UTRECHT UNIVERSITY
INSTITUTE FOR HISTORY AND FOUNDATIONS OF SCIENCE

Empirical Equivalence and Underdetermination of Theory Choice

A Philosophical Appraisal and Two Case-Studies

Dissertation

Pablo Acuña
Supervisor: Dennis Dieks

June 2014

The illustration in the cover, *Alice and Tweedledum and Tweedledee*, is taken from the original drawings by John Tenniel in Lewis Carroll's *Through the Looking Glass.*

Empirical Equivalence and Underdetermination of Theory choice:
a philosophical appraisal and two case-studies

Empirische Gelijkwaardigheid en Onderdeterminatie Theorie Keuze:
een filosofische beoordeling en twee *case-studies*
(met een samenvatting in het Nederlands)

Proefschrift ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 27 august 2014 des middags te 12.45 uur

door

Pablo Thomas Acuña Luongo

Geboren op 6 mei 1980 te Santiago, Chili

Promotor: Prof. dr. D. Dieks

This thesis was accomplished with financial support from CONICYT (Comisión Nacional de Investigación en Ciencia y Tecnología, Chile).

INTRODUCTION

The problem of empirical equivalence and underdetermination of theory choice is one of the most debated issues in the philosophy of science. This is hardly surprising, for the very rationality of theory choice is at risk if the problem is not solvable: if the problem proves intractable, there would be no way to provide an account of rational and objective theory acceptance that is grounded on empirical evidence. Furthermore, since the problem is usually understood in a universal way—in the sense that for any conceivable theory there is a predictively equivalent rival that leads to underdetermination—the threat to the objectivity ad rationality of theory choice that it is supposed to entail applies to science as a whole.

In spite of all the attention and pages that have been devoted to this problem, it is awkward that most of that attention and those pages deal with the subject only from an abstract and conceptual point of view. None of the solutions provided has been directly tested in a case of 'real-life' science. This is rather curious, since there are well-known cases of predictive equivalence in the history of science that could be used as case-studies for the problem at hand. In this thesis I will deal with the case of Einstein's special relativity *vs*. Lorentz's ether theory, and with the case of 'standard' quantum mechanics vs. David Bohm's quantum theory. Even though there is plenty of excellent historical and conceptual works available regarding both examples, those works have not been used in order to evaluate how the different philosophical positions regarding the problem of empirical equivalence and underdetermination fit in real science. It is true, though, that philosophers of physics have proposed analyses of these two cases in which they consider and defend some reasons that may be rationally invoked to make a decision. However, none of them has approached those cases as a specific instances of the general problem of empirical equivalence and underdetermination of theory choice. Thus, there is a twofold gap in the philosophical treatment of the problem. First, philosophers of science have not carefully tested their views in case-studies that can be found in the history of science. Second, when philosophers of physics discuss the two examples that will be considered here, they do not extract general conclusions or lessons about the general problem.

In this work I attempt to fill this twofold gap. I offer a description and evaluation of the problem at issue, and I carry out two tests of this evaluative description by means of two case-studies: Einstein's special relativity and Lorentz's ether theory, and standard quantum mechanics vs. Bohm's theory. The evaluative description of the problem of empirical equivalence and underdetermination here defended also yields contributions to important issues in the philosophy of physics, for it allows accurate assessments of the examples considered. That is, the arguments here presented will permit us to understand exactly what is the justification for a choice favoring Einstein's theory in the first example, and will provide us with an accurate appraisal of the situation given in the contend between Bohm's theory and standard quantum mechanics.

In chapter one I undertake a philosophical assessment in which I provide a clear and precise exposition of the problem, an examination of the most important possible solutions that have been proposed in the relevant literature, and an evaluation of these solutions in order to determine which of them work and which do not. The solution I defend is based on an argument introduced by Larry Laudan and Jarret Leplin. Even though I think that their line of reasoning is essentially correct, I introduce important provisos and qualifications which clarify the real scope and nature of the solution that these authors offer. Laudan and Leplin state that the problem at issue gets *refuted*, but I argue that this is an evaluation that goes off the mark. I claim that their argument, rightly understood, shows that the regular practice of science can provide results that *might* break the empirical equivalence between the theories and/or the underdetermination of the choice to be made. However, the problem remains as a possible scenario, and its solution—as provided by the development of science—is not guaranteed from the outset.

In the second chapter I offer a historical and conceptual outline of the Einstein *vs.* Lorentz case. This outline includes an exposition of the main tenets of both theories, and a comparison between them in order to show that they are different and rivals, but at the same time predictively equivalent. After the presentation and assessment of both theories, I explain and evaluate the usual arguments for a choice favoring special relativity that have been provided, and I introduce a couple of new arguments that, to my knowledge, have not been considered. I argue that even though all of the arguments that have been addressed so far in the available literature are of a non-empirical nature, the case of Einstein vs. Lorentz can be decided in terms of empirical evidence, in spite of the predictive equivalence that stands between the theories.

In chapter three I undertake an analysis of the case of Bohm's theory vs. standard quantum mechanics. I proceed in a similar way as in chapter two. I provide a careful examination of both theories, and I explain in detail the features that make them predictively-equivalent-but-rivals. Then I consider and evaluate the reasons that are usually proposed in the relevant literature in order to support both theories. Based on this analysis, I claim that, unlike the case of Einstein vs. Lorentz, in this case there is no way to determine an evidentially based decision. The case of Bohm vs. quantum mechanics constitutes a persistent instance of empirical equivalence and underdetermination.

In a brief final chapter I present the results that the analyses and arguments provided in the previous chapters yield. I propose an evaluative description of the problem at issue given by five statements. Then I show how the results of the two case-studies addressed clearly illustrate that these five statements correctly grasp the nature and scope of our problem. Besides, I show that, in turn, these five statements allow a clear and accurate evaluation of both the case of Einstein vs. Lorentz and of the case of Bohm vs. quantum mechanics.

The research and writing of this thesis was financed by a doctoral scholarship provided by Comisión Nacional de Investigación en Ciencia y Tecnología—Chile (CONICYT). During the four-year period (2010-2014) that I was funded by CONICYT I completed both the MSc and PhD programs in History and Philosophy of Science at Utrecht University. A revised version of my master thesis—which was awarded with the Utrecht University Best Graduate Thesis Prize 2012-2013, and which came in second place in the international contest Hanneke Janssen Memorial Prize 2013 for theses in philosophy of physics—is included as a part of this dissertation. The results of chapters 1 and 2 have appeared in published form in (Acuña & Dieks 2014), (Acuña 2014a) and (Acuña 2014b), and I expect to submit the results of chapter 3 for publication in the near future.

Acknowledgments

I am deeply grateful to my supervisor, Dennis Dieks, for all his support, help and fruitful criticisms along the writing of this thesis. To be guided in this big learning adventure by such a remarkable scholar was a great honor. Moreover, that in all our meetings and discussions he always received me with a warm and friendly smile is something that I cannot thank enough. I learned not only a good deal of philosophy of science and philosophy of physics from Dennis, but also that generosity and kindness constitute the best possible context for scholarship.

I owe much also to Roberto Torretti. His work has been of great inspiration ever since my early days in the world of philosophy, and his generous suggestions and criticisms concerning some subjects in this work were certainly crucial to improve it. Our friendship and constant intellectual dialogue are two 'results' of this thesis that I deeply treasure.

I am also very grateful to my good friends and mates in the HPS program. All the nice conversations (about physics, epistemology, football, life in Holland, and even about the weather) with Vincent Schoutsen, Chao Kang Tai, Giulia Paparo, Nick Evans, Steve Shapiro and Tom Sterkenburg contributed to the completion of this work.

I would like to express my special gratitude to Alfred van Herk and his wife Marjia. Their beautiful friendship helped me not to feel always so far from home. In conversations with Alfred I always return to the conviction that physics and philosophy are valuable because of the pure will and awe to understand.

Thanks a lot also to my "Hungarian block-family": János, Márta and Máté in Budapest. Their friendship and generous hospitality is simply invaluable for me. Köszönöm szépen!

Quiero agradecer también a todo mi *clan* en Chile. A mi madre por siempre estimular mis ganas de saber todo lo que hay por saber. A mi padre por la certeza de que si Colo-Colo gana, todo está bien. A Nico, Luis, Iván, Elías y Sebastián por todas las risas por Skype, y por siempre recordarme que al fin y al cabo "no somos nada". A mi hermana Mariana por seguirme en el camino de "cachar el mote". Sin el amor de mis padres-hermanos-primos-amigos nada de esto sería posible ni valioso.

Finally, I want to thank Kata. However, I don't really know how to do it. There are just no words that could express my gratitude towards her for being a part of my life during all these years.

For Kata, for all the joy, for all the peace

A cloud of eiderdown
Draws around me
Softening a sound.
Sleepy time, and I lie,
With my love by my side,
And she's breathing low.
And the candle dies...

When night comes down
You lock the door.
The book falls to the floor.
As darkness falls
The waves roll by,
The seasons change
The wind is wry.

Now wakes the owl
Now sleeps the swan
Behold the dream
The dream is gone.
Green fields are calling
It's falling, in a golden door.

And deep beneath the ground,
The early morning sounds
And I go down.
Sleepy time, and I lie,
With my love by my side,
And she's breathing low.

And I rise, like a bird,
In the haze, when the first rays
Touch the sky.
And the night wings die...

Roger Waters

TABLE OF CONTENTS

CHAPTER 1
A PHILOSOPHICAL APPRAISAL

I will begin with a philosophical treatment of the problem of empirical equivalence and underdetermination of theory choice. The general goal of this chapter is to provide an accurate description and evaluation of our problem. To achieve that, I will undertake an examination of its presuppositions, precise meaning, and consequences. This chapter is divided in five sections. In the first one I provide a precise formulation of the problem in terms of an argument constituted by two premises and a conclusion. I also undertake an analysis of what is precisely at stake in the problem, and of what are the presuppositions underlying both the premises. In the second section I argue that two possible solutions that have been provided do not work: the view that predictively equivalent theories are simply two different formulations of the same theory, and Jarret Leplin's attack on the logical soundness of the argument which constitutes the problem. In the third section I argue that although recourse to non-empirical or theoretical features (virtues or flaws) are useful and fruitful in the sense of providing rational reasons in order to mark a preference between the theories involved, such reasons are not enough in order to ground a fully objective and uniquely determined choice—non-empirical features can provide only a partial solution to the problem. In the fourth section I offer a detailed exposition of some interesting and compelling arguments that Larry Laudan and Jarret Leplin proposed in 1991 with respect to the problem of empirical equivalence and underdetermination, and I add an argument by Richard Boyd as a useful complement for that solution. In the fifth section I deal with three remaining challenges given by some artificially generated examples of empirical equivalence which are usually discussed also as examples of underdetermination. I argue that none of these remaining challenges implies difficulties beyond the scope of the solution provided in section three. The sixth and final section of this chapter consists on a re-evaluation of the solution provided by Laudan & Leplin (and Boyd). I argue that these authors assign a scope to their position which goes too far. They claim that the thesis of underdetermination as a consequence of empirical equivalence has been *refuted*, but I show that even though their argument is essentially correct—in the sense that science has the tools to dissolve it—empirical equivalence as a source of underdetermination of theory choice remains as a possible scenario. Thus, this reassessment I argue for gives us a correct description of the nature and status of the problem at stake[1].

1.1 DEFINING THE PROBLEM

The problem of underdetermination (UD), as a consequence of empirical equivalence (EE) between two competing theories, is at face value, a very deep and difficult one. If two different theories entail exactly the same observational consequences, they are also equivalent from a confirmational point of view—provided that the standard hypothetic-deductive model of evidential support is adopted. That is, everything which confirms or disconfirms one of the theories also confirms or disconfirms the other. Therefore, there are no possible evidential resources to accept one of them and to reject the other. This conclusion seems to threat the objectivity and even the very rationality of the decision to be made between the competing theories. Moreover, the argument is commonly presented as establishing that *any* theory has an actual or virtual empirically equivalent competitor, so in this case *any* theory whatsoever is underdetermined, and, consequently, theory choice as a general feature of scientific practice could be seen as ungrounded.

[1] The main contents and results of this chapter can be found in published form in (Acuña & Dieks 2014) and (Acuña 2014a).

Laudan and Leplin present the problem of EE and UD in the following way: 'By the 1920s, it was widely supposed that a perfectly general proof was available for the thesis that there are always empirically equivalent rivals to any successful theory. Secondly, by the 1940s and 1950s, it was thought that—in large part because of empirical equivalence—theory choice was radically underdetermined by any conceivable evidence' (Laudan and Leplin 1991, 449). As this formulation shows, the problem arises from a connection between EE and UD of theory choice in terms of empirical evidence. This UD, in turn, is supposed to imply deeply problematic consequences—the most notorious one being that the status of the ultimate basis for the acceptance of theories, empirical evidence, gets epistemically eroded:

> The idea that theories can be empirically equivalent, that in fact there are indefinitely many equivalent alternatives to any theory, has wreaked havoc throughout twentieth century philosophy. It motivates many forms of relativism, both ontological and epistemological, by supplying apparently irremediable pluralisms of belief and practice. It animates epistemic skepticism by apparently underwriting the thesis of underdetermination. *In general, the supposed ability to supply an empirically equivalent rival to any theory, however well supported or tested, has been assumed sufficient to undermine our confidence in that theory and to reduce our preference for it to a status epistemically weaker than warranted assent*. (ibid., 450, my emphasis)

A simple schematic presentation of the problem at issue can be given in terms of an explicit 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 rationale for this statement comes from two main sources. Some authors claim that given any theory that entails certain observational consequences, there are algorithmic procedures which produce another theory with exactly the same observational consequences. On the other hand, what is commonly known as *the Quine-Duhem thesis* is usually taken as providing support for this premise. Duhem has shown that the logic of evidence is holistic: a theoretical hypothesis can entail empirical consequences only with the help of auxiliary hypotheses, so that, *logically speaking*, any evidence could be accommodated by any theory given the necessary arrangements on the auxiliary assumptions. 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'. This last conclusion leads to the effectiveness of the first premise: theory T, compounded by hypothesis H and auxiliary assumptions A, entails e; and another theory T' compounded of hypothesis H' and auxiliary assumptions A' also entails e, so that T and T' are empirically equivalent.

The second premise of the argument says that *only observational statements that can be logically derived from a theory can count as empirical evidence to support it*[2]. This statement is supported by the traditional hypothetic-deductive model of evidential confirmation. Roughly and briefly speaking, this model, introduced by the logical positivists, asserts that an observational report can count as evidence for a certain hypothesis if and only if a sentence expressing that report is entailed by the hypothesis at issue (along with auxiliary assumptions, of course). In spite of the many problems that the logical positivist program had to face, this model of confirmation—surely because of its simplicity and *prima facie* obviousness—has remained a milestone in the philosophy of science.

These two premises entail the problematic conclusion. If there is an EE rival to any theory, and if a theory gets empirically 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 implies 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 threat. Given this diagnosis of the problem, it is quite clear that any attempt to solve it will have to be put in terms of criticism of one

[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.

of the premises, or of both. Obviously, such criticism will be directed to a critical assessment of their supporting theses—the effectiveness of algorithms to produce empirically equivalent theories, and the holistic nature of confirmation.

EE and UD can be understood as a problem in several (though connected) contexts. For example, the logical possibility of predictive equivalence[3] is often taken as a challenge for scientific realists—that is, as an argument against the idea that (converging) truth is a goal that science can reasonably pursue. Therefore, it is important to underscore that I will only tackle the problem of EE and UD in the context of the challenge that it poses with respect to the *rationality* and *objectivity* of theory choice. As Laudan and Leplin state,

> A number of deep epistemic implications, roughly collectable under the notion of "underdetermination", have been alleged for empirical equivalence. For instance, it is typical of recent empiricism to hold that evidence bearing on a theory, however broad and supportive, is impotent to single out that theory for acceptance, because of the availability or possibility of equally supported rivals. Instrumentalists argue that the existence of theoretically noncommittal equivalents for theories positing unobservable entities establishes the epistemic impropriety of deep-structure theorizing, and with it the failure of scientific realism. Some pragmatist infer that only nonepistemic dimensions of appraisal are applicable to theories, and that, accordingly, theory endorsement is not exclusive, nor, necessarily, preferential (1991, 459-60).

Here I will deal only with the first and third dimensions of the problem that Laudan and Leplin mention in this quote, not with the second one[4].

1.2 DISCARDING SOME WAYS OUT

In this section I will briefly review an attempted way to avoid the problem of EE and UD, namely, the thesis that if two theories are predictively equivalent means that the theories are nothing but two different formulations of the same theory, and an argument proposed by Jarret Leplin that intends to show that the EE thesis itself is inconsistent with UD—if the EE and the UD theses cannot be both true, then there would be no problem. I argue that both these approaches are unsuccessful.

1.2.1 EE as different formulations of the same theory

If the empirical equivalence between two theories can be shown not to be a case of a competition, but a case in which the very same theory is presented in two different ways, it is obvious that the UD does not even come up. The "choice" to be done can be grounded on pragmatic considerations such as simplicity or the like, since no theory is really being rejected. Choosing one or the other formulation is not an epistemic issue.

This was the position that logical positivists held. Its rationale comes from the verificationist criterion of meaning. The meaning of a term, of a sentence, or of a theory, is nothing but the method to verify it. Such a method, in the case of scientific theories, is given by their observational consequences. Therefore, if two theories have exactly the same empirical consequences, they have exactly the same meaning. From this semantic point of view, two empirically equivalent theories are just synonyms, so that the choice between them has nothing to do with evidential or epistemic conditions.

[3] I use the expressions 'empirical equivalence' and 'predictive equivalence' in a fully synonym way.
[4] For a general assessment of the problem of EE and UD as a problem for the realist, see (Psillos 1999).

