time. On the other hand, since Lorentz's theory is incompatible with general relativity and thus cannot be included in it, the empirical support of Einstein's gravitational theory cannot flow to Lorentz's ether theory³⁵.

³⁵ Unlike the relation between Lorentz's theory and quantum physics, the connection between special relativity and general

This is an instance of Laudan and Leplin's argument explained in section 4.2. Given two EE theories, H_1 and H_2 , such that only H_1 can be encompassed by a more general theory T , the independent empirical evidence for T flows to H_1 but not to H_2 —and thus the UD of the choice gets broken. In our example H_1 corresponds to special relativity, H_2 to Lorentz's theory, and T to general relativity.

The case of special relativity vs. Lorentz's theory illustrates that the way out of the problem of EE and UD reviewed above is actually a part of scientific practice. This example also shows that the reassessment of the solution proposed above is in principle correct. Even though the arguments considered show that the regular practices of science are capable of solving the problem of UD, they are not able to *refute* the problem.

Two features of the Lorentz vs. Einstein case clarify and illustrate this further. First, there was a period in which the choice between Einstein's theory and Lorentz's was indeed underdetermined. Before it became generally recognized that classic electrodynamics—and *a fortiori* Lorentz's theory—was incompatible with the results of quantum physics there were no empirically grounded reasons to make a choice. That is, the Einstein vs. Lorentz case testifies that EE and UD can actually happen in science.

Second, the incompatibility between quantum physics and classic electrodynamics is a *contingent* feature of science. Quantum physics could have been compatible with Lorentz's theory and not with Einstein's—in that case the choice would have been the opposite—, it could have been compatible with both of them, or it could have been never formulated—and in these two cases the UD would not have been removed. In other words, that the further development of science was to break the UD was not assured from the outset. In the case of general relativity something similar holds. That Einstein's gravitational theory was to encompass special relativity, but not Lorentz's theory, was not assured in advance. Moreover, and following Bangu, Lorentz or another scientist might have formulated an alternative encompassing theory that could have restored the UD situation.

This leads us to the conclusion that even though EE and UD is not a universal and necessarily everlasting problem affecting all theories at all times, it can arise as a problem after all. However, if it does come up, it is a problem for which science, in principle, has the tools to find a way out; and just as in all problems that science attempts to solve, a breakthrough might eventually be found—but it also might not. The problem of EE and UD is a *contingent* one. If with respect to a certain stage of the development of science we have a pair of EE theories for which inter-theoretical connections are not able to break the UD of the choice to be made, we can only have recourse to non-empirical features in order to justify a preference, but, as stated above, such features are not able to assure a uniquely determined and evidence-based decision. Nevertheless,

relativity was not *historically* relevant. General relativity got completed in 1916—and empirically tested in 1919—and by then special relativity was already generally accepted by the scientific community, whereas Lorentz's theory had been put aside. Our claim that the connection between special and general relativity grounds a reason to choose Einstein's theory instead of Lorentz's is thus only conceptual, not historical.

since the problem is contingent, we can still keep the hope that future development of science might provide a way out grounded on empirical evidence³⁶.

References

- Balashov, Y. and Janssen, M. 2003. Presentism and Relativity, *British Journal for the Philosophy of Science*, 54, 327-346.
- Bangu, S. 2006. Underdetermination and the Argument from Indirect Confirmation, *Ratio*, 19, 269-277.
- Boyd, R. 1973. Realism, Underdetermination, and a Causal Theory of Evidence, *Noûs*, 7, 1-12.
- Bunge, M. 1961. The Weight of Simplicity in the Construction and Assaying of Scientific Theories, *Philosophy of Science*, 28, 120-141.
- Brush, S. 1999. Why Was Relativity Accepted?, *Physics in Perspective*, 1, 184-214.
- Cuvaj, C. 1968. Henri Poincaré's Mathematical Contributions to Relativity and the Poincaré Stresses, *American Journal of Physics*, 36, 1102-1113.
- Darrigol, O. 2006. The Genesis of the Theory of Relativity, in T. Damour, O. Darrigol, B. Duplantier and V. Rivasseau (eds), 2006, *Einstein 1905-2005: Poincaré seminar 2005*, Basel: Birkhäuser, 1-31.
- De Regt, H. W. and Dieks, D. 2005. A Contextual Approach to Scientific Understanding, *Synthese*, 144, 137-170.
- Duhem, P. 1954. *The Aim and Structure of Physical Theory*, New Jersey: Princeton University Press. Chapter VI reprinted in S. Harding (ed.), 1976, *Can Theories Be Refuted? Essays on the Duhem-Quine thesis*, Dordrecht: Reidel, 1-40.
- Goldberg, S. 1967. Henri Poincaré and Einstein's Theory of Relativity, *American Journal of Physics*, 35, 934-944.
- Grünbaum, A. 1960. The Duhemian Argument, *Philosophy of Science*, 27, 75-87. Reprinted in S. Harding (ed.), 1976, 116-131.
- Hempel, C. 1945. Studies in the Logic of Confirmation (II), *Mind*, 54, 97-121.
- Hirosige, T. 1976. The Ether Problem, the Mechanistic Worldview, and the Origins of the Theory of Relativity, *Historical Studies in the Physical Sciences*, 7, 3-82.