Insofar as this view depends on the verificationist criterion of meaning, if such criterion is shown to be untenable, the related view with respect to EE also falls. As it is widely known, this is exactly what happened during the last century. However, if the view that empirically equivalent theories are the same theory can be supported by arguments which do not depend in the logical positivist semantic criterion, this way out can be reintroduced. Some attempts along this line of thinking have been done, especially within the semantic conception of scientific theories[5]. For example, John Norton (2008) argues that whenever we face a case in which empirical equivalence can be asserted between two theories we might suspect that they are nothing but variant formulations. His argument states that in order to determine the equivalence a set of conditions must be met: 1) we must have a tractable description of their observational consequences; 2) each of the theories must make essential use of its own theoretical language in the entailment of their observational consequences;

> 3. If we are able to demonstrate observational equivalence of the two theories, the theoretical structures of the two theories are most likely very similar. While it is possible that they are radically different, if that were the case, we would most likely be unable to demonstrate the observational equivalence of the two theories. For the theoretical structures are what systematizes the two sets of observational consequences, and a tractable demonstration of observational equivalence must proceed by showing some sort of equivalence in these systematizing structures.
> 4. The two sets of theoretical structures may be inconvertible without loss; or they may not be. In the latter case, there would be additional structures present in one theory but not in the other. However, any such additional structure will be unnecessary for the recovery of the observational consequences. That follows since the additional structure has no correlate in the other theory, yet the other theory has identical observational consequences. Thus any additional structures will be strong candidates for being superfluous, unphysical structures. (Norton 2008, 34-5)

This list of requirements, Norton argues, implies that if two theories are so demonstrated to be empirically equivalent, then there must be some structural similarity, so that interconveribility must be possible. Cases in which no structure is lost, like matrix mechanics and wave mechanics in quantum theory are clear and straightforward examples of theory identity[6]. Now, when the interconversion yields some structure loss, the extra structure in the less economic theory might be considered as representing nothing physical, for such an extra structure it is not needed for the entailment of observable predictions in the more economic theory. Thus, we could then conclude that the theories are not really rivals, but formulations of the same theory, but with one of the formulations carrying some extra and unnecessary structural baggage. However, Norton himself points out that the debate about the superfluous character of the remaining structure is a very complex one. In any case, his point consists in showing that, at least in principle, EE strongly *suggests* theoretical identity. This does not mean that EE *means* theoretical identity right away. Anyways, Norton's argument shows that examples of EE must be scrutinized in order to establish a real rivalry between the theories. If we can straightforwardly show that the extra structure is superfluous and that the excision of such a structure from the theory leaves us with two different formulations of the same theory, then there is no problem of UD.[7]

---

[5] The semantic or structuralist conception, unlike logical positivists, states that theories are extra-linguistic entities. They are constituted by a set-theoretical predicate –the structure– that is satisfied by a model of the predicate in the real world. This approach is sometimes called semantic insofar as it construes theories as what their formulations refer to when their formulations are given a formal semantic interpretation. However, structuralists argue that the *mapping* relation between the model in the real world and the structure, unlike logical positivist's concept of *correspondence rules*, is not a part of the theory. Another important feature of this approach is that it accepts the thesis of *theory ladenness of observation*: all terms are theory laden, but if they count as theoretical or observational is a context-dependent issue. The same concept can play an observational or a theoretical role depending on what theory it is a part of.

[6] This is, of course, the 'traditional' view on the matter. F. Muller (1997) challenges this view and states that, when originally formulated, matrix mechanics and wave mechanics were not empirically equivalent—and therefore not the same theory. According to Muller, they became empirically and ontologically identical with the work of von Neumann in 1932.

[7] (Norton 2008, 33-40). For other considerations of theoretical identity between empirically equivalent theories from a semantic-structural point of view, see also (French 2011b), (Yalcin 2001), and (Mormann 1995).

Norton's view seems to be much more promising and cogent then the logical positivists' one. However, it depends on whether the semantic-structural conception of theories is adequate enough—an assessment of this issue goes quite beyond the scope of this work. Moreover, Norton's stance is different from the logical positivistic view in an important way. For the latter, the semantic criterion *necessarily* implies that empirically equivalent theories are nothing but alternative formulations of the same theory. In Norton's case, the equivalence only *suggests* that it could be a case of alternative formulations. That the suggestion proves correct in some cases does not mean that we have a solution of the general problem of EE and UD, for there may be other cases in which the suggestion turns out wrong. Actually, and in spite of Norton's argument, it might turn out that the most relevant cases of EE are indeed cases of two different and rival theories.

More generally, since the first premise of the argument states that for any theory T there is an empirically equivalent rival T', the issue of whether the two theories are really rivals and different becomes relevant. Therefore, a general criterion for identity or non-identity between theories would be most useful. However, it is not currently available. But that does not mean that the problem of EE and UD cannot be posed or that it cannot be solved. The current situation is that whether two EE theories are identical or not must be approached case by case, and there is no reason to presuppose, as the logical positivists did, that all of the cases will be instances of theoretical identity. As P. D. Magnus affirms,

> 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)

1.2.2 Leplin's inconsistency argument

Jarret Leplin proposes a solution which consists on showing that EE cannot lead to UD insofar as the EE thesis and the UD thesis are mutually inconsistent. That is, for logical reasons, there is no problem at all. His proof of the inconsistency between the theses at issue relies on the fact that in order to determine if two theories are EE, it is needed that the auxiliary assumptions that permit the entailment of the observational consequences may be specified and well established. However, if UD is the case, then the auxiliary hypotheses required are also underdetermined, and the set of the possible and available auxiliaries gets unclear and undetermined. Therefore, there are no clear and well established auxiliaries to perform the entailment of observational consequences from the theories we want to asses. Thus, what are the observational consequences of a theory, if UD is the case, is impossible to determine. If UD is true, if EE cannot be decided; and if EE is true, UD cannot be true:

> *The truth of UD 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 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. (Leplin 1997a, 154-5; see also Leplin 1997b)

If the available auxiliary assumptions are not firmly established and justified by a certain criterion, they cannot be safely used, Leplin argues, in order to derive observational consequences from theoretical hypotheses. If UD were to affect auxiliary hypotheses, they could never be better supported than the theoretical ones that are to be evaluated, leading to a radical holism in which it would be completely

impossible to asses any theoretical hypothesis (more or less) directly; a *Duhemian nightmare* would be the case:

> 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. (ibid, 155)

If this were the case, then auxiliaries would be as underdetermined as theoretical hypotheses. Therefore, there would be no epistemic standard to establish some hypotheses as justified assumptions to be used to derive observational consequences from theoretical hypotheses, "*there will be no fact of the matter as to what the empirical consequences of any theory are*" (ibid); and consequently, it cannot be stated whether two theories are empirically equivalent or not.

What Leplin believes to have shown is that the first premise (the fact that for any theory there is an empirically equivalent rival) is inconsistent with the conclusion of UD, *only insofar as the former is supported by the Duhemian holist thesis*. If UD is the case in this sense, then it is impossible to determine the set of observational consequences of any theory. However, we may recall that the support for the first premise comes also from algorithms that are supposed to generate an EE rival given any theory, and Leplin concedes that if we grant that the class of observational consequences is already specified for any theory, and we introduce a certain method to produce an observationally equivalent theory, then the UD problem raises again. In other words, given a certain theory T whose observational consequences O are established, if there exists an algorithm to be applied to the theoretical hypotheses of T such that a new theory T' results which also entails O; then the problem of UD comes up anyway:

> The only general strategy for upholding EE that is capable of coping with this difficulty is algorithmic. This strategy exists in many versions. Usually, the empirical-consequence class O of an arbitrary theory T is supposed to be given, and the algorithm operates on T to produce another theory T' whose consequence class is also O. (ibid, 158)

That is, Leplin's criticism shows that the EE premise and the UD conclusion are inconsistent only if we understand EE as the result of a Quine-Duhem holism. If we understand it as the result of the operations of certain algorithms which presuppose the class of observational statements, the inconsistency is not the case.[8]

Leplin's thesis that UD implies that there is no way to firmly establish the degree of empirical confirmation possessed by a hypothesis that is to be used as an auxiliary in the derivation of empirical consequences from other hypotheses can be understood in two ways. First, we can take UD as a consequence of the Quine-Duhem holism thesis. As it is widely known, Quine and Duhem showed that a theoretical hypothesis cannot be disconfirmed in isolation, what is disconfirmed via *modus tollens* in the light of negative evidence is a conjunction of the hypothesis and auxiliaries, and we can always blame the latter rather than the former for the entailment of the wrong prediction. A modest interpretation of this holist thesis is simply that, just like empirical confirmation, the empirical disconfirmation of scientific hypotheses is fallible. However, and mainly because of Quine's famous paper, the thesis is many times understood as implying a radical form of UD, in the sense that given a certain hypothesis H, *no observation whatsoever is capable of disconfirming a rival hypothesis* H'. Now, since this form of UD means that we cannot

[8] Leplin's proviso to his own argument is somewhat awkward. I understand that he is presupposing that the class of observational consequences O of the theory T is already specified, so that the algorithmically tailored theory T' and its correspondent O' does not require the introduction of new or further auxiliary hypotheses. However, the entailment of O from T does require auxiliary hypotheses!

establish that H is better confirmed than H', none of them can be safely used as an auxiliary hypothesis in the derivation of observable consequences from other theoretical hypotheses—and since this form of UD is supposed to hold for all scientific hypotheses H, then it would be impossible to determine the class of observable consequences of other theories.

But this interpretation of the Quine-Duhem holism thesis is unfounded (see Laudan 1990). That there exists the *logical* possibility of blaming auxiliaries rather than the tested hypothesis for a wrong prediction does not mean that the dynamics of empirical testing is impotent in rejecting hypotheses. As Duhem himself stressed out in his seminal book, there is a *good sense* that the rational scientist possesses that indicates him when a hypothesis has been empirically disproven, in spite of the ever present logical possibility of blaming the auxiliaries for the wrong prediction. That is, since empirical confirmation is not a matter of mere logic, then the holist thesis does not imply that hypotheses can never be refuted. Now, if we interpret the Quine-Duhem holism thesis in this way, and not as entailing the radical form of UD that denies the possibility of firmly establishing the degree of confirmation possessed by hypothesis to be used as auxiliaries, then UD *as a consequence of EE* is actually a real menace. If science has methods to establish that some hypotheses are well-confirmed (and better confirmed than other rival hypotheses), than the process of determining the class observational consequences of a theory by using such well-confirmed hypothesis is justified. Now, if two theories have the same class of observational consequences, and if the derivation of observational consequences is the only way to empirically confirm a theory, the choice between the theories is underdetermined by the empirical evidence.

On the other hand, Leplin's argument can be also understood as stating that the UD that as resulting from EE is what disallows the possibility of firmly establishing the degree of empirical confirmation of hypotheses to be used as auxiliaries. Understood in this way, the argument certainly puts logical pressure on the *universality* which is usually assigned to the problem. If the thesis of EE holds for all possible hypotheses and theories, then the process of firmly establishing the degree of confirmation of hypotheses that are to be used as auxiliaries gets jeopardized. Thus, we may take that EE and UD are inconsistent theses only if the former is taken as a *universal* thesis. However, this is not enough to consider Leplin's argument as a solution of the EE and UD of theory choice, for if we take for granted that the confirmation process is justified (as it is), and if the EE thesis is understood as holding for *some* hypotheses and theories, then in the cases where EE occurs, there may be no way to rationally or objectively determine a choice between the theories involved. That is, EE and UD form a set of inconsistent theses only if the EE thesis is taken in a universal sense. However, if we take it as a condition that can affect certain pairs of theories, then the problem is still there. Leplin interprets his argument as a solution of the problem of EE and UD surely because he thinks that the problem *is a problem for the realist*, and this is so when the problem takes the universal form. But, as I mentioned above, here I am considering it as a problem for the rationality and objectivity of theory choice, this form of the problem can remain even if the scope of the first premise gets restricted. Actually, as we will see below, the universality of the EE thesis can be indeed refuted, but there are no reasons to think that EE and UD cannot occur at all[9].

1.3 A PARTIAL SOLUTION

A straightforward way out of the problem would be to weaken the second premise in the problematic argument by having recourse to non-empirical features of the EE theories involved. If one of the theories is simpler or proves to have more explanatory power than its rival, for example, one has a reason to

[9] For different criticisms of Leplin's argument see (Douven 2000), and (Sarkar 2000).

prefer this theory after all. However, both simplicity and explanatory power are features that are controversial and hard to assess unambiguously. In the case of simplicity, the very definition of the concept is far from clear—it looks like a feature of theories 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, objective terms.

Something similar holds in the case of explanatory power. What a scientific explanation is, or must be, is an open philosophical question. There are well known arguments for the position that explanation is an essentially context dependent concept. Bas van Fraassen (1980, 134–57), 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 and on 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 theory[10].

These remarks already illustrate that non-empirical features will often be pragmatic and context-dependent, 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. However, the objectivity of theory choice cannot be rescued in this way. Even if it were possible to state completely unambiguously that a theory possesses more explanatory power than its rival in a certain context, there might 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 both at the same time[11]. The limitation of this kind of pragmatic solution is therefore that, even though pragmatic non-empirical features may 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, fortunately, 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 after all.

These remarks should not be taken to imply that non-empirical virtues and empirical evidence are two completely unrelated concepts—I am not arguing that non-empirical virtues must necessarily be merely pragmatic, subjective and context dependent, and totally unconnected to empirical evidence. It is possible to conceive of these non-empirical virtues as ultimately grounded in an evidential basis—so that the-

[10] See also (De Regt & 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.

[11] 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. I 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.

oretical or aesthetic features would be a better label than non-empirical. James McAllister (1989) has offered an interesting account of such features along this line. McAllister argues that 'indicators of beauty' that can work as relevant criteria for theory choice are based on meta-inductions on the aesthetic aspects of empirically successful theories of the past. That is, the criteria that define theoretical virtues may be indirectly based on the empirical success of theories:

> 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).

This account of aesthetic virtues might look as a way to justify them as a full solution to our problem—they might count as objective evidence after all. However, on closer inspection it appears that this possible connection between empirical evidence and theoretical-aesthetic features is not enough for considering the latter as a source for objective decisions in cases of EE. First, if so-called theoretical virtues are inductively linked to the empirical success of past theories, they are ipso facto epistemically subordinated to empirical evidence when it comes to theory choice—if (dis)confirming evidence goes against the prevailing canons of theoretical beauty it is clear that the former will be more important[12].

Second, I agree with McAllister when he emphasizes that 'indicators of truth', the evidential criteria which are decisive for justified theory choice are determined by the basic goals of science. Thus, indicators of truth are of a different character and are much more stable than indicators of beauty[13]. Actual, present empirical success remains the ultimate criterion for theory choice even after scientific revolutions, whereas the canons of theoretical beauty are governed by past performance and are intrinsically related to a specific state of science and to individual credos of scientists. Therefore, 'indicators of beauty' are not likely to provide objective and uniquely determined choices 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)[14].

In other words, that theoretical virtues may derive from meta-induction on empirically successful theories does not imply that there will be a uniquely defined canon of beauty for all scientists. Moreover,

---

[12] '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 anteceding 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).

[13] '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' (ibid., 38). In order to retain neutrality regarding the realism-antirealism schism, we can replace 'indicators of truth' for 'indicators of empirical success'.

[14] Furthermore, in times of scientific crisis there is no unique canon of beauty (if there ever is). A good example is given by the four-dimensional formulation of special relativity by Hermann Minkowski. Some scientists (such as Sommerfeld and Laue) considered the chrono-geometric formulation as expressing aesthetic virtues (based on simplicity, mainly), whereas others (e.g., P. Frank, at least for some time) considered it as expressing a non-empirical flaw (given the loss of intuitive visualizability involved). See (Illy 1981) and (Walter 2010).

within a single canon of beauty different theoretical features will usually play a role, and it may happen that in a pair of EE theories one of them scores better than its rival according to one feature, but that the situation is inverted according to a different feature. There is no clear and objective ranking of importance to classify the different theoretical features within one single canon. Scientists' and philosophers' ranking of aesthetic virtues connects to their individual epistemological and metaphysical commitments, with the consequence that canons of theoretical virtues are not fully objective—even if they are meta-inductively grounded.

Apart from the objectivity problem, another obstacle for meta-inductively supported theoretical virtues as a full solution of the problem of EE and UD comes from the nature of the inductive argument itself. As McAllister stresses, the meta-induction at issue is strictly Humean, in the sense that it is based upon mere past correlation, without any guarantee of empirical success now:

> The present account of indicators of beauty is intended to resemble the Humean explanation of the origin of notions of cause: just as Hume believed the inductive apprehension of causal links to be unsupportable by nomological data but a nonetheless ineluctable product of a driven mind, aesthetic canons in science boast no systematic relation to truth but spring from the psychological concerns of scientists. Neither Hume's account nor this concludes that notions thus formed are of no value: as causal links are a convenience of the Humean life, so indicators of beauty may here aid theory-construction and choice. It is, however, important to remember the contingent nature of concepts generated by Humean inductions and to avoid attributing to them any necessity. (ibid, 40–1)

The Humean nature of the induction implies that even though 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 theoretical virtues, it is still possible, as we will see below, that the future development of science may break the EE and/or the UD favoring—in terms of empirical evidence—the theory we rejected; and this is so even if the rejected theory is inferior in all contexts and in all aspects with respect to theoretical 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 theoretical 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. This risk can be avoided only if an epistemological inherent connection between theoretical virtues and empirical success were provided. However, if possible at all, an argument like this would necessary rely on very strong—and quite likely very doubtful—metaphysical presumptions.

1.4 LAUDAN & LEPLIN'S SOLUTION

In a very influential paper Jarret Leplin and Larry Laudan undertook an attempt to provide a solution to the problem of EE and UD which I consider essentially correct—though, as I will argue, its meaning and scope must be reassessed. Their argument consists in a critical assessment of both the premises that lead to the UD conclusion. I will now provide a careful exposition of each of the steps of their argument, along with the main criticisms it has received.

1.4.1 The first premise

The first step in Laudan & Leplin's argument consists on a critical revision of the first premise of the problematic argument stated in section 1.1. Since their line of reasoning is considers several stages, I will divide its analysis in subsections.

a) 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 of these theses 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 & 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 development[15]. 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 concerns 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' (ibid, 454).

The upshot is that if two theories completely coincide in their predictions now, it does not follow that they are essentially 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_t) and (T_2, A_t)—where A stands for the auxiliary assumptions—can be considered as 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

[15] The authors 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).

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.

b) Algorithms and theoreticity

Two logical results that have been very relevant in connection with the problem of EE are Craig's theorem and Ramsey's sentence. These results were originally interpreted by logical positivists as showing that theoretical terms are unnecessary in scientific theories[16]. Based on this interpretation, they can be understood as providing algorithms that given an input theory T deliver an empirically equivalent theory T' in which the theoretical conceptual baggage of T has been excised. Kukla has proposed an algorithm that precisely describes the logical maneuver involved: '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). Laudan and Leplin dismiss algorithms like this because the output they produce is not really a genuine rival to T, but simply an *instrumentalized* version of T—in the case of Kukla's algorithm we could call it the *antirealist* version of T. 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 parasitic 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, 456–7).

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 theories—for the theoretical terms are required to derive the empirical consequences of T^*, but they are denied in the latter theory (see the example of intentional psychology below). 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 complains that if parasitism is used against these algorithmic structures, then a principle that is actually applied in scientific practice gets violated (Kukla 1996, 149-50). 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.

[16] 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).

But this argument clearly rests on a misunderstanding. 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.

Kukla's defense of algorithms does not stop here though. He mentions yet another candidate for an EE-algorithm that does not fall prey to the parasitism rejection: 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). This example is enough, he argues, to prove that there exist algorithmic procedures that are capable of producing non-parasitic, predictively equivalent propositional structures. Therefore, if they are going to be rejected, their outputs must be shown to be pseudo-theories on the basis of theoreticity criteria:

> 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. 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).

From Laudan and Leplin's (1993) response to Kukla's challenge three such criteria can be extracted: non-superfluity, plausibility and testability. 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. The algorithm candidate that Kukla proposes includes the postulation of a hypothesis like this: that the laws of nature are intermittent and depend on the presence/absence of observers—by definition, the consequences of T^* do not belong to the class of observable consequences of the algorithmic theory *T'*, so the intermittency hypothesis is empirically superfluous.

Kukla replies that if non-superfluity is to be considered as a criterion of theoreticity it would follow that 'normal' theories like *T* should also be rejected, for they contain the equally superfluous hypothesis that the laws of nature continue to hold when nobody is looking. But this argument is unconvincing. If *T* is a theory in which the content of the laws of nature postulated is such that they hold 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. That is, according to the laws of *T* the behavior of the world is not affected by the presence/absence of observers. 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 the laws in T^* and in *T'* explain '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, the term 'plausibility' fits. In order to be considered as genuinely scientific, a hypothesis must possess a minimum degree of plausibility—which is normally judged on the basis of 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 or theoretical 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 is rather unlikely that any algorithm could do so, given the connection between plausibility and scientific creativity and ingenuity—it follows that it is nothing but a promissory note for an algorithm[17].

One final requirement of theoreticity I 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 & Leplin 1993, 13).

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. I will not deal with a detailed consideration of what these goals 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 intelligible[18] in the sense of making them fit in with general empirically based background knowledge.

It must be underscored that even though theoreticity conditions can be considered as a priori in the sense that they work as pre-given constraints on theories in order to make them genuinely scientific, 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-

[17] 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.

[18] I consider, unlike constructive empiricists, that explanation and understanding are essential aspects of science, see (De Regt & Dieks 2005).

testable and superfluous in a given state of science may become testable and non-superfluous with the introduction of new auxiliary assumptions. 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 theories. Hypotheses are not superfluous, untestable and/or implausible in themselves, but with respect to a concrete state of scientific knowledge[15].

We may recall that, as I said above, support for the universality of first premise of the problematic argument was also given by the Quine-Duhem thesis. But now we may notice that theoreticity constraints also block the holist thesis as providing support for the universal scope of the first premise of the problem. As Adolf Grünbam 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'.

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 imply that there are no automatic algorithms capable of producing an EE rival given any theory *T*[19]. Therefore, the problem that arises from EE and UD is not necessarily universal. 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 ineffective, and theoreticity constraints also block the Quine-Duhem thesis as a logical tool to obtain EE. Moreover, EE and UD, if present, are not necessarily everlasting features, 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 corresponding time-indexed UD of the choice between them is still possible. Although algorithms (including the 'Q-D algorithm') 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 I will argue for below[20].

---

[19] 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. In 'real life' science plausibility for a new hypothesis is usually originated in scientists' creativity and ingenuity, so it is difficult to see how an algorithm could contain a receipt for this property to be included in their output.

[20] It is important 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 preclude that their 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 problem. These remarks prevent a possible objection. The reader might complain that in section 1.3 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 mentioned, theoreticity constraints block algorithms and so undermine the first premise of the problem. I am not using theoreticity as a criterion to make a choice between EE 'real life' theories. For example, even if the degree of plausibility of a certain hypothesis or theory may not be objectively addressed in some cases of 'real-life' science, in the case of 'algorithmic theories' it is clear that the algorithms involved do not include any receipt to provide their outputs with the mentioned property.

1.4.2 The second premise

The second part of Laudan and Leplin's argument is directed against the second premise of the problem, namely, that only observational statements that can be logically derived from a theory can count as empirical evidence to support it. Laudan and Leplin claim that this statement is an overly simplified and inaccurate view of the dynamics of 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.

If we hold on to two basic and very plausible principles of confirmation, then it follows that the class of observational statements that can confirm a theory does not reduce to the class of observational consequences that theory entails. Suppose that a theory T implies two logically independent theoretical hypotheses H_1 and H_2, and in turn H_1 entails the observational consequence e_1. Then, if e_1 is true it counts as empirical evidence for H_1, for T, and also for H_2, even though e_1 is not a logical consequence of H_2. This confirmation scheme relies on the two principles that Laudan and Leplin invoke. The first principle states that evidence for a hypothesis h is also evidence for the statements that imply h[21]—this is why in our example e_1 confirms T. The second principle says that evidence for a hypothesis h is also evidence for the statements that h entails[22]—this is the reason why e_1 confirms H_2 [23].

We can now apply the confirmation scheme and the principles just explained to cases EE. Laudan and Leplin illustrate their point by the following scheme: 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, 461-2). This scheme 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 EE between two theories is not a sufficient condition for UD of the choice to be made 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)

The two underlying principles are quite sound as providing standards of confirmation. For example, in his classic paper on confirmation, Carl Hempel considers them as logical requirements that any plausible account of confirmation has to satisfy (1945, 102 and ff.). On the other hand, it is not difficult to realize that these principles are actually applied in scientific practice. If we reject them, many paradigmatic cases of confirmation of theories by empirical evidence should be queried.

---

[21] Hempel (1945, 104), dubs this principle 'the converse consequence condition'.

[22] This is Hempel's 'special consequence condition' (ibid, 103).

[23] By defending this scheme of confirmation I am not presupposing that evidence is just a matter of logical relations. That is, one could still come up with the question 'how good an evidence is e_1 for H_2?' The pattern of evidence just sketched only means that e_1 can in general *be* evidence for H_2, but *how good* is that evidential support might depend on contextual features or on the specific meanings of the statements involved, for example. But that e_1 can, in general, *be* evidence for H_2 is enough for the view that is being defended here.

This is already enough to deny that the only observable statements that can provide empirical confirmation for a theory are statements that can be logically derived from the theory. Anyways, Laudan and Leplin's attack on the second premise of our problem can be 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)[24]

Boyd's (adapted) view[25] 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 feature.

The epistemic justification of the principle we have extracted from Boyd's argument is a very basic goal of science: mutual consistency between accepted theories. Suppose that T and T' are EE, that T is consistent with another wellconfirmed 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

[24] A *universal force*, roughly speaking, is a force that acts equally on all physical objects and that it cannot be shielded against. A *differential force*, on the contrary, can be shielded against and does not act equally on all physical objects. See (Reichenbach 1958, §6). The argument asserts the conventionality of physical geometry insofar as the empirical equivalence between the theories renders the choice between G and G' as a matter of convention.

[25] 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.

argument should be taken as a principle in the dynamics of empirical confirmation. This is a very plausible stance of course: 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 were inconsistent, we would hardly call such a set 'knowledge'. 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 scientists.

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. First, Samir Okasha (1997) objected that Laudan and Leplin's argument falls prey to a problem that Hempel had already noticed in 1945. Okasha correctly claims that the epistemic support for Laudan and Leplin's argument is given by the following—prima facie very plausible—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, respectively. Now, the pattern of empirical confirmation that Laudan and Leplin propose in the context of EE presupposes that both principles are at work at the same time. 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 *xi)* e does not confirm H_2. However, as Okasha reminds us, Hempel 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 $(S \,\&\, T) \rightarrow S$. By special consequence, S confirms T, since $(S \,\&\, T) \rightarrow T$. 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 an abstract exercise in logical entailment relations. However, 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, 104-5). 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.

It is 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 by a hypothesis is not a necessary condition for that statement to confirm the hypothesis. Inter-theoretical relations are also crucial features to be considered.

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_2: 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, 273–4)

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 does not 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, the joint actual existence of T and T^* 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 with regard to 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 the eventual breakdown of UD may be undone again 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 that it cannot return.

1.5 REMAINING CHALLENGES

EE between theories can be instantiated in four different ways: *i)* by algorithms, *ii)* by accommodating auxiliary hypotheses according to the Duhem-Quine thesis, *iii)* by the regular practice of science, and *iv)* by concrete artificial examples. The universal scope of the first premise of the problem is supported by i) and ii). If there exist algorithms that are able to produce EE theories given any theory T, or if it is always possible to accommodate evidence by means of manipulation of auxiliary hypotheses, then it follows that EE is a condition that holds for any theory whatsoever. As we have seen, neither *i)* nor *ii)* really work as possible sources of EE . In the case of *iii)*, Laudan and Leplin's argument shows that EE is a time-indexed feature and that it might get broken by future scientific or technological developments; and also that the UD between EE theories can be broken by means of non-consequential empirical evidence, even if the predictive equivalence remains .

In this section I will tackle the remaining source of EE, namely, concrete examples of artificially generated pairs of empirically equivalent theories. These examples are neither the outcome of the application of algorithms, nor obtained by manipulation of auxiliary hypotheses given an actual theory T. They are not the result of the practice of real science either. Rather, they have been cooked up and exploited by philosophers of science in order to speculate about their epistemological consequences. I will address an examination of three examples of artificially generated EE theories that have received attention in the philosophy of science literature: Bas van Fraaseen's alternative formulations of Newton's mechanics; the theories involved in the Poincaré-Reichenbach 'parable'; and the case of predictively equivalent total theories or systems of the world.

1.5.1 Van Fraassen's alternative formulations of Newton's theory

In *The Scientific Image* Bas van Fraassen introduced an argument for his constructive empiricism that involves an example of EE. He presents Newton's theory as a theory about the motion of bodies in space and the forces that determine such motions. The crucial feature that grounds van Fraassen's argument is that Newton's theory is supposed to be committed to the view that physical objects exist in absolute space. Thus, by reference to absolute space the concepts of absolute motion and absolute velocity become meaningful. Then, van Fraassen proposes

> let us call Newton's theory (mechanics and gravitation) TN, and $TN_{(v)}$ the theory TN plus the postulate that the center of gravity of the solar system has constant absolute velocity v. By Newton's own account, he claims empirical adequacy for $TN_{(0)}$; and also that if $TN_{(0)}$ is empirically adequate, then so are all the theories $TN_{(v)}$. (van Fraassen 1980, 46)

Newton's most famous argument for the existence of absolute space is given by the thought experiment of the rotating bucket. In order to make sense of the acceleration of the rotating water in the bucket, the reality of absolute space has to be asserted, Newton argued. Van Fraassen's line of reasoning is that if absolute space exists, as Newton believed, then the concept of absolute motion of objects in space gets defined and so does the concept of absolute velocity. However, since—unlike absolute acceleration—absolute velocity has no observable effects, there are infinitely many predictively equivalent rival formulations of TN, each of them assigning a different specific value to the absolute velocity of the solar system's center of gravity.

According to van Fraassen, this entails a problem for the realist. The realist is committed to the view that only one of these alternative formulations is the true theory, but the realist's choice cannot be determined on evidential grounds[26]. For the constructive empiricist, van Fraassen argues, there is no such problem. In his/her case there is no commitment to the truth of the theory, but only to its empirical adequacy. Therefore, for the constructive empiricist it is enough to accept the empirical content of the theory as empirically adequate and assume a dodging attitude with respect to its non-empirical content—including the value for the absolute velocity of the solar system, of course. In other words, the empirical equivalence of the alternative formulations of Newton's theory does not necessarily put the constructive empiricist in the position of having to make a choice[27].

A systematic consideration of van Fraassen's challenge shows that the real problem is not EE. It is true that Newton endorsed absolute space and that his preferred alternative was $TN_{(0)}$. However, rather than a case of EE, what is behind van Fraassen's example is a situation where there is a superfluous hypothesis within TN. A hypothesis is superfluous if it is not logically relevant for the derivation of any empirical consequences of the theory it forms a part of; and a hypothesis being superfluous is a strong indication that it represents nothing physical—an ontologically empty hypothesis, we could say. Therefore, the fact that the predictive equivalence between van Fraassen's alternative formulations is grounded on the stipulation of a specific value for a superfluous parameter—absolute velocity—indicates that we have a problem with the foundations of $TN_{(v)}$,rather than a genuine problem of EE.

The problem of the superfluity of the concept of absolute velocity in Newton's theory has actually been solved and, *a fortiori*, the specious problem of EE gets dissolved. The key concept is a structure known as *neo-Newtonian space-time*[28]. The basic elements of this structure are event-locations—the spatiotemporal locations where physical events (can) occur. A temporal separation—that can be zero—is defined for all pairs of event-locations, and this is an absolute relation in the sense that it is not relative to particular frames of reference, states of motion, etc. A class of simultaneous event-locations—those for which their temporal separation is zero—forms a *space*[29], and the structure of each space is that of Euclidean three-dimensional space.

The feature that differentiates Newtonian absolute space and neo-Newtonian spacetime is the way in which the spaces are connected or 'glued-together'. In absolute Newtonian space points conserve their

[26] From the viewpoint of the semantic conception of scientific theories, that van Fraassen endorses, the realist is committed to the view that *only one of the models that satisfy $TN_{(v)}$ correctly represents the world*. In the case of Newton, that model is given by $TN_{(0)}$, though the absolute velocity of the solar system is not a phenomenon.

[27] In semantic temrs, the constructive empiricist stance is that to accept $TN_{(v)}$ as empirically adequate means that $TN_{(v)}$ *has a model which is empirically adequate*, i.e., it possesses an empirical substructure isomorphic to all phenomena. Making a choice is possible for a constructive empiricist, and he/she could do it based on pragmatic features of one of the alternative formulations. However, such a choice does not have an epistemic import, according to van Fraassen's view.

[28] Neo-Newtonian spacetime is the result of the work of P. Frank in 1909, and E. Cartan and K, Friedrich in the 1920s. For a technical exposition of neo-Newtonian space-time and references to the seminal works of Frank, Cartan and Friedrichs, see (Havas 1964). For simpler expositions see (Sklar 1974) and (Stein 1970).

[29] In neo-Newtonian space-time simultaneity is an equivalence relation: every event is simultaneous with itself, if a is simultaneous with b, then b is simultaneous with a, and if a is simultaneous with b and b is simultaneous with c, then a is simultaneous with c. Therefore, it is possible to divide the class of all events in equivalence classes under the relation of simultaneity—classes that have no members in common and that taken together exhaust the class of all events.

spatial identity through time, and it is thus meaningful to ask whether a certain point or event-location at time t_1 is identical with some point or event-location at time t_2. In neo-Newtonian spacetime this question makes no sense, since the notion of spatial coincidence is only defined for simultaneous event-locations.

This difference in structure has a straightforward effect on the way that velocity is defined in each case. In neo-Newtonian space-time it is coherent to ask for the velocity of a particle between two events in its history, but only if we are talking about its velocity with respect to some particular object or frame of reference—we can ask if the distance of the particle with respect to another object or frame is the same as its distance to that same object or frame at an earlier time, of course. But since absolute spatial coincidence through time is not defined, the concept of 'absolute velocity' is meaningless in neo-Newtonian space-time. Since points or event-locations do not conserve their identity through time, we cannot ask if the distance of an object with respect to a certain point in space at time t_2 has changed, or not, with respect to the distance between the object and that same point at an earlier time t_1.

Even though 'absolute position' and 'absolute velocity' are undefined, the concept of 'absolute acceleration' is well defined in neo-Newtonian space-time, but this definition does not require reference to absolute space. First we need to introduce the three-place relation of 'being inertial' between three non-simultaneous event-locations a, b and c. The relation holds if there is a possible path for a particle such that three events in its history are located at a, b and c, and if the particle is at rest in some inertial frame—a frame in which no inertial forces act upon any physical system at rest in it. More generally, a collection of events conforms an inertial class of events if they are all locations of events in the history of some particle that moves free of forces, a particle that moves inertially.

We can now explain the absolute acceleration of a particle along a time interval. Take the particle at the beginning of the interval and find an inertial frame in which the particle is at rest. At the end of the interval we find the new inertial frame in which the particle is at rest. Then we find the relative velocity of the second frame with respect to the first one at the end of the interval. Even though there is no such thing as the absolute velocity of the first inertial frame, we do know that, by definition, its velocity—with respect to any other inertial frame—has not changed throughout the interval. Therefore, the relative velocity of the second frame with respect to the first one gives us the absolute change of velocity throughout the interval, since the particle was at rest with respect to the first frame at the initial instant, and at rest with respect to the second frame at the end. We take this absolute change of velocity and divide it by the time separation between the initial and final event-locations and we obtain the absolute acceleration of the particle over the interval, and by applying the usual limiting process of differential calculus on the time interval we generate the concept of instant absolute acceleration. That is, absolute acceleration, within the context of a neo-Newtonian space-time, is defined not as relative to absolute space, but as relative to any inertial frame.

Now we can go back to van Fraassen's challenge. As I mentioned above, the formulation of Newtonian mechanics in terms of neo-Newtonian space-time can be understood as the solution for an unease about its foundations—the superfluous concept of absolute velocity. That is, the example that van Fraassen offers is not a genuine case of EE between rival theories. The problem is simply that the presence of the superfluous parameter v in TN manifested in that alternative, apparently incompatible formulations could be given. Neo-Newtonian space-time solves this problem. It allows a more satisfactory formulation of TN in which the superfluous parameter has been swept away, so that there is no EE arising from different values assigned to v. In other words, the EE equivalence between van Fraassen's formulations was not the sickness, but just a symptom. Therefore, van Fraassen's challenge cannot be fruitfully used in order to extract conclusions related to the problem of EE and UD. These remarks, of course, do not

intend a refutation of constructive empiricism. The point is only that this particular example has no relevant consequences regarding the problem of EE and UD[30].

1.5.2 The Poincaré-Reichenbach argument

In *Science and Hypothesis*, Henri Poincaré introduced an argument for the conventionality of geometry that has been considered as an example of EE. He designed a 'parable' in which a universe given by a Euclidean two-dimensional disk is inhabited by flatlanders-physicists. The temperature on the disk is given by $T(R^2 - r^2)$, where R is the radius of the disk and r is the distance of the location considered to the center of the disk—therefore, the temperature at the center of the disk is TR^2 and at the edge it is 0° absolute. The inhabitants of this world are equipped with measuring rods that contract uniformly with diminishing temperatures, and all such rods have length 0 when their temperature is 0°. The two-dimensional physicists proceed to measure distances in the disk with their rods in order to determine the geometry of their world; but they assume, falsely, that the length of their rods remains invariant upon transport—the flatlanders themselves also contract with diminishing temperature. Accordingly, the result they obtain is that they live in a Lobachevskian plane of infinite extent. For example, they measure that the ratio of a circumference to its radius is always greater than 2π. They obtain the same result by using measurements performed with light rays, for their universe is characterized by a refraction index $1/(R^2 - r^2)$; but they falsely assume that light beams travel along geodesics in their world, and that the index of refraction of vacuum is everywhere the same.