³⁶ Notice that although non-empirical features can provide reasonable grounds for making a choice, a decision based on this kind of features runs the risk of contradicting empirical evidence. If we face a case of EE and UD, and if we choose one of the theories in terms of non-empirical virtues, it is still possible that the future development of science may break the EE and/or the UD favoring—in terms of empirical evidence—the theory that we rejected; and this is so even if the rejected theory is inferior in all contexts and in all aspects with respect to pragmatic non-empirical features. It is true that theory acceptance on the basis of empirical evidence is also risky. Any well-confirmed theory might prove empirically wrong in the long run. But a choice based on non-empirical features implies a risk in the sense that further development of science could demonstrate that, according to the *ultimate* criterion of theory choice—empirical evidence—our choice is wrong. Thus, rather than a *choice*, what non-empirical features can rationally ground is a (pragmatic) *preference*. In the context of McAllister's account of non-empirical virtues, the risk comes from the fact that the meta-induction is Humean: there are no necessary connections between aesthetic features and empirical success, only contingent correlations. See (McAllister 1989, 40-41).

- Hoefer, C. and Rosenberg, A. 1994. Empirical Equivalence, Underdetermination, and Systems of the World, *Philosophy of Science*, 61, 592-607.
- Janssen, M. 1995. *A Comparison Between Lorentz's Ether Theory and Special Relativity in the Light of the Experiments of Trouton and Noble*, PhD Dissertation, University of Pittsburgh.
- Janssen, M. 2002a. Reconsidering a Scientific Revolution: The Case of Lorentz versus Einstein, *Physics in Perspective*, 4, 421-446.
- Janssen, M. 2002b. COI Stories: Explanations and Evidence in the History of Science, *Perspectives on Science*, 10, 457-522.
- Janssen, M. 2003. The Trouton Experiment, $E = mc^2$, and a Slice of Minkowski Space-time, in A. Ashtekar, R.S. Cohen, D. Howard, J. Renn, S. Sarkar and A. Shimony (eds), 2003, *Revisiting the Foundations of Relativistic Physics: Festschrift in honor of John Stachel*, Dordrecht: Kluwer, 27-54.
- Janssen, M. 2009. Drawing the Line between Kinematics and Dynamics in Special Relativity, *Studies in History and Philosophy of Modern Physics*, 40, 26-52.
- Janssen, M. and Stachel, J. 2004. The Optics and Electrodynamics of Moving Bodies, preprint, Max Planck Institute for the History of Science.
- Kukla, A. 1993. Laudan, Leplin, and Underdetermination, *Analysis*, 53, 1-7.
- Kukla, A. 1996. Does Every Theory Have Empirically Equivalent Rivals?, *Erkenntnis*, 44, 137-166.
- Kukla, A. 2001. Theoreticity, Underdetermination, and the Disregard for Bizarre Scientific Hypotheses, *Philosophy of Science*, 68, 21-35.
- Laudan, L. 1990. Demystifying Underdetermination, in C. W. Savage (ed.) *Scientific Theories XIV*, Minneapolis: University of Minnesota Press, 267-297.
- Laudan, L. and Leplin, J. 1991. Empirical Equivalence and Underdetermination, *The Journal of Philosophy*, 88, 449-472.
- Laudan, L. and Leplin, J. 1993. Determination Underdeterred: reply to Kukla, *Analysis*, 53, 8-16.
- Leplin, J. 1997a. *A Novel Defense of Scientific Realism*, Oxford: Oxford University Press.
- Leplin, J. 1997b. The Underdetermination of Total Theories, *Erkenntnis*, 47, 203-215.
- Magnus, P. D. 2003. Underdetermination and the Problem of Identical Rivals, *Philosophy of Science*, 70, 1256-1264.
- McAllister, J. 1989. Truth and Beauty in Scientific Reason, *Synthese*, 78, 25-51.
- McCormmach, R. 1970. H.A. Lorentz and the Electromagnetic View of Nature, *Isis*, 61, 459-497.
- Miller, A. I. .1973. A Study of Henri Poincare's *Sur la dynamique de l'électron*, *Archive for History of Exact Sciences*, 10, 207-328.
- Miller, A. I. 1981. *Albert Einstein's Special Theory of Relativity: emergence (1905) and early interpretation (1905-1911)*, New York: Springer.

Norton, J. 2008. Must Evidence Underdetermine Theory?, in M. Carrier, D. Howard, and J. Kourany (eds), 2008, *The Challenge of the Social and the Pressure of Practice: science and values revisited*, Pittsburgh: University of Pittsburgh Press, 17-44.

Okasha, S. 1997. Laudan and Leplin on Empirical Equivalence, *British Journal for the Philosophy of Science*, 48, 251-256.

Psillos, S. 1999. *Scientific Realism: how science tracks truth*, London: Routledge.

Suppe, F. 1974. The Search for Philosophical Understanding of Scientific Theories, in F. Suppe (ed.), 1974, *The Structure of Scientific Theories*, Urbana: University of Illinois Press, 3-241.

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