The parable also tells us that one particularly smart and revolutionary scientist in the disk comes up with the correct theory about the geometry of their world. Even though they are not able to observe effects of the temperature gradient $(R^2 - r^2)$ and of the refraction index $1/(R^2 - r^2)$, our brilliant physicist notices that, by assuming the reality of such unobservable features, the result is that the geometry of their universe is that of a finite Euclidean disk. The scientific community on the disk does not have the resources to make an evidentially based decision between the theories, and Poincaré's point is that the only way they can determine a specific geometry for their world is in terms of a *convention*. Poincaré also states that in our three-dimensional world we are, in principle, in the same situation. Empirically equivalent theories of our world that differ in the geometry they pose are analogously attainable. Therefore, the geometry of the physical world is a matter of convention also for us.

Two remarks can be made at this point about Poincaré's argument. First, it is clear that it is not an argument directly aiming to extract conclusions about the problem of EE and UD; but an argument concerning the epistemology of geometry. This feature indicates that if we are going to take it as a concrete example of EE and UD some provisos must be introduced. Second, it is also clear that the example of empirically equivalent theories it considers is of a peculiar kind. The theories are not about the 'real' physical world. The universe of the flat disk is a mental construction and, as such, it can be arranged and manipulated so that it totally complies with the description given by each of the theories. The world described by the theories is an *ad hoc* world. But this feature of the argument suggests that the example of EE involved is not a very serious or threatening one. The choice between the theories is underdetermined because the whole situation can be conceptually manipulated in the required way.

[30] The reader might complain that since the alternative formulations of $TN_{(v)}$ are based on a theory that forms part of real physics means that van Fraassen's argument is a case in which EE is supposed to arise from the actual practice of science, not an artificial example. However, notice that a choice between formulations of $TN_{(v)}$ was never an issue for the scientific community, there never was a scientific debate about what is the correct value of v. What did happen was a debate concerning the meaningfulness of v—Leibniz's arguments in the *Leibniz-Clarke correspondence*, for example. This debate was not grounded on a problem of EE and UD of theory choice, it was a debate about the ontology of space. This is yet another indication that van Fraassen is exploiting a problem with the foundations of Newton's theory in order to create a (specious) artificial case of EE.

Hans Reichenbach, in *The Philosophy of Space and Time*, introduced a sort of generalization of the argument. He presented it as a theorem showing that from any spacetime theory about the *real* physical world it is possible to obtain an alternative theory which is predictively equivalent but that assigns a different geometry:

> Mathematics proves that every geometry of the Riemannian kind can be mapped upon another of the same kind. In the language of physics this means the following:
> *Theorem* θ: 'Given a geometry G' to which the measuring instruments conform, we can imagine a universal force F which affects the instruments in such a way that the actual geometry is an arbitrary geometry G, while the observed deviation from G is due to a universal deformation of the measuring instruments'." (Reichenbach 1958, 32-3)[31].

Under this formulation, the argument for the conventionality of geometry has a more substantial upshot on the problem of EE and UD. Reichenbach claims that the parable that Poincaré introduced can be effectively applied to 'real' spacetime theories. For example, it could be stated that general relativity is empirically equivalent to a Newtonian-like theory of gravitation in which the curvature of spacetime is replaced by the action of a universal force. This complies with the first remark I made above regarding Poincaré's parable. Under Reichenbach's formulation, the argument for the conventionality of geometry can, in principle, be considered as an instance of EE involving theories about *our* world.

However, we still need to be precise about in what sense this argument, that primarily concerns the epistemology of geometry, affects the problem of EE and UD. For this purpose it is useful to take a look at what exactly Reichenbach is arguing for. The conventionalist stance he defends is weaker than Poincaré's. According to Reichenbach, what is a matter of convention regarding geometry are not, bottom line, the geometric features of the physical world, but the specific 'language' in which those features are expressed. This argument relies on the concept of *coordinative definition*, that is, arbitrary definitions that settle units of measurement and which ground the particular conceptual systems that underlie physical theories:

> Physical knowledge is characterized by the fact that concepts are not only defined by other concepts, but are also coordinated to real objects. This coordination cannot be replaced by an explanation of meanings, it simply states that *this concept* is coordinated to *this particular thing*. In general this coordination is not arbitrary. Since the concepts are interconnected by testable relations, the coordination may be verified as true or false, if the requirement of uniqueness is added, i.e., the rule that the same concept must always denote the same object. The method of physics consists in establishing the uniqueness of this coordination, as Schlick has clearly shown. But certain preliminary coordinations must be determined before the method of coordination can be carried any further; these first coordinations are therefore definitions which we shall call *coordinative definitions*. They are *arbitrary*, like all definitions; on their choice depends the conceptual system which develops with the progress of science.
> Wherever metrical relations are to be established, the use of coordinative definitions is conspicuous. If a distance is to be measured, the unit of length has to be determined beforehand by definition. This definition is a coordinative definition. (Reichenbach 1958, 14-5)

Now it becomes clear why I said that Reichenbach's conventionalist view is a 'weak' one. What is at stake in the EE between theory $T = F + G$ and $T' = F' + G'$—where F denotes the set of forces that affect physical objects according to T, and F' is that same set plus a universal force f' that accounts for the deviation from geometry G according to T'—is only a divergence regarding the particular coordinative definitions that are presupposed by the theories. That is, we are in a situation analogous to a decision concerning whether Lionel Messi's height is 1,69 meters or 5 feet and 7 inches. In the case of Poincaré's disk, there are two different coordinative definitions at stake: one states that distances measured by rods

[31] Recall that *universal force*, roughly speaking, is a force that acts equally on all physical objects and that it cannot be shielded against. A *differential force*, on the contrary, can be shielded against and does not act equally on all physical objects.

have to be corrected according to a certain law, whereas in the other the measuring rods are rigid bodies that always express correct distances. Reichenbach's view on the conventionality of geometry is 'linguistic', we could say. T and T' are two versions of the same theory expressed in different geometrical *languages.* To state that T is truer or more correct than T', or vice versa, is analogous to say that 'meter' is a more correct unit of measurement than 'foot'[32].

If Reichenbach is right, then the case of EE between T and T' that the argument involves is a harmless one. The choice between the theories is just a matter of the language we pick to express the same physical theory. Under Reichenbach's view the conventionality of geometry has no special upshot on the problem of EE and UD as defined above. It is true that the choice between T and T' can be done only in terms of pragmatic considerations such as simplicity—empirical evidence, by definition, cannot settle the case. However, this is not a scientific or epistemological problem at all, for the choice does not involve incompatible rivals that differ in the way they describe the world. If we follow Reichenbach's line of thought, a genuine case of EE and UD would happen only if the theories involved postulate incompatible geometrical features for the world *provided that in both theories the universal forces are set to the zero value*. There is nothing in Reichenbach's argument to believe that this cannot happen, but it does not involve any example of this kind either.

This easy way out of the problem works only if Reichenbach is right, of course. His position regarding the epistemology of geometry is, clearly, quite close to the verificationist criterion of meaning endorsed by most of logical positivists. As it is known, this criterion has been shown to be untenable, and Reichenbach's view of the meaning of geometrical statements as reducible to coordinative definitions falls prey, *mutatis mutandis*, to the typical objections that have been leveled against logical positivistic semantics. That is, there are good reasons to think that Reichenbach's position is wrong, and, *a fortiori*, that the case of EE involved in his argument might be a relevant example with respect to the problem of UD of theory choice.

However, it turns out that even if we consider the case of $T = F + G$ vs. $T' = F' + G'$ as a genuine case of EE, this does not necessarily imply that we are dealing with a case of UD. The reason is given by the evidential status of the 'universal forces'. We can understand Reichenbach's theorem as stating that spacetime theories can have alternative empirically equivalent formulations by means of universal forces, and we can assume—unlike Reichenbach—that such alternatives are *genuine* rivals. However, that there exists an EE rival that postulates the reality of universal forces is not, *ipso facto*, an indication that the choice to be made is underdetermined by the empirical evidence.

All 'real' physical theories that invoke forces as the cause for dynamical effects postulate these forces as associated to *observable* effects; but the universal forces involved in Reichenbach's arguments are not at all like these 'typical' forces. They are, in principle, not associated to any empirically detectible effect.

---

[32] Reichenbach also argues that the *default* language is the geometry in which universal forces are set to the zero value. If we do so, then the question regarding the specific geometry of the physical world becomes really meaningful, not only a matter of linguistic definitions: 'The forces which we called universal are often characterized as forces *preserving coincidences*; all objects are assumed to be deformed in a way that the spatial relations of adjacent bodies remain unchanged. […] It has been correctly said that such forces are not demonstrable, and it has been correctly inferred that they have to be set equal to zero if the question concerning the structure of space is to be meaningful. It follows from the foregoing considerations that this is a *necessary* but not a *sufficient* condition. Forces *destroying coincidences* must also be set equal to zero, if they satisfy the properties of the universal forces […]; only then is the problem of geometry uniquely determined. […] We can define such forces as equal to zero because a force is no absolute datum. When does a force *exist*? By force we understand something which is responsible for a *geometrical change*. If a measuring rod is shorter at one point than at another, we interpret this contraction as the effect of a force. The existence of a force is therefore dependent on the coordinative definition of geometry. If we say: actually a geometry G applies, but we measure a geometry G', we define at the same time a force F which causes the difference between G and G'. The geometry G constitutes the zero point for the magnitude of a force. If we find that there result several geometries G' according as the material of the measuring instrument varies, F is a differential force; in this case we gauge the effect of F upon the different materials in such a way that all G' can be reduced to a common G. If we find, however, that there is only one G' for all materials, F is a universal force. In this case we can renounce the distinction between G and G', i.e., we can identify the zero point with G', thus setting F equal to zero. This is the result that our definition of the rigid body achieves' (Reichenbach 1958, 27-8).

The reality of usual, differential forces in physical theories is evidentially supported by the observable effects they cause, but this is not the case with universal ones. That is, in the case of $T' = F' + G'$ there is a hypothesis which is not, in principle, evidentially warranted. Therefore, we can conclude that $T = F + G$ possesses a higher degree of evidential support than T'. That is, we are simply applying the principle of confirmation that we extracted from Boyd (1973):

> 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 [universal] force f' required by F' [the class of the forces postulated by our T'] is dramatically unlike those forces about which we now know—for instance, it fails to arise as the resultant of fields originating in matter or in the motions of matter. Therefore, it is, in the light of current knowledge, highly implausible that such a force as f' exists.
> Furthermore, this estimate of the implausibility of "F'&G'" reflects *experimental* evidence against "F'&G'", even though this theory has no falsifies observational consequences. (Boyd 1973, 7-8)[33].

In this context, Boyd's passage is illuminating in two respects. First, it is not only the unobservability of a universal force what makes it bizarre and lacking evidential support. It is also a very implausible concept, in the sense that it is not alike at all to usual forces in another crucial respect: there is nothing in Reichenbach's theorem to let us know about its physical underpinning. Usual forces have a source, for example—typically charges and massive objects—; but what is the source of universal forces? Second, the quote underscores that the problematic nature of universal forces is not just a matter of theoretical uneasiness. Universal forces are bizarre not only from the point of view of formal *a priori* or pragmatic considerations. The difficulties with them are also based on lack of *empirical evidence* to support their reality. Let me clarify this point with yet another quote, this time from a paper by John Norton:

> I must note that the notion of a universal force, as a genuine, physical force, is an extremely odd one. They are constructed in such a way as to make verification of their existence impossible in principle. The appropriate response to them seems to me not to say that we must fix their value by definition. Rather we should just ignore them and for exactly the sorts of reasons that motivated the logical positivists in introducing verificationism. Universal forces seem to me exactly like the fairies at the bottom of my garden. We can never see these fairies when we look for them because they always hide on the other side of the tree. I do not take them seriously exactly because their properties so conveniently conspire to make the fairies undetectable in principle. Similarly I cannot take the genuine physical existence of universal forces seriously. Thus to say that the values of the universal force field must be set by definition has about as much relevance to geometry as saying the colors of the wings of these fairies must be set by definition has to the ecology of my garden." (Norton 1994, 165)[34]

---

[33] Kyle Stanford offers a similar account of the matter: 'While Eddington, Reichenbach. Schlick and others have famously agreed that general relativity is empirically equivalent to a Newtonian gravitational theory with compensating 'universal forces', the Newtonian variant has never been given a precise mathematical formulation (the talk of universal forces is invariably left as a promissory note), and it is not at all clear that it can be given one (David Malament has made this point to me in conversation). The 'forces' in question would have to act in ways no ordinary forces act (including gravitation) or any forces could act insofar as they bear even a family resemblance to ordinary ones; in the end, such 'forces' are no better than 'phantom effects' and we are left just with another skeptical fantasy. At a minimum, defenders of this example have not done the work needed to show that we are faced with a credible case of non-skeptical empirical equivalence' (Stanford 2001, S6, footnote 6).

[34] In his paper Norton introduces this comment only as a sort of side-remark: 'As an aside from my main argument…' (ibid). His main goal is to disprove the conventionality of geometry, and his main argument is that universal forces finally reduce to 'correction-terms' in suitable gauge transformations that preserve the physical meaning of invariants in general covariant formulations of spacetime theories. This implies that the underlying metric in the spacetime theories involved is not affected at all by the introduction of universal forces. As Norton himself explicitly acknowledges, this is a refutation of a strong version of the conventionalist thesis. This argument leaves the weak 'linguistic-definitional' version untouched—which is Reichenbach's stance—though Norton states that such a version is trivial.

I totally agree with Norton's view regarding universal forces[35]. However, this stance, I think, should not be taken as an ultimate rejection of them as a possible part of scientific theories. Hypotheses are not testable or untestable in *a priori* terms. For example, new available auxiliary hypotheses could be conjoined to a certain untestable hypothesis and turn it into a testable one. I see no reason why this might not happen in the case of universal forces. That is, *so far as we know*, there is not empirical evidence for the reality of such forces, but future findings might provide good reasons to postulate them in physical theories. A future theory could include observable effects that, at least indirectly, support the reality of a universal force.

It is important to underscore that these remarks hold for universal forces as such, that is, independently of their involvement in Reichenbach's argument. Actually, it seems that this particular argument requires, by definition, that universal forces are not related to any observable effects. The EE between T and T' seems to have as a condition that the universal forces are totally undetectable. However, as I just mentioned—and putting Reichenbach's argument aside—, there might be possible physical theories in which universal forces do relate to observable features. At least this possibility has not been disproven.

The answer to the question of whether Reichenbach's argument involves a challenging case of EE is thus negative. The reason is that, so far as we know, there is no evidential support for the reality of universal forces. Therefore, even though we could concede that Reichenbach's example involves genuine EE and rivalry, this does not mean that we are facing a case of UD, for the theory in which universal forces are absent has more evidence in its favor than its rival. Moreover, the fact that Reichenbach's argument requires that the universal forces involved are totally undetectable suggests that this particular example cannot provide a case of UD, no matter what particular form these forces take within the theory they are a part of. If universal forces are to have any special consequences with respect to EE and UD, it will not be through Reichenbach's example[36].

1.5.3 *Total theories* or *systems of the world*

The last case of EE in terms of artificial examples I will address is given by *total theories* or *systems of the world*. Such theories are defined by providing an account of all possible phenomena, past present and future, in opposition to regular 'local' theories that hold for a determinate realm of appearances:

> The thesis of underdetermination of theory choice by evidence is about empirically adequate total science; it is a thesis about what Quine calls 'systems of the world'—theories that comprehensively account for all observations—past, present and future. It is a thesis about theories that entail all and only the true observational conditionals, all the empirical regularities already confirmed by observation and experiment." (Hoefer and Rosenberg 1994, 594).

Hoefer and Rosenberg accept the solution proposed by Laudan & Leplin's, but they correctly affirm that it cannot work in the case of total theories—that's why they state that the problem of EE and UD is

[35] Reichenbach's followers could reply that, since they define force as something which is responsible for a geometrical change, and therefore it essentially depends on the coordinative definitions underlying a physical geometry (see footnote 8 above), then the reality of universal forces is also a matter of convention—and their introduction becomes justified. But this answer only shifts the problem. The usual physical meaning of 'force' is much more substantial than a mere stipulation about the presence or absence of geometrical changes. Reichenbach's conventional definition of force is quite debatable.

[36] There is an alternative way to tackle this example of EE. Rather than as an artificial instance of EE, we could consider Reichenbach's theorem as an *algorithm* to instantiate EE. The algorithm would be something like this: *given a space-time theory T, it is possible to construct a rival theory T' by introducing a universal force f' with a suitable value as to assign a different metric to the world, while conserving all of T's empirical consequences*. But according to the general criticism of algorithms explained in section 1.4, it is clear that this procedure is not enough to provide a genuine scientific theory. More precisely, *f'* cannot be any force whatsoever, it must show a minimum degree of plausibility and it must be testable and relevant to entail consequences, but the algorithm says nothing about how to achieve that.

a problem only for total theories. Since Laudan and Leplin's argument makes essential reference to background science, that is, to *other* theories, if we are dealing with systems of the world such other theories are, by definition, not available. All possible auxiliary hypotheses are included in the EE total theories involved, and there cannot be more general theories in which to encompass any system of the world. Therefore, EE in the case of total theories seems to pose a special challenge.

I think that it is true that if a pair of predictively equivalent theories of this kind were given, then the UD involved could not be overcome. However, we do not need to worry about this example of EE either. Even though the very definition of a system of the world precludes that UD could be broken in terms of empirical evidence if EE is given, this definition is problematic in the sense that there is no way for us to know whether a specific theory counts as a system of the world or not.

There are several ontological and epistemological difficulties with the concept. First, if we are going to take systems of the world seriously, it would have to be shown that the world admits a description by a theory like that. This question involves a metaphysical issue of course: is the set of all natural phenomena regular and coherent enough as to be describable in terms of one single theoretical framework? Second, and taking for granted that this it is possible to describe the world by one single theoretical framework, is human science capable to provide an alternative, rival, predictively equivalent system of the world? If we discard algorithms and bizarre, parasitic theories this sounds like an extremely unlikely scenario.

It could be argued that the possibility of a total theories-EE scenario has not been disproven, and that this is enough to take the problem seriously. We can concede this, but the problems with the concept of a system of the world do not end here. Recall that the definition involves the property of being empirically adequate for all possible phenomena, past, present and future; but how in the world could we know that a certain (total) theory will be empirically adequate with respect to all future phenomena? Notice that the problem is not that we cannot know whether a certain total theory is true (or empirically adequate) or not; the problem is that since we can never know that a certain theory is empirically adequate with respect to future phenomena implies that *we cannot know whether a certain theory is really a system of the world*. That is, the very definition of the concept at issue precludes us to know that any candidate-theory is really a total one or not.

Analogously, we cannot know whether a certain theory has all possible phenomena under its scope. It is true that by its form and content a certain theory can claim to be valid in a total way—for all possible phenomena—but the fact that a certain theory intends to be a total one does not necessarily mean that it is. Our world is not like the universe in Poincaré's parable, we cannot accommodate it in a way such that it complies with our theoretical framework. There might always be realms of phenomena that are not accounted for in a theory, even if such a theory intends to be a system of the world. For example, assume that we are facing a case of EE between two total theories. In spite of what the theories say, nothing precludes the possibility that new kinds of phenomena—that have never been observed before and that cannot be accounted for by any of the theories involved—get detected. This already shows that we can never know if the theories involved are total or not. Besides, if such unexpected phenomena are indeed detected, then the problem of EE and UD at issue could be solved à la Laudan and Leplin—the auxiliary hypotheses provided by a new theory that explains the unexpected phenomena could break the predictive equivalence, for example.

The upshot of these remarks for the problem of EE and UD is clear. It is true that if two total theories are EE then the UD of the choice would be a big problem[37]. However, from the point of view of human scientific knowledge, the very concept of a system of the world is problematic. It is impossible to know whether a certain theory qualifies as a total one. At most, philosophers can speculate about their epistemological and/or metaphysical consequences on a high level of abstraction, but total theories do not

[37] If one of the theories includes implausible universal forces, for example, the alternative theory might be better supported by evidence in spite of the EE. That is, the EE between systems of the world would be a big problem granted that both the theories are genuinely scientific and have solid foundations.

present a serious case of EE and UD in the context of the philosophy *of science*. The situation is thus analogous to Descartes' evil-genius argument. It is an interesting and serious issue in metaphysics and general epistemology, but it does not have any particular or relevant consequences for the philosophy of science[38].

1.6 THE REAL STATUS OF THE PROBLEM OF EE AND UD

Laudan & Leplin's argument is essentially correct. However, it must be reassessed in order to correctly appreciate what it achieves. I turn, such a reassessment allows us to understand what is the real situation concerning the problem of EE and UD. Laudan & 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. Based on this analysis they conclude that

> This contextuality [time-index] 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 (Laudan and Leplin 1991, 454, my emphasis).

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 its observational consequences. 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. Accordingly, Laudan and Leplin draw the following conclusion:

> Results that test a theory and results that are obtainable as empirical consequences of a theory constitute partially nonoverlapping sets. Being an empirical consequence of a theory is neither necessary nor sufficient to qualify a statement as providing evidential support for the theory. Because of this, it is illegitimate to infer from the empirical equivalence of theories that they will fare equally in the face of any possible or conceivable evidence. 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'[39]—stands refuted. (ibid, 466; my emphasis)

In order to attain an accurate evaluation of the implications of EE and UD, these conclusions must be carefully assessed. Laudan and Leplin address the EE/UD problem in its universal and general form: the

---

[38] Samir Okasha has offered an objection to the cogency of the very concept of a total theory, but along a different line of reasoning. He claims that since the theoretical-observational distinction is not absolute, but context-dependent—a certain term in a theory counts as theoretical, but the same term in a different theory can count as observational—neither the observational content nor the theoretical apparatus of a system of the world can be defined: 'If we are even to understand this suggestion [that EE between two total theories leads to UD], let alone endorse it, we must have a criterion for deciding which side of the divide an arbitrarily chosen statement falls on. But such a criterion is precisely what the minimal, context-relative theory/data distinction does not give us. If that distinction is all we have to go on, we can get no grip on what it means for our 'global theory' to be underdetermined by the 'empirical data', nor indeed on what a 'global theory' is even supposed to be' (Okasha 2002, p. 318).

[39] After the results of the previous section, we know that in this passage Laudan and Leplin cannot be using the term 'systems of the world' in its canonical meaning.

threat that EE holds for all scientific theories and that, consequently, UD infects all theory choice. As we have seen, their arguments are effective in denying the generality of the problem—it is not true that for all theories there is an EE rival, and EE is not a sufficient condition for UD. However, the conclusions that the epistemological morals derived from EE are otiose and that the thesis of UD as founded on presumptions of EE stands refuted, can be understood as implying that EE and UD are epistemologically idle issues[40]. So understood, these conclusions do not provide a reliable analysis of the seriousness of EE and UD. First, even if EE is a time-indexed feature and theoreticity requirements can block algorithms, that two theories may be EE remains a possible scenario, for a genuinely scientific EE rival to a given theory might be formulated after all. Besides, 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. Second, it is true that a more general theory could break the UD by means of '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. Recourse to intertheoretic 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.

That two genuinely scientific theories are EE (with respect to a given state of science) remains a possible situation. If (also with respect to a given state of science) it happens to be the case that both theories stand on equal footing in terms of compatibility with background knowledge, and if there are no general theories to encompass them, it does follow that the choice between the theories is underdetermined by empirical evidence. That is, in a case like this—a case whose possibility Laudan and Leplin's arguments do not deny—the morals of EE are not epistemically otiose, for UD would indeed follow from EE. The problem of UD as a consequence of EE has not been dispelled.

From Laudan and Leplin's proposal it can be accepted that the problem of UD, just as EE, is a time-indexed feature, and this is certainly an important clarification. But that the problem may disappear with time does not imply that there is no problem at all. EE leading to UD can happen in science, and that there is a solution *à la* Laudan and Leplin is not guaranteed in any specific case. What has been achieved is a demonstration that strategies and methods typically used in science might be effective in overcoming the problem if it comes up. In other words, Laudan and Leplin's proposal does not prove that EE and UD do not constitute a problem, though it certainly 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. But, after all, this is a feature that all the problems that science is to solve share.

Anyways, that EE and UD constitute a problem that *science* may solve in individual cases does not mean that we are dealing with an epistemologically idle situation. Actually, the very fact that the solution of the problem is not very different from the solution of other scientific problems implies that important epistemic features are involved. That the scientific solution is contingent, in the sense that the conditions required for the breakdown of EE and/or UD might or might not obtain, means that recalcitrant cases of UD of theory choice are possible. Moreover, elaborating on Bangu's objection and on the contingency of Boyd's solution, we note that a problem of EE and UD that gets eventually solved might become problematic again—variations in well confirmed background knowledge or the introduction of further suitable encompassing theories might reintroduce the problem. It is clear that a situation like this would pose important questions concerning the epistemological status of theory acceptance.

We can schematically summarize the results of this chapter in the following statements:

[40] Actually, even if we interpret Laudan and Leplin's argument as directed only against the universality and generality of the problem, they do not even mention that a remaining 'local' problem of EE and UD still stands—let alone that this remaining problem has important epistemic dimensions.

i) even though algorithms and the Q-D thesis are ineffective in providing an EE rival theory T' given any theory T, and although EE is a time-indexed condition between two theories, EE between scientific theories that results in UD is still a possible scenario;

ii) that recourse to non-empirical features can work as a partial solution of the problem, in the sense that such features can provide rational motives in order to prefer one of the theories, but cannot ground a uniquely determined and fully objective choice;

iii) that Laudan and Leplin are right in that the UD of the choice to be made between EE theories can be removed by the breakdown of the regular practice of science: by new auxiliary hypothesis or observation methods that can break the EE, or by non-entailed empirical evidence (grounded on intertheoretic connections of the theories involved) that can break the UD;

iv) these possible solutions do not count as a complete removal of the problem: that the EE or the UD will be thus eliminated is a contingent matter, there is no warrant that the development of science will be such that the problem gets solved in every case, so that recalcitrant UD as a result of EE is a possible scenario; and

v) the problem of EE and UD, when present, is a problem for *science* to solve, and, just as in any other scientific problem, that a solution will be found is not guaranteed from the outset.

Now that we are in possession of a precise and accurate evaluative description of the problem of EE and UD, we can turn to the analysis of the two case-studies announced.

CHAPTER 2

LORENTZ'S ETHER THEORY VS. SPECIAL RELATIVITY

The first case-study of EE and UD to be addressed is given by the rivalry between Einstein's theory of special relativity and Hendrik Antoon Lorentz's ether theory. This chapter is divided in five sections. In the first one I provide a schematic outline of the scientific context which motivated Lorentz to invent his ether theory. This overview helps to grasp a better understanding of Lorentz's scientific work, which I present in section two—from a chronological point of view, and paying special attention to Poincaré's amendments and contributions to the ether theory. The third section is devoted to a concise exposition of Einstein's special relativity. In the fourth section I show that the theories at issue are indeed predictively equivalent—but only if the crucial work of Poincaré is considered—and that the theories are different and contenders. In the final section I explain and evaluate the reasons that can be invoked in order to make a choice between them[1].

2.1 THE QUEST FOR THE ETHER[2]

The historical appraisal of the origin of Einstein's special relativity theory (SR) is a field in which very different interpretations have been provided. The 'textbook view' commonly suggests a close connection between Einstein's motivation to create his theory and what Tetu Hirosige (1976) labels as 'the ether problem'. More specifically, the fact that most of the textbook expositions of SR refer to the negative results of the Michelson-Morley experiment of 1887 suggests that Einstein's theory was the final solution to the ether problem by showing its superfluity. Beginning in the 1960's, historians of physics such as Hirosige, Holton, Schaffner and Miller compellingly argued that this is not an adequate historical claim. Einstein's motivations for SR were not intrinsically linked to the ether problem. However, they have also shown that in order to understand all of the relevant issues concerning the rise of SR, it is necessary to take a close look at the development of the ether problem. For example, this view allows one to clearly see that Lorentz's theory, in his 1904 version, was a satisfactory solution of the problem. From these general remarks it is obvious that Lorentz's theory is one of the final stages of the quest of the ether, so that an adequate historical and conceptual understanding of the former requires an examination of the latter, examination which I will now undertake.

2.1.1 Stellar aberration and the nature of light

During the 1720's James Bradley performed astronomical observations set out in order to find stellar parallax. Since the Earth changes its position along its translational motion, the distant stars should change their apparent position in the course of a year. This effect is a function of the ratio between the diameter of the translational motion of the Earth and the distance to the star considered. The last quantity is too big for the parallax effect to be detected by the experimental equipment available to Bradley. However, he observed another kind of systematic change in the apparent position of the star he was looking at. He noticed it could not be stellar parallax since the pattern of this change was a function of the velocity variations rather than positional shifts. More precisely, the effect he observed was proportional to the

[1] The main results of this chapter have been published in (Acuña 2014b).
[2] This section is based on (Hirosige 1976), (Schaffner 1972), (Darrigol 2005) and (Janssen & Stachel 2004).

ratio between the velocity of the Earth in its orbit around the Sun and the velocity of light, that is, proportional to v/c.

This effect was readily explainable in terms of the then prevailing particle-emission theory of light. If on a rainy windless day a person walks covered by an umbrella, the way in which she should hold it in order to stay dry depends on how fast she is walking, on the direction that she is moving, and on the velocity of the falling raindrops. More precisely, the apparent direction of the falling rain depends on the ratio between the person's velocity and the velocity of the raindrops –both velocities considered with respect to the Earth. Analogously, a telescope set out to look at a star has to be tilt, even if the star is right overhead, as an effect of the Earth's velocity along its orbit and the velocity of the light particles entering the telescope.

However, there was a sense in which Bradley's discovery put a challenge for the particle-emission theory of light. The calculations of the velocity of light based on stellar aberration were consistent with the measurements made by Ole Römer in 1670 based on the changes in the periods between successive eclipses of Jupiter's moon Io. He explained those changes in terms of the time it takes for the light to travel from Jupiter to the Earth. The consistency between the value that Römer obtained and the value derived from Bradley's observation of the aberration effect suggested that there is something like *the* velocity of light. This concept is problematic within a particle theory of light, for the measured velocity of the light-particles depends on the velocity of the source with respect to the measuring receiver. Moreover, there was no reason to think that an emitting object only emits streams of light particles with one single velocity. Therefore, the most natural assumption in a particle-emission theory was that light-particles coming from the distant stars should be received on Earth within a wide range of different velocities. From the point of view of a wave theory of light, however, the concept of *the* velocity of light was quite natural, for the velocity of waves only depends on the properties of the medium they move in, not in the state of motion of the emitting body –though the relative motion of the *receiver* of the wave with respect to the medium should affect the value of the velocity measured.

In 1810 Francois Arago tested this assumption. He covered half of a telescope with an achromatic prism to refract the light coming from a star, and aimed it to a star on different dates along a year in order to be sure that the velocity of the star with respect to the Earth was different every time due to the translational motion. He found no associated changes in the patterns of refraction –the rays and their refraction always respected Snell's law, what indicates that the velocity of the incoming light was always the same (with the angle of the light calculated with respect to the apparent position of the star, not with respect to its real position with the aberration effect corrected). Arago concluded, in order to save the phenomena from the view of a particle theory, that stars do emit light with different velocities, but that in order to be perceived by an observer, the ratio between the relative velocities of the source and the receiver must lie within a specific range. This explanation was considered even by Arago himself as highly implausible, so he looked for further opinions.

2.1.2 Fresnel vs. Stokes

Arago turned his attention to the wave theory of light to try to find a more suitable explanation for stellar aberration and the results obtained with respect to the velocity of light. By 1815, he knew that Augustine Fresnel was working in that field, so he encouraged him to develop an explanation. In 1818 Fresnel provided a theory based on two main assumptions. The first was the 'immobile ether hypothesis', i.e., that the Earth moves through a stationary ether without 'carrying' it along. This assumption offers a simple explanation in terms of a wave conception of light. If the Earth were to drag some amount of ether along its orbital motion, the consequent motion of the ether would affect the path of light coming from the distant stars, so that the stellar aberration effect would not be expected. But if the ether stays still in spite of the motion of the Earth across it, the light emitted by the stars would follow a rectilinear path, so

that the aberration effect follows quite naturally. This simple explanation in terms of a stationary ether had been already introduced by Thomas Young in 1804.

However, Arago's observations showed that the motion of a transparent dense medium did not affect the refraction pattern of light coming from a star when it enters this medium. In other words, his findings imply that the glass lenses of telescopes directed to a star do not alter the path of the incoming light, but it was known that transparent dense mediums, such as glass, refract light and alter its path in a specific angle. Therefore, the aberration effect should be affected depending on the state of motion of the Earth: the aberration pattern should be different if the incoming starlight entered the telescope in different phases of the Earth's translational motion. Arago's experiment showed that this alteration of the aberration effect did not occur. Consequently, Young's explanation only works if we suppose that the telescopes used are hollow. The assumption of the stationary ether was not enough by itself to explain stellar aberration from the perspective of a wave theory of light.

Fresnel's second assumption comes to solve this problem. It states that transparent dense mediums such as glass *drag* a part of the ether within them when moving across it. The glass of a telescope picks up a fraction of the light's velocity coming into it, so that the expected alteration of the aberration effect gets canceled. The quantitative expression for Fresnel's dragging coefficient f is $1 - {}^{1}/_{n^2}$, where n is the refraction index of the transparent medium.

The physical interpretation that Fresnel proposed for his dragging coefficient was that a moving transparent body, with a refraction index greater than 1, drags along the excess of ether inside it with respect to the density of the ether outside it:

> Following Young, Fresnel assumed that the ether density in a transparent medium was proportional to the square of the medium's index of refraction. For any classical wave, the speed of propagation is given by $\sqrt{T/\rho}$, where T is the tension and ρ is the density [of the medium]. If the tension is assumed to be constant, as Fresnel did, the velocity c/n is proportional to $1/\sqrt{\rho}$. Hence, $\rho \propto n^2$. Fresnel further assumed that, in optically dense media, only the ether density in excess of that pervading all space would be carried along by the medium. Let the density outside the medium be ρ and let the density inside be $\rho' = n^2\rho$. On average the ether inside the medium moving through the ether with velocity $\mathbf{v}$ will then move with velocity $\left(\frac{\rho'-\rho}{\rho'}\right)\mathbf{v} = \left(1 - \frac{\rho}{\rho'}\right)\mathbf{v} = \left(1 - \frac{1}{n^2}\right)\mathbf{v}$. (Janssen & Stachel 2004, 13-4)

The introduction of Fresnel's coefficient was very successful in explaining both stellar aberration and Arago's experiment from the standpoint of a wave theory of light. Moreover, it was also capable to explain further phenomena in the context of experiments performed with terrestrial sources of light. Without the coefficient, laboratory experiments on refraction should yield deviations of the order v/c with respect to Snell's law, and that deviation should be interpreted as a function of the motion of the Earth with respect to the ether. The introduction of Fresnel's coefficient precludes that deviation, and therefore, any possibility of detecting the motion of the Earth with respect to the ether by this kind of experiments. Many tests of this sort were carried out, and the results were always consistent with Fresnel's theory[3]. Yet another source of empirical support for Fresnel's coefficient was a prediction he made as early as 1818. The coefficient holds for any medium with a value for its refraction index n greater than 1. Therefore, if the observations of a star are made with a telescope filled with water, Fresnel's coefficient corresponding to water would cancel the effect of refraction of this element. That is, the water in the telescope should not affect the measured angle of aberration. In experiments carried out in the early 1870s, George Airy confirmed this prediction.

However, the physical interpretation provided by Fresnel himself was not quite satisfactory. Many objections were made against it. Maybe the simplest and deepest one was given by Willhelm Veltmann's

[3] See (Janssen & Stachel 2004, 12-3).

experimental results during the early 1870s. It was originally assumed that Fresnel's coefficient presupposed a refraction index n as referring to an average frequency of light, but Veltmann found out that the coefficient should be applied individually to each frequency. Since the index depends on the specific color-frequency of light, then in Fresnel's view transparent bodies should drag different amounts of excess-ether for every color. Objections like this determined the attitude of the scientific community towards Fresnel's theory. Its huge empirical success grounded the view that any optical theory committed to a stationary ether should include the coefficient. All refraction experiments showed that optical phenomena followed the same laws as if the Earth were still with respect to the ether—at least up to first order of v/c. Nevertheless, the true nature of the physical mechanism underlying Fresnel's coefficient was highly dubious. As we will later see, the claim for a satisfactory account had to wait until Lorentz's work.

Beyond the empirical success of Fresnel's theory, there was yet another problematic feature in it which led to the formulation of a rival theory. By the 1840s it was already known, by means of experiments showing polarization effects in light, that the ether should be an elastic solid medium of high rigidity. If light is a polarized wave it has to be a transverse one, and then its propagation medium cannot be a gas or a fluid, for these can only carry longitudinal waves. On the other hand, the extremely high value of the speed of light required the medium to possess a high degree of rigidity. Then, if the ether must have the features stated, how could it be conceived that a massive object such as the Earth moves across it without altering it at all?

George Gabriel Stokes, in 1845-6, proposed an alternative theory which was able to avoid this problem. The main assumption was simply that the Earth drags along the ether that surrounds it as it moves in its orbit. However, his theory demanded a complex explanation for the behavior of light approaching the Earth, for any motion of the ether should affect the path of light, and then the aberration effect could not be explained without further considerations. Following a hydrodynamic analogy, Stokes described the ether as behaving as a rigid solid for high-frequency waves –such as light– but as a fluid for relatively slow massive objects moving across it –such as the Earth. Stokes' ether was a sort of fluid with large viscosity. Being a fluid, it allows the Earth to move through it –but dragging the part which surrounds the planet. Its large viscosity makes it to behave just like a solid body with small shear elasticity and large plasticity. This feature explains, according to Stokes, why it is able to behave as a rigid solid with respect to high frequency waves. This description gives a more realistic model of the ether and of the nature of its interaction with the Earth, but in order to explain stellar aberration Stokes had to assume a complex description of what happens at the border between the immobile ether far from the Earth and the mobile ether which is dragged by the Earth's motion: the ether has to be an incompressible fluid in irrotational motion with a velocity potential with respect to the motion of the Earth, so that the bending of the wave fronts of starlight –when they cross the border between the immobile ether and the dragged one– produces the observed pattern of aberration:

> In Stokes' view, the ether was a jelly-like substance that behaved as an incompressible fluid under the slow motion of immersed bodies but had rigidity under the very fast vibrations implied in the propagation of light. In particular, he identified the motion of the ether around the earth with that of a perfect liquid. From Lagrange, he knew that the flow induced by a moving solid (starting from rest) in a perfect liquid is such that a potential exists for the velocity field. From his recent derivation of the Navier-Stokes equation, he also knew that this property was equivalent to the absence of instantaneous rotation of the fluid elements. Consequently, the propagation of light remains rectilinear in the flowing ether, and the apparent position of stars in the sky is that given by the usual theory of aberration. (Darrigol 2005, 4-5)

At this point it is relevant to pay attention to yet another empirical test of Fresnel's coefficient. I already mentioned that in most of these tests the coefficient operates as a canceling certain optical features which otherwise would be obtained. That is, its confirmation was mainly related to negative results. One im-

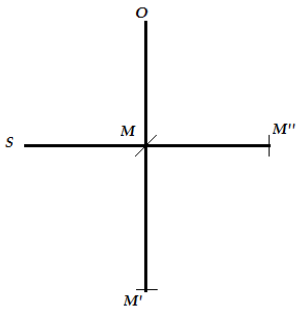
portant exception came along with Hypolite Fizeau's experiment of 1851. Its main objective was to measure the value of the speed of light in the laboratory, rather than by means of astronomical observations. The experiment consisted in a device made out of two connected tubes in which water was made to flow in opposite directions. Fizeau examined the effect of the flowing water on light that was made to pass through it: he wanted to find out what happens when light emitted from the same source was made to pass through water flowing in opposite directions. The effect he found was a shift on the interference pattern of the light rays after passing through the water, a shift whose value was quite consistent with what should be expected on the assumption of Fresnel's coefficient. Fizeau's experiment was thus considered as a more direct and successful test of the coefficient than the ones based on the null effect of the motion of the Earth through the ether in refraction experiments.

This test had effects on Stokes' theory. In order to account for Fizeau's experiment, it needed to include Fresnel's coefficient. But, as we saw, one of its main attractions was that it did not need it in order to explain stellar aberration and the absence of ether-wind effects in refraction experiments carried on terrestrial labs. There were no deviations from Snell's law because the ether surrounding the Earth was dragged, so that they were at rest with respect to each other. Now, in spite of the relative rest between the Earth and the surrounding ether Fresnel's coefficient had to be considered anyway. According to this, Fizeau interpreted his experiment as supporting Fresnel's immobile ether theory over Stokes'. In any case, the latter theory still had the advantage of providing a more reasonable account of the interaction between massive objects, light waves, and a solid ether.

2.1.3 The Michelson-Morley experiment

The situation ca. 1860 was then that of a hard competition between Fresnel's and Stokes' theories with respect to the problem of the ether. Fizeau's experiment turned the balance somewhat in Fresnel's favor. However, Stokes' theory had the attraction mentioned in the previous paragraph which compensated its complex explanation of stellar aberration. Besides, Fresnel's theory was empirically very successful, but the mechanism underlying the partial drag coefficient was quite unclear. Therefore, an experiment capable to decide between the theories in a more definitive way was to be most welcome. J. C. Maxwell, in the entry for *Ether* in an edition of the *Encyclopaedia Britannica*, made a suggestion for an experiment to measure the velocity of the Earth with respect to the ether in a terrestrial laboratory that consisted in looking for variations in the speed of light travelling back and forth between two mirrors. He noticed, however, that the related effects were too small to be measured, of the order of v^2/c^2, and the alternative method he suggested in order to obtain expected effects of first order of v/c required astronomical data about the periods between eclipses of Jupiter's moons along 12 years, data which by the time were not available with the precision required.

Albert Michelson took the challenge set by Maxwell's suggestion. He designed an 'interferometer', a device two 'arms' perpendicularly connected at M. In one of the ends of the arms, there is a source of light S, and in the opposite end M'' of the same arm there is a mirror. In the other arm, in one of the ends there is an 'observer' O which measures interference patterns produced by two light beams, whereas in the opposite end M' there is also a mirror. In M there is yet another mirror, a 'beam-splitter' placed at a suitable angle that partly reflects and partly transmits light. Finally, the distance $MM'' = MM' = l$. If a light beam is emitted in S and it is split in M, a reflected beam travels back and forth along MM', and another transmitted beam travels back and forth along MM''. The two beams meet again at M and are transmitted and travel together along MO, where the pattern of interference they create is measured:



Suppose that the ether is moving with respect to the interferometer with velocity v parallel to OMM' — or that the earth is moving with velocity v across the ether in the opposite direction, of course. In this case the time it takes for the beam traveling along MM', i.e., in the direction parallel to v, is $\frac{l}{c+v}+\frac{l}{c-v}=\frac{2lc}{c^2-v^2}\approx\frac{2l}{c}\left(1+\frac{v^2}{c^2}\right)$. In the case of the light beam travelling perpendicularly to v along MM'', its travel time is given by $\frac{2l}{\sqrt{c^2-v^2}}\approx\frac{2l}{c}\left(1+\frac{1}{2}\frac{v^2}{c^2}\right)$. From the comparison between the two expressions it follows that the time it takes the beam to travel in the parallel direction to the relative motion of the ether and the Earth $MM'M$ is larger than the speed of the beam traveling perpendicularly to that motion $MM''M$. The value of the time difference is approximately $\frac{l}{c}\frac{v^2}{c^2}$. This time difference multiplied by the frequency f of the light used gives the phase difference of the beams which determines the interference pattern measured in O. If the frequency f is expressed as c/λ — where λ is the wavelength — and is so multiplied by the time difference expression, one obtains $\frac{l}{\lambda}\frac{v^2}{c^2}$ for the phase difference. Even though the quantity v^2/c^2 is minute, the ratio l/λ between the length of the arms and the wavelength of the light used can be made very large. This is why Michelson's interferometer is able to measure an effect of second order.

Of course, the experiment cannot assume what is the direction of motion of the Earth across the ether. Moreover, only *changes* in the phase difference can be observed as changes in the interference pattern at O. For these reasons Michelson's interferometer was designed to be rotated. By rotating it in 90°, the roles of the arms get inverted, so that the change in phase difference and the corresponding interference pattern to be observed is twice the amount given in the expression above.

One final important remark about the design of the experiment is that Michelson made a considerable mistake. He calculated a time for the travel of the beam perpendicular to v of $2l/c$, just as if the interferometer were at rest with respect to the ether, instead of $\frac{2l}{\sqrt{c^2-v^2}}$, which is the value that does consider the relative motion between the interferometer and the ether. The result of this mistake was that he miscalculated the time difference between the two trips by a factor of 2, and this overestimation reflected in the value of the interference pattern shift he expected.

Michelson carried out the experiment in Potsdam in 1881. The result he observed was by far within the range of expected disturbances due to the ambient. Michelson himself interpreted it as a refutation of a theory of an immobile ether, for no effect of the motion of Earth was detected, even in the order of v^2/c^2. He also suggested that his experiment could be interpreted as a crucial one and favoring Stokes over Fresnel.

At this point is where Hendrik Antoon Lorentz gets involved in the quest for the ether. In 1886 he published a paper in which he deeply and compellingly criticized the foundations of Stokes' theory. He showed that the assumption of an ether in irrotational motion and the assumption that the ether surrounding the Earth is fully dragged are inconsistent, and he proposed a theory that mixed elements of both Fresnel's and Stokes':

> He made the following assumptions: first, that the ether surrounding the earth is in motion and that this ether has a velocity potential; second, that the motions of the ether and the earth can be different from each other at the earth's surface; third, that when the ether moves through a transparent body, the elementary waves of light in this body are dragged along the direction of the relative motion of the body with respect to the ether with the velocity kv. Here v denotes the relative velocity of the body to the ether, and $k = 1 - 1/n^2$, n being the refractive index of the body. Finally, Lorentz made no assumptions about opaque bodies. With these assumptions and neglecting terms higher than the first order of v/c, Lorentz examined the path of light rays with regard to the earth—the relative rays, as he called them—to show that all phenomena occur as if the earth were at rest and the relative rays followed the path of light rays with regard to the ether, that is, the path of the absolute rays. In other words; except for the Doppler effect, there is no detectable effect of the motion of the earth upon optical phenomena […].
> [Lorentz claims that] If we regard the atoms of matter as a local modification of the ether, we may expect that the ether freely penetrates material bodies however thick they might be. Lorentz considered this problem so important that he urged physicists not to be content with considerations of probability or simplicity, but to decide on the basis of experiment whether the ether at the surface of the earth is at rest or in motion. (Hirosige 1976, 26-7)

Hirosige's summary of Lorentz's 1886 theory (or better, Lorentz's *sketch* of a theory) illuminates some important features. First, it contained an account of Fresnel's coefficient in terms of light rays being carried rather than excess-ether, Lorentz's view provided a new rationale for the physical process underlying the coefficient. I will later show how this was done in terms of electromagnetic considerations. Second, assumptions 1 and 2 imply that Lorentz's definitive response to the question of the ultimate state of motion of the ether remains open and waiting for empirical testing—even though he favored an immobile ether theory—but at the same time the theory offers an account of the absence of observed effects related to this issue. Finally, he offers a rationale for the problem that motivated Stokes' theory, namely, the interaction between massive bodies and the ether. If atoms are considered as a sort of state of the ether, then it is quite natural to suppose that they will move through it without disturbing it (in terms of motion). In this view one can find a seed of the 'electromagnetic view of nature' that Lorentz later endorsed.

Lorentz also put special attention on Michelson's mistake. He stated that the 1881 experiment could not at all be considered as refuting Fresnel and supporting Stokes. Therefore, his 1886 work operated as one of the motivations to repeat the interferometer experiment with increased accuracy. He expected that it would finally show the empirical success of Fresnel's view, fulfilling his will to have an empirically based decision on the state of motion of the ether.

Alfred Potier also drew attention on Michelson's mistake, so that the inconclusiveness of his experiment became blatantly apparent. On the other hand, Lord Rayleigh and William Thomson encouraged Michelson to repeat it, but this time first performing Fizeau's experiment with a higher degree of accuracy. He followed the advice and in 1886, in collaboration with Edward Morley, carried out an improved version of it. The results they obtained strongly confirmed Fresnel's coefficient, and they even took it as a confirmation of the immobile ether hypothesis –the opposite conclusion to the one Michelson had obtained in 1881.

Their next step was, of course, to repeat the interferometer experiment. This time they got the right calculations and designed a much more sensitive and reliable device: it was capable to be rotated in a much smoother way and the light beams were sent back and forth many times along their paths, so that the ratio between l and λ got largely augmented and the expected shift in the interference pattern to be measured increased tenfold. Once again, the result was negative. They repeated the experiment some months later in order to discard the almost fantastic possibility that at the first time the overall velocity of the Earth with respect to the ether had been quite small. But the result was negative as well.

The resulting situation was thus quite dramatic. Stokes' theory had been severely undermined by Lorentz's criticism. Moreover, also in 1887, Hertz succeeded in detecting the electromagnetic waves predicted by Maxwell. This discovery led to the inclusion of optics into electrodynamics, and it turned out that it was very difficult –if possible at all– to incorporate any ether drag in Maxwell's theory and at the

same time to have an account of phenomena such as aberration and Fizeau's effect. On the other hand, Fizeau's experiment as carried out by Michelson and Morley strongly confirmed the reality of Fresnel's coefficient, but its physical explanation could not be that of an inner partial ether drag, because of the very same reasons just outlined. Furthermore, the immobile ether thesis to which Fresnel's theory was committed was deeply threatened by the negative result of the Michelson-Morley experiment. Hence, none of the two available alternatives was able to successfully face the radical problem of the ether.

The depth and difficulty of the problem immediately underscores how important the solution that Lorentz later provided was. Totally committed to the unification of optics and electromagnetism that Maxwell's electrodynamics brought, he faced the task of creating a theory under the assumption of an immobile ether capable to offer an account for all the negative results of the experiments so far performed in order to measure the motion of the Earth through the ether. I now turn to this subject.

2.2 LORENTZ'S THEORY[4]

What at the time was called Lorent'z *Theory of the Electron*, was the outcome of a scientific enterprise that Lorentz started in 1892 (but with its basic roots settled in 1886), and finished in 1904—though he made later important remarks and revisions until 1916. Therefore, different and progressive stages of its development can be distinguished, and this distinction is quite useful in order to understand his work in a deeper and more accurate way. The stages which I will differentiate in this work (closely following the 'standard view' of the historians who have written about the subject) are two: from the seeds of the theory of 1886 up to the *Versuch* of 1895; and from the formulation of what Janssen calls the 'generalized contraction hypothesis' in 1899 up to its definitive inclusion as a part of the theory in 1904. In between both periods, and after the second, important criticisms and reinterpretations introduced by Henri Poincaré must be considered if one is to consider Lorentz's work as predictively equivalent to Einstein's SR.

2.2.1 Stage one: 1886-1895

Hendrik Lorentz's first major scientific work was his doctoral dissertation of 1875. In it he tackled a problem of optics which was first acknowledged by Helmholtz in 1870. Once the luminiferous ether had been depicted as an elastic solid medium, and under Maxwell's analogy between motions in a dielectric and motions in the ether, Helmholtz noticed that the assumption of an ether with those properties implied that, at the limit between two transparent media, the boundary conditions needed to explain reflection and refraction of light were inconsistent to each other. Fresnel's theory, for example, gave the correct formulas at the price of overlooking this problem. Lorentz attempted the task of solving this difficulty from a point of view in which he flirted with the 'action at a distance' approach for charges and currents that constituted the mainstream view in continental Europe at the time. The relevance of this early work is that in it, and in his following published paper of 1878 where he dealt with an electromagnetic explanation of dispersion, the roots of the basic ontology of his ether theory got settled: the divorce of ether from matter. That is, since the very beginning of his career, Lorentz was committed to a dualist ontology in which the optic (and electromagnetic) ether was a substance of an essentially different kind than 'regular' matter[5].

After 1878 Lorentz turned to problems of kinetic theory and thermodynamics, but in 1886 he returned to electrodynamics and optic issues. As I mentioned above, in that year he published his *On the Influence*

[4] My presentation of Lorentz's theory is based mainly on (Miller 1998) and (Janssen 1995).
[5] For Lorentz's concept of a purely electromagnetic ether, see (Nersessiann 1984) and (McCormmach 1970b). For Lorentz's work before 1886, see (Darrigol 1994).

of the motion of the Earth on Light Phenomena—published in Dutch and the following year in French—where he made two very important remarks: that Stokes' theory was founded on inconsistent assumptions, and that Michelson's mistake of 1881 made the experiment completely inconclusive. Therefore, he concluded that the latter was not at all a reliable source of empirical support for Stokes' theory. In that same work he sketched the outlines of a hybrid theory which assumed features both form Stokes' and Fresnel's. However, it was clear that he favored a plan for a definitive theory in which the ether was immobile and fully transparent to the motion of massive bodies: "It seems to me that the latter view [immobile transparent ether] is at least as simple as the former, if not simpler. It may be that what we call an atom is nothing but a modification of the state of this medium; then one could understand that an atom could move without dragging the ether" (from Lorentz 1886, quoted in Darrigol 1994, 274-5).

Notice that in the last quote a subtle twist in Lorentz basic ontological view is contained. Ether and matter are divorced entities; however, when he writes that *an atom is a modification of the ether*. Here one can already recognize his commitment to an electromagnetic view of nature: the hypothesis that the ultimate nature of reality is electromagnetic. Charges, the ether and electromagnetic forces are the main constituents of physical reality and from them mechanical features emerge.

Yet another relevant feature in Lorentz's publication of 1886 was the derivation of Fresnel's coefficient and a more satisfactory rationale for its underlying physical mechanism. He claimed that the coefficient was not connected to an ether drag, rather, it was the outcome of the interaction between the molecules of the transparent body in which the light entered and the ether surrounding them. That is, his explanation was purely electromagnetic and permitted a conception in which the ether is completely immobile: there is no theoretical necessity for any excess-ether drag or of any kind of partial drag—even though Lorentz's final position about the state of motion of the ether was still open in 1886[6].

As it can be noticed, Lorentz work of 1886 was a sort of first draft of a consistent theory that was able to account for up to first order optical phenomena—mostly 'negative' ones—which merged elements both from Fresnel and Stokes. An essential question that he was not yet able to conclusively answer was that of the true state of motion of the ether: 'to what degree the ether participates in the motion of bodies that traverse it … is of interest not only for the theory of light. It has acquired a more general importance since the ether probably plays a role in electric and magnetic phenomena' (from Lorentz 1886, quoted in Miller 1998, 18). In spite of his clear sympathy for a completely immobile and transparent ether, which is apparent in his new explanation of Fresnel's coefficient, he considered that this issue was still open and needed to be decided on the basis of empirical data: 'In my opinion we cannot permit ourselves to be guided in such an important problem by considerations concerning the degree of probability or simplicity of one hypothesis or the other, but to address ourselves to experiment in order to ascertain the state of rest or motion of the ether at the earth's surface' (ibid, 21-3).

---

[6] Arthur Miller offers an explanation of the matter and a very interesting remark connected to Einstein: "Lorentz (1886) used Huygens' principle [Miller depicts a nice figure explaining the operation of Huygens' principle in Lorentz's reasoning] and Fresnel's hypothesis to deduce the velocity u_r of light that traversed a medium of refractive index N that was at rest on the earth as $\boldsymbol{u}_r = \boldsymbol{c}/_N - \boldsymbol{v}/_{N^2}$ (Eq. 1.17), where the source could have been either on the earth or in the ether. For N=1, Eq. (1.17) reduced to Eq. (1.13) [$\boldsymbol{u}_r = \boldsymbol{c} - \boldsymbol{v}$, which is the regular explanation for aberration without considering any refraction of light, with $\boldsymbol{c}$ being the velocity of light and $\boldsymbol{v}$ that of the Earth], and for $N \neq 1$, Eq. (1.17) explained Arago's experiment and an equivalent one by George Bidell Airy. Lorentz (1886) continued by noting that from the viewpoint of the geocentric system we could say that 'the waves are entrained by the ether' according to the amount $-\mathbf{v}/N^2$ […]. On the other hand, an observer at rest in the ether measured the velocity of the light that was propagating through the medium at rest on the moving earth to be $\boldsymbol{c}' = \boldsymbol{u}_r + \boldsymbol{v}$ (1.18). Lorentz (1886) noted that the ether-fixed observer could interpret Eq. (1.18) as the 'entrainment of the light waves by the ponderable matter'. Consequently, although the phenomenon of stellar aberration depended on only the relative velocity between the earth and the star, ether-based theories of optics described it in two different ways depending on whether the source or observer were in motion. […] Einstein considered redundancies of this sort as 'asymmetries which do not appear to be inherent in the phenomena'. In summary Fresnel's hypothesis of a dragging coefficient explained: 1) the dependence of the velocity of light on the velocity of the medium through which it propagated; 2) to first order in v/c, where v was the earth's velocity relative to the ether; optical phenomena were unaffected by the earth's motion". (Miller 1998, 18-20)

The last two quotes clearly explain the motivation of Lorentz's subsequent work. From the first one, one can see that his approach will be that of Maxwell's unification of optics and electromagnetism. From the second one, one can see that the negative result of the Michelson-Morley experiment implied a huge new challenge for the project of an electrodynamical theory able to explain all of the observations regarding the relative motion of the Earth and the ether: the theory should also be able to explain negative results for an experiment of the second order of v/c.

In 1892 Lorentz published two works where he attempted both tasks. The first, *The Electromagnetic Theory of Maxwell and its Application to Moving Bodies*—originally in French—was a big study on Maxwell's theory that included a new term—he later dubbed it as 'local time'—that enabled him to predict negative results for *any* kind of experiments to measure the relative motion between the Earth and the ether, not only for refraction effects tests—up to first order of v/c, though. The negative result of the Michelson-Morley experiment was of course out of the scope of this explanation. Lorentz coped with the special case of second order experiments in a paper entitled *The Relative Motion of the Earth and the Ether*. In it he introduced yet another concept, the hypothesis of a 'length contraction' of bodies moving across the ether that precluded the measurement of second order effects of ether-wind. Later on, in 1895, he published his famous *Attempt of a Theory of Electric and Optic Phenomena in Moving Bodies*—in German—commonly known as the *Versuch*. In it he presented in a more systematic way the results of both the mentioned works of 1892. For simplicity, and following Janssen 1995, I will immediately refer to the *Versuch* as the next step in the first stage of Lorentz's theory.

The basic framework on which Lorentz developed his theory in the *Versuch*, and also in his previous work of 1892 on Maxwell's theory, was the assumption that the sources of electromagnetic disturbances in the immobile, non-mechanical, and purely electromagnetic ether, were microscopic charged particles able to freely move through it. On this assumption he presented the following Maxwell equations for the electric field ***E*** and the magnetic field ***B***:

$$\mathrm{div}_0\mathbf{E} = \rho/\varepsilon 0, \qquad \mathrm{div}_0\mathbf{B} = 0, \qquad \mathrm{curl}_0\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t_0}, \qquad \mathrm{curl}_0\mathbf{B} = \mu_0\rho\mathbf{u}_0 + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t_0};$$

where $\mathbf{E}$, $\mathbf{B}$, the charge density ρ, and the current density $\rho\mathbf{u}_0$, are all quantities that are functions of the spatial and time coordinates $(\mathbf{x}_0, t_0)$ of a reference system S_0 which is *at rest with respect to the ether* (the subscript 0 in $\mathbf{u}_0$ indicates that $\mathbf{u}$ is a velocity with respect to S_0). Then Lorentz shows that if the Galilean transformations ($\mathbf{x} = \mathbf{x}_0 - \mathbf{v}t_0$; $t = t_0$) are applied to these equations in order to obtain the ones that hold for a system S *in motion with respect to the ether*, then the formulas for $\mathbf{E}$ and $\mathbf{B}$ in S become:

$$\mathrm{div}\,\mathbf{E} = \rho/\varepsilon 0, \qquad \mathrm{div}\,\mathbf{B} = 0, \qquad \mathrm{curl}\,\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t_0} + v\frac{\partial \mathbf{B}}{\partial x}, \qquad \mathrm{curl}\,\mathbf{B} = \mu_0\rho(\mathbf{u}+\mathbf{v}) + \frac{1}{c^2}\left(\frac{\partial \mathbf{E}}{\partial t} - v\frac{\partial \mathbf{E}}{\partial x}\right)^{7}$$

It is quite apparent that these field equations, valid in a moving frame, are not Maxwell's equations for a frame at rest in the ether. This is just another way of saying that the motion of the Earth across the ether should yield observable effects. We saw that Fresnel's coefficient was an explanation for the non-existence of these effects, but only in the case of refraction phenomena, and up to first order of v/c. This is the context and motivation on which Lorentz introduces his famous auxiliary quantity of 'local time'. With the help of this *mathematical tool*—for he did not assign any physical meaning to it—he was able to introduce a new set of transformations for the system in motion with respect to the ether, such that the resulting equations have the same form of Maxwell's equations for the ether-rest frame—if terms of second

[7] The time derivative $\partial/\partial t_0$ of the first set of equations has been replaced by the differential operator $\partial/\partial t - v\,\partial/\partial x$, and in which the velocity $\mathbf{u}_0$ of the first set of equations has been replaced by $\mathbf{u} + \mathbf{v}$, where $\mathbf{u}$ is a velocity with respect to S—and with v being the velocity of S with respect to S_0 in both replacements.

and higher orders of v/c are neglected. That is, the transformations are such that Maxwell equations become *Lorentz-invariant*.

It is important to remark that Lorentz line of thought involves three different reference frames: S_0, in which Maxwell equations hold; S, whose field equations are not Maxwell's; and S', the *auxiliary* frame in which the equations obtained through the coordinate transformations introduced by Lorentz hold. This point illustrates more clearly that Lorentz did not assign any kind of physical meaning to his transformations. S' is an auxiliary frame which does not reflect measured quantities. Considering this remark, the structure of Lorentz reasoning is thus: on the field equations valid for S_0, Galilean transformations are applied so that the equations for S are obtained; and then Lorentz transformations are applied to the latter so that the Maxwell's Lorentz-invariant equations (up to first order) valid for the frame S' are obtained. The quantitative expression of the transformations is

$$\mathbf{x}' = \mathbf{x} = \mathbf{x}_0 - \mathbf{v}t_0, \qquad t' = t - (v/c^2)x = t_0 - (v/c^2)x \ [8]$$

As Janssen points out, from a modern point of view the derivation of the field equations for S' is simply a part of a proof that, to first order, Maxwell's equations are invariant under the transformation that Lorentz obtained. From the modern perspective, the quantities $\mathbf{x}'$, t', $\boldsymbol{E}'$, and $\boldsymbol{B}'$; belong to the *Lorentzian* frame S' moving with velocity v with respect to S_0, just as the corresponding unprimed quantities belong to the *Galilean* frame S which moves with velocity v with respect to S_0.

Armed with his 'local time' and the new transformations it permits, Lorentz formulates his famous *theorem of corresponding states*, which explicitly states that the transformations constitute a general proof that, to fist order, no effects of the relative motion among the Earth and the ether will be observed. That is, Lorentz 1895 version of the theorem provides a first order solution for the problem of the ether:

> If there is a solution of the source free Maxwell equations in which the real field **E** and **B** are certain functions of $\mathbf{x}_0$ and t_0, the coordinates of S_0 and the real Newtonian time, then, if we ignore terms of order v^2/c^2 and smaller, there is another solution of the source free Maxwell equations in which the fictitious field **E**' and **B**' are those same functions of $\mathbf{x}'$ and t', the coordinates of S and the local time in S.[9]

Since from a modern standpoint, as Janssen points out, what Lorentz did in 1895 is understood in a quasi-relativistic way, one should be careful and subtle about the differences between Lorentz's and the modern conception of the matter. Two remarks show that Lorentz's work was not at all a relativistic theory in the modern sense. First, the fact that the frame S' in which the Lorentz-invariant equations hold is *auxiliary* implies that the explanation of the negative results of the experiments which aim to measure the effects of the relative motion between the Earth and the ether cannot be given in terms of the *measured quantities* in S'. S' is not a physically *real* frame; it results as an auxiliary one from the application of the transformations that, in turn, are mathematical tools. For example, many experiments of the mentioned kind are based on observation of interference patterns—in how they change or in the fact that they do not. The explanation that the Lorentz transformations give, Lorentz-like interpreted, is not that in the moving frame the measurements of the time for the trips of light beams which produce the interference patterns will be the same as the corresponding measurements in the ether-rest frame; but that the *structure* of the patterns in both frames are the same: if at a certain place in one of them there is darkness in the pattern, there will be darkness in the corresponding state of the other frame. In other words, the fact that the theorem of corresponding states implies that *the patterns of light and darkness in the corresponding*

[8] Janssen (1995, 3.1.1), points out that Lorentz also introduced the fictitious fields $\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}$, and $\mathbf{B}' = \mathbf{B} - \frac{1}{c^2}\mathbf{v} \times \mathbf{E}$. The adjective *fictitious* is yet another indication that he considered his new transformation as a mathematical tool devoid of any physical meaning.

[9] This is Janssen's paraphrase of Lorentz's formulation (1995, section 3.1.1).

frames will be the same is not equivalent to the statement that *the value of each of the measured quantities involved will be the same*. The difference in meaning between these two statements indicates the nature of the difference between the Lorentzian and the relativistic approach:

> Consider some field configuration In S_0 and its corresponding state in S. according to the theorem of corresponding states, the same functions that give the real fields **E** and **B** as a function of the real coordinates $\mathbf{x}_0$ of S_0 and the real time t_0 for the configuration in S_0 will give the fictitious fields **E'** and **B'** as a function of the coordinates $\mathbf{x} = \mathbf{x}'$ of S and the local time t' for the corresponding state of that configuration in S. Suppose the configuration in S_0 is such that at a point P with coordinates $\mathbf{x}_0 = \mathbf{a}$ it is dark. That means that the fields **E** and **B** vanish at this point, not just at one instant, but over a stretch of time that is long compared to the period of the light waves described by the fields **E** and **B**. It follows that the fictitious fields **E'** and **B'** [see note 14] will vanish at **x=a** in the corresponding state in *S*. Since the relation between the real and the fictitious field is linear, this means that the real fields **E** and **B** in the corresponding state will also vanish at $\mathbf{x} = \mathbf{a}$. It follows that the patterns of light and darkness in the moving frame and the patterns of light and darkness in the frame at rest are the same. (Janssen 1995, 3.1.2)[10]

The modern reader might complain that the relativity of simultaneity plays a role in the issue and that it defies the possibility of this Lorenztian explanation. It is true that, in a relativistic explanation, the relativity of simultaneity underlies the possibility of an explanation in terms of the identity of the measured quantities involved. However, the patterns of light and darkness are such that they are meaningful from the perspective of time periods which are large compared to the periods of the waves used. Therefore, the fact that the sets of simultaneous events in both frames are different is harmless for the Lorentzian explanation[11].

The second important remark about the difference between the relativistic and the Lorentzian standpoint is that Lorentz did not conceive his transformations as symmetric operators. That is, to 'go back' from S' to S_0, the transformation to be applied is not a symmetrical Lorentz transformation, but its 'inverse function'. This is yet another indication that the real measured quantities belong to the system S_0, and that the coordinates and fields in S' are mere auxiliary quantities which are the outcome of the application of a mathematical tool. I will show below that it was Poincaré, and of course Einstein, who introduced the right relativistic interpretation.

Coming back to Lorentz theory itself, we have that yet another feature which illustrates its big importance, in the context of unified electrodynamics and the problem of the ether, is given by the formal and general derivation of Fresnel's coefficient. Consider a medium with refractive index n at rest in the ether in which a plane wave propagates with velocity c/n along the x-axis of a frame which is also at rest with respect to the ether. The components of the fields which describe this wave depend on x and t through the expression governing the phase of the wave $t - \frac{x}{c/n}$. Therefore, in its corresponding state, i.e., in a system in motion across the ether with velocity v in the x-direction, the components of the auxiliary

[10] This is the only place in the literature that I went through in which this subtle and important remark is underlined. Janssen refers to brief hints of it in (McCormmach 1970b, 471) and in (Darrigol 1994, 288). In the first case it is just a sentence that can be so interpreted after acknowledging the issue. In the case of Darrigol, he gives an explanation of the null results of the experiments which is quite similar to the one that Janssen offers. However, he does not stress the subtle but important difference with respect to a relativistic point of view.

[11] 'the stationary nature of patterns of light and darkness plays a crucial role in this argument. Without this property, the x-dependence of local time would lead to serious complications. Suppose that in two points P_0 and Q_0 of S_0, the fields vanish with respect to the real Newtonian time. In the corresponding points P and Q of the moving frame S, the field then will vanish simultaneously with respect to the local time. Since the local time depends on x, this means that they will *not* vanish simultaneously with respect to the real Newtonian time. This would invalidate Lorentz's conclusion with regard to patterns of light and darkness. Fortunately, patterns of light and darkness, by their very nature, are stationary situations. The concepts of light and darkness only have meaning on time scales that are large compared to the periods of the light waves used. So, when at P and Q it is dark at the same instant in local time, it will also be dark at both points at the same instant in real time' (Janssen 1995, 3.1.2).

fields which constitute the wave depend on the expression $t' - \frac{x}{c/n}$. Considering the Lorentz transformation for time $t' = t - (v/c^2)x$, then the auxiliary fields depend on t via:

$$t - (v/c^2)x - \frac{x}{c/n} = t - (v/c^2 + n/c)x.$$

From this expression it can be inferred that the velocity of the wave with respect to the moving medium is:

$$\frac{1}{\frac{v}{c^2} + \frac{n}{c}} = \frac{\frac{c}{n}}{\frac{v}{nc} + 1} \approx \frac{c}{n}\left(1 - \frac{v}{nc}\right) = \frac{c}{n} - \frac{v}{n^2},$$

and in order to obtain the formula for the velocity of the wave, *with respect to the rest frame in the ether*, the velocity v of the moving frame must be added, so that:

$$\frac{c}{n} - \frac{v}{n^2} + v = \frac{c}{n} + v\left(1 - \frac{1}{n^2}\right)$$

The velocity of the wave with respect to the ether is then $c/n + v\left(1 - 1/n^2\right)$, in agreement with Fresnel's coefficient $\left(1 - 1/n^2\right)$ [12]. It is important to compare this derivation with the electromagnetic treatment that Lorentz gave to the issue in 1886. This time the derivation does not refer to any specific electromagnetic assumption, and this is why it can be considered as more general. The crucial factor is now of course *local time*. That the Fresnel coefficient can be obtained in this purely mathematical way from the Lorentz transformations is yet another case in which it is apparent that the theorem of corresponding states is a fundamental (first order) solution for the problem of the ether. The fact that refractive phenomena will not produce any observable features revealing the relative motion of the Earth and the ether is just a specific consequence of the theorem. This remark clearly indicates the generality and unification-power of the theory that Lorentz was attempting. Once again, and in modern terms, it was a theory whose aim was to obtain Lorentz-invariance for Maxwell equations. Lorentz's *Versuch* of a theory of 1895 achieved this goal up to first order of v/c.

Unfortunately for Lorentz, the Michelson-Morley experiment was designed to measure effects of second order. Therefore, it was out of the scope of his theorem of corresponding states of 1895. That part of his theory was incapable of providing an explanation of it. With this in mind[13], Lorentz returned in the *Versuch* to the length contraction hypothesis that he had introduced in 1892. We already saw in the analysis of the Michelson-Morley experiment that the time required for a light ray to travel back and forth along one of the arms of the interferometer in the direction parallel to the direction of its motion through

[12] This derivation is from (Janssen & Stachel 2004, 25). Janssen also offers a slightly different one in his (1995, section 3.1.3), along with a derivation of the classical expression for the Doppler effect and of the classic formula for the aberration effect.

[13] The following quote clearly illustrates Lorentz's concern about the issue: "Fresnel's hypothesis, taken conjointly with his coefficient $1 - 1/N^2$, would serve admirably to account for all the observed phenomena were it not for the interferential experiment of Mr. Michelson, which has, as you know, been repeated after I published my remarks on its original form, and which seems decidedly to contradict Fresnel's views. I am totally at a loss to clear away this contradiction, and yet I believe if we were to abandon Fresnel's theory, we should have no adequate theory at all, the conditions which Mr. Stokes has imposed on the movement of aether being irreconcilable to each other. Can there be some point in the theory of Mr. Michelson's experiment which has as yet been overlooked?" From a letter to Lord Rayleigh dated August 18, 1892 –shortly after finishing *The Electromagnetic Theory of Maxwell and its Application to Moving Bodies*. Quoted in (Miller 1998, 27-8).

the ether is $\frac{l}{c+v}+\frac{l}{c-v}=\frac{2lc}{c^2-v^2}$, whereas the corresponding time required for the ray traveling perpendicularly to the interferometer's motion through the ether is $\frac{2l}{\sqrt{c^2-v^2}}$, and the different value of these expressions yielded a change in the interference pattern produced as the device was rotated. Lorentz assumed that as a body moves through the ether it gets contracted by a factor $\sqrt{1-v^2/c^2}$, so that in the first expression for the travel-time of the light ray the length l becomes $l\sqrt{1-v^2/c^2}$, and therefore the whole expression becomes $\frac{l\sqrt{1-v^2/c^2}}{c+v}+\frac{l\sqrt{1-v^2/c^2}}{c-v}=\frac{2l\sqrt{c^2-v^2}}{c^2-v^2}=\frac{2l}{\sqrt{c^2-v^2}}$. It is clear that Lorentz's contraction hypothesis implies that the travel time for both rays is the same, and in this case the Michelson-Morley experiment yields a null result.

This hypothesis was clearly introduced in order to account for one particular experiment, but at least Lorentz provided an argument to make it plausible. In 1892 he had already shown that the electromagnetic forces **F'** in a frame in motion with respect to the ether and the electromagnetic forces in a rest system with respect to the ether are related in the following way:

$$\mathbf{F}'_{x'}=\mathbf{F}_x \qquad \mathbf{F}'_{y'}=\frac{\mathbf{F}_y}{\sqrt{1-v^2/c^2}} \qquad \mathbf{F}'_{z'}=\frac{\mathbf{F}_z}{\sqrt{1-v^2/c^2}};^{14}$$

where the primed coordinates belong to the moving frame across the ether and the unprimed ones belong to the frame at rest in the ether. This result can be interpreted as stating that if a system at rest in the ether is in equilibrium under a configurations of forces **F**, then that same system, if in motion across the ether, is in equilibrium under a configuration of forces **F'**.

He then assumed that what he called 'molecular forces' determine the shape and length of a body, and that these forces act by intervention of the ether; but he also stated that the nature of molecular forces was totally unknown, so that his assumption was not directly assessable. However, if it is assumed that the molecular forces behave just as the electromagnetic forces do, then the length-contraction obtains. That is, if a system at rest in the ether is in its 'equilibrium shape' under a configuration of molecular forces $\mathbf{F}_m$, then the same system when in motion across the ether will be in 'equilibrium shape' under a configuration of molecular forces $\mathbf{F'}_m$. And if the transformation from $\mathbf{F}_m$ to $\mathbf{F'}_m$ is the same as the transformation for electromagnetic forces, then when at motion in the ether the system gets contracted[15]. Lo-

[14] Actually, Lorentz's expression was $\mathbf{F}'_{y'}=\mathbf{F}_y(1+p^2/2V^2)$, where $p=v$ and $V=c$; and to first order of v/c, $(1+p^2/2V^2)\cong 1/\sqrt{1-v^2/c^2}$. The same holds for $\mathbf{F}'_{z'}$, of course. A review of how Lorentz derived this result is presented in (Janssen 1995, section 3.2.5). See also (Miller 1998, 26-9).

[15] 'Let A be a system of material points carrying certain electric charges and at rest with respect to the ether; B the system of the same points while moving in the direction of the x-axis with the common velocity p through the ether. From the equations developed by me, one can deduce which forces the particles in system B exert on one another. The simplest way to do this is to introduce still a third system C, which just as A, is at rest but differs from the latter as regards the location of the points. System C, namely, can be obtained from a system A by a simple extension by which all dimensions in the direction of the x-axis are multiplied by the factor $(1+p^2/2V^2)$ and all dimensions perpendicular to it remain unaltered. Now the connection between the forces in B and C amounts to this, that the x-components in C are equal to those in B whereas the components at right angles to the x-axis are $1+p^2/2V^2$ times larger than in B.
We will apply this to molecular forces. Let us imagine a solid body to be a system of material points kept in equilibrium by their mutual attractions and repulsions and let system B represent such a body whilst moving through the ether. The forces acting on any of the material points of B must in that case neutralize. From the above, it follows that the same cannot then be the case for the system A whereas for system C it can; for even though a transition from B to C is accompanied by a change in all forces at right angles to the axis, this cannot disturb the equilibrium, because they are all changed in the same proportion. In this way it appears that if B represents the state of equilibrium of the body during a shift through the ether then C must be the state of equilibrium when there is no shift. But the dimensions of B in the direction of the x-axis are $(1-p^2/2V^2)$ times the corresponding dimensions of C whereas the dimensions along right angles to the x-axis are the same in both systems. One obtains, therefore, exactly an influence of the motion on the dimensions equal to the one in which, as appeared above, is required to explain Michelson's experiment' (from *The Relative Motion of the Earth and the Ether* (1892), quoted in Janssen 1995, section 3.2.6).

rentz did not offer this reasoning as a *proof* of the contraction hypothesis, but only as a plausibility argument for it: 'one may not of course attach much importance to this result; the application to molecular forces of what was found to hold for electric forces is too venturesome for that' (from *The Relative Motion of the Earth and the Ether*, quoted in Miller 1998, 29).

The hypothesis of length contraction as an explanation for the null result of the Michelson-Morley experiment had been already introduced by G. F. Fitzgerald in 1889. It is interesting to take a look at the way in which he justified the hypothesis. The plausibility argument that Fitzgerald offered was based upon Oliver Heaviside's 1888 discovery that the electromagnetic field around a moving charge gets shrunk along its direction of motion. The quantitative expressions that he determined for this effect were:

$$\mathbf{E} = \frac{q}{r^2}\frac{(1-v^2/c^2)}{(1-v^2/c^2\,sin^2\theta)^{3/2}} \qquad\qquad \mathbf{H} = \mathbf{E}v\sin\theta,$$

where **E** is the electric field directed radially outward of the charge, **H** the magnetic field in circles centered around the line of motion, q is the charge, v is the velocity through the ether, c the speed of light, r the distance of the charge to a point, and θ the angle to the line of motion. With respect to this result Hunt comments:

> Note especially the $(1 - v^2/c^2)$ factor and the way the field lines bunch up around the 'equator' as the speed increases. This compressed field is in fact the same as the Fitzgerald-Lorentz contraction of the electrostatic field of a charge at rest, in exact accordance with Einstein's theory of relativity. The surface of electrical equilibrium, called a Heaviside ellipsoid, is an oblate spheroid contracted along the line of motion by a factor of $\sqrt{1 - v^2/c^2}$, although it was not until 1892 that this fact was fully clarified by Heaviside's friend G. F. C. Searle. All of this follows directly from Maxwell's equations and shows quite clearly that 'relativistic' effects were already implicit in Maxwell's theory. […]
> Fitzgerald replied [to Heaviside] that he was 'very glad to hear that you have solved completely the problem of the moving sphere' and remarked that, as the formula suggested, the velocity of light might be a physical limit to speed. He also mentioned the possible application of Heaviside's work to 'a theory of the forces between molecules', indicating that Fitzgerald already thought that intermolecular forces might be essentially electromagnetic. Indeed, since he believed that all physical forces, as well as matter itself, arose from the various motions of a single ether, Fitzgerald regarded Heaviside's formula for how electromagnetic forces varied with a velocity as a valuable guide to how other forces were likely to be affected by motion through the medium. (Hunt 1988, 71-2)[16]

In 1889 Fitzgerald made his insight concrete and proposed a length contraction factor of $\sqrt{1 - v^2/c^2}$ in a paper published in the journal *Science*[17]. This journal was a rather obscure one by the time, so the hypothesis did not have an immediate impact in the community. It only became more prominent when Lodge referred to it in his 1892-3 publications. It was via these works that Lorentz got acquainted with it, and in the *Versuch* he mentioned that Fitzgerald had independently arrived at the same result he obtained. The fact that both Lorentz and Fitzgerald introduced the same length contraction hypothesis, and the fact that they both justified it in the same way reinforce Lorentz's view (and also Fitzgerald's, of course) that from an electromagnetic view it was a rather plausible physical feature.

A second important remark about the Lorentz-Fitzgerald contraction consists in that Lorentz immediately noticed that a longitudinal contraction was not the only possible dynamical explanation for the null

[16] Hunt argues that Fitzgerald might have arrived to the contraction hypothesis even without knowing about the Michelson-Morley experiment, or even if it had never been performed. I remain neutral about this thesis, but his review of how Fitzgerald conceived the hypothesis is quite interesting in connection with Lorentz.

[17] "The length of material bodies changes according as they are moving through the ether or across it by an amount depending on the square of the ratio of their velocities to that of light. We know that electric forces are affected by the motion of the electrified bodies relative to the ether, and it seems a not improbable assumption that the molecular forces are affected by the motion and that the size of bodies alters consequently". From Fitzgerald's *The Ether and the Earth's Atmosphere*, quoted in (Hunt 1988, 75).

result of the Michelson-Morley experiment. A transverse dilation of bodies when moving across the ether in the suitable amount, or a combination of both effects, would also do[18]. However, by 1904 he got committed to a purely longitudinal effect for reasons connected to his model of the electron[19]. Moreover, in 1905-6 Poincaré stated that for theoretical reasons –consistency with his 'relativity principle' and the mathematical properties of the Lorentz transformations– the effect should be a purely longitudinal contraction. I will return to this issue below.

Finally, it is important to underscore a feature of Lorentz's theory which can be a source of confusion. In his *The Electromagnetic Theory of Maxwell and its Application to Moving Bodies* of 1892, Lorentz considered a set of coordinate transformations that included a spatial one in addition to the temporal one that in 1895 he called 'local time'. However, it is quite clear that the former transformation was not an expression of the length contraction hypothesis, but a purely mathematical tool connected to his quest of invariance for the Maxwell's equations, in the same sense that 'local time' was. Miller is very clear in this respect: after applying the Galilean transformation to the Maxwellian wave equations, Lorentz remarked that they no longer had the same form, so that in order to obtain invariance, he

> proposed an additional coordinate transformation on the inertial coordinates (x_r, y_r, z_r, t_r) in order that the Eq. (1.42) [the wave equation for the system S_r in motion through the ether which results of the application of the Galilean transformations] possessed the proper form of a wave equation […]:
>
> $$x' = \gamma x_r \qquad y' = y_r \qquad z' = z_r \qquad t' = t - (v/c^2)\gamma^2 x_r \ ;$$
>
> where $\gamma = 1/\sqrt{1 - v^2/c^2}$. (I called the primed reference system Q')[20]. Lorentz considered the transformation from S_r to Q' as a purely mathematical coordinate transformation—for example, he introduced x' as a "new independent variable", and similarly for t'. (Miller 1998, 26)

It is clear that in this context γ has nothing to do with length contraction. Moreover, these transformations, including γ, were used by Lorentz in his derivation of the electromagnetic forces that hold for a frame in motion with respect to the ether and that underlie his plausibility argument for the length contraction hypothesis. The question is then why γ was not included in the transformations that ground the theorem of corresponding states of 1895. Miller suggests that:

> Whereas a Galilean transformation from S to S_r failed to yield a proper wave equation, a further transformation from S_r to Q' resulted in a wave equation for a disturbance that depended on the emitter's motion, thereby violating an ether-based wave theory of light. Although Lorentz did not comment explicitly on this result for Q', we can assume that he noticed it because he wrote that calculations in the remainder of (1892a) [*The Electromagnetic Theory…*] were only to first-order accuracy in v/c, because this approximation facilitated further calculations, and it led to a "*theorème générale*." To first order in v/c the equations for the electromagnetic field quantities of the molecules constituting matter had the same form in S as in a reference system connected with S_r through the equations:
>
> $$x' = x_r \qquad y' = y_r \qquad z' = z_r \qquad t' = t - (v/c^2)x_r$$
>
> […] Hence, to order v/c the mathematical coordinate system Q' becomes in its spatial coordinates identical with the spatial Galilean coordinates, and the time coordinate mixes the Galilean absolute time $t_r(= t)$ with the Galilean spatial coordinate x_r. (ibid, 27)

[18] On this issue, see (Brown 2001).

[19] By 1895, it was clear that Lorentz sympathized with an only-longitudinal contraction effect, but he left open the possibility of the mentioned alternatives. In 1904 his position became more definite and definitive.

[20] Notice that Lorentz's 'two-steps method' is at work. S is the rest ether frame in which Maxwell's equations hold, S_r is a frame in motion with respect to the ether and in which the equations that hold is the result of the application of Galilean transformations, which are not invariant. Q' is the auxiliary frame in which the equations of S_r have been Lorentz-transformed and that are Lorentz-invariant.

In other words, the coordinate transformation for x that included γ, considered only as a mathematical tool, yielded a wave equation dependent on the state of motion of the emitter. That problem would be solved by adding the length contraction hypothesis to the γ-including transformation for x, but Lorentz took that step only in 1899, as I will show below.

I said that this feature of Lorentz's work of 1892 can be confusing because of the example of Zahar (1973, 211-2). He seems to understand that the γ of *The Electromagnetic Theory* is interpreted as the contraction factor in 1895—and also in *The Relative Motion of the Earth and the Ether* of 1892. This leads to an interpretation in which γ and the contraction hypothesis get conflated, but I think that Miller is very clear in that they are two very different things: the former is a mathematical tool, a coordinate transformation; whereas the latter is a physical hypothesis. They will only get more closely connected by Lorentz in 1899 and 1904, even though remaining logically independent. Actually, the fact that the *Versuch* offered two different and disconnected explanations for the null result of ether wind experiments of first and second order of v/c was the aim of a criticism that Poincaré made about Lorentz's theory. Now I turn to it and to some other observations that the French scientist and epistemologist introduced with respect to Lorentz's work.

2.2.2 Interlude: enters Poincaré

Henri Poincaré got involved in the development of Lorentz's work by underscoring that the explanations for the negative results of ether-wind experiments of first and second order it provided were two different and disconnected parts of the theory. His dissatisfaction about it was grounded on his view that physical science should be built upon certain principles that might be respected. In this case the relevant one is his *principle of relativity*,

> according to which the laws of physical phenomena must be the same for a stationary observer as for an observer carried along in a uniform motion of translation; so that we have not and cannot have any means of discerning whether or not we are carried along in such a motion. (Poincaré 1958, 94)[21]

Based on this principle, Poincaré believed that the result of any ether-wind experiment should be negative –a result that he did not qualify as surprising– and that the explanation for it must be based on the very core of a physical theory rather than on a compilation of different hypotheses and assumptions. It was in this sense that he criticized the structure of Lorentz's theory:

> I must explain why I do not believe, in spite of Lorentz, that more exact observations will ever make evident anything else but the relative displacements of material bodies. Experiments have been made that should have disclosed the terms of the first order; but the results were nugatory. Could that have been by chance? No one has admitted this; a general explanation was sought, and Lorentz found it. He showed that the terms of the first order should cancel each other, but not the terms of the second order. Then more exact experiments were made, which were also negative; neither could this be the result of chance. An explanation was necessary and was forthcoming; they always are; hypotheses are what we lack the least. But this is not enough. Who is there who does not think that this leaves to chance that this singular concurrence should cause a certain circumstance to destroy the terms of the first order, and that a totally different but very opportune circumstance should cause those of the second order to vanish? No; the same explanation must be

[21] This formulation is from an article entitled *L'État Actuel et l'Avenir de la Physique Mathématique* that he originally published in 1904. Charles Scribner shows that this view can be traced in Poincaré as early as 1895: "Experiment has revealed a multitude of facts which can be summed up in the following statement: it is impossible to detect the absolute motion of matter, or rather the relative motion of ponderable matter with respect to the ether; all that one can exhibit is the motion of ponderable matter with respect to ponderable matter". From *L'Éclairage Électrique*, quoted in (Scribner 1964, 673). It was only in 1904, in the passage that I quoted, when he first dubbed his principle as 'the principle of relativity'. Notice that in the passage of 1895 it is quite clear that the status of this principle is not *a priori* or ultimate, in the sense that it does not require further explanation; the principle is grounded on experience Poincaré never quit to this view.

> found for the two cases, and everything tends to show that this explanation would serve equally well for the terms of the higher order and that the mutual destruction of these terms will be rigorous and absolute. (Poincaré 1952a, 172) [22]

A second important issue in which Poincaré was important in the development of Lorentz' theory consists in his interpretation of 'local time'. The analysis that he provides of this concept is such that, unlike Lorentz's, it has a definite physical meaning: local time is a *measured* quantity in a frame in motion with respect to the ether, whereas the real time can only be measured in the ether-rest frame. He provides his analysis by means of the case in that time measurements and the determination of simultaneity are established through the interchange of light signals between two observers. At time $t_A = 0$ in his watch observer A sends a light signal to observer B, and when the latter receives the signal at time t_B, $B's$ clock must be set to AB/c in order to get synchronized with $A's$ –where AB is the distance between the observers and c is the speed of the light signal. If at $t_b = AB/c$ B sends back a light signal to A, then A will receive it a time $2AB/c$. At this point is when Poincaré introduces his relevant observation:

> In fact they mark the same hour at the same physical instant, but on the one condition, that the two stations are fixed. Otherwise the duration of the transmission will not be the same in the two senses, since the station A, for example, moves forward to meet the optical perturbation emanating from B, whereas the station B flees before the perturbation emanating from A. The watches adjusted in that way will not mark, therefore, the true time; they will mark what may be called the *local time*, so that one of them will gain on the other. It matters little, since we have no means to perceive it. All the phenomena which happen at A, for example, will be late, but all will be equally so, and the observer will not perceive it, since his watch is slow; so, as the principle of relativity would have it, he will have no means of knowing whether he is at rest or in absolute motion. (Poincaré 1958, 99)[23]

Poincaré does not explicitly mention the ether as the referential 'object' with respect to which the light does travel with the same velocity in all directions[24]. However, if one reads his writings it is clear that he is consistent in using the expression 'absolute motion' as meaning 'motion with respect to the ether'. Therefore, the *true* time is *measured* only in the ether-rest frame, whereas any motion with respect to it determines that the time to be *measured* will be the *local* one. Thus, Lorentz's auxiliary quantity in his 1895 coordinate transformations offers an explanation, up to first order, of why the observers do not notice their motion across the ether:

> The proof goes as follows. When B receives the signal from A, he sets his watch to zero (for example), and immediately sends back a signal to A. when A receives the latter signal, he notes the time τ that has elapsed since he sent his own signal, and sets his watch to the time $\tau/2$. By doing so he commits an error $\tau/2 - t_-$, where t_- is the time that light really takes to travel from B to A. This time and that of the reciprocal travel are given by $t_- = AB/(c+u)$ and $t_+ = AB/(c-u)$, since the velocity of light is c with respect to the ether. The time τ is the sum of these two traveling times. Therefore, to first order in u/c, the error committed in

[22] From *Sur les Rapports de la Physique Expérimentale et de la Physique Mathématique*, originally published in 1900.
[23] From *L'État Actuel*... (1904).
[24] One must be careful about this point. In an article of 1898, Poincaré states that the assumption of the light speed being the same in all directions—in the ether-rest frame—is a *convention* that cannot be verified by any experiment. See *The Measure of Time*, the English translation of that article, in Poincaré 1958, 27-36.
With respect to the ether and the speed of light, the following passage is maybe more clear:
"If they are carried along in common motion… [Suppose] now that A, for example were overtaking the light that went to B, while B receded from the light that went to A. if the observers are thus carried along in a common translation and they do not suspect it, their regulation [of their clocks] will be defective; their clocks will not indicate the same time; each of them will indicate the *local time* proper to the place where they find themselves.
The two observers will have no means of perceiving if the stationary ether always transmits the advancing light signals with the same velocity. … The phenomena that each of them would observe would be either advanced or retarded; they would not occur at the same moment as if the translation did not exist, but as if when one were to observe a badly regulated clock, one could not perceive [the motion]… . The appearances would not be altered". Quoted in (Goldberg 1967, 940).

> setting the watch A is $\tau/2 - t = (t_+ - t\_)/2 = uAB/c^2$. At a given instant of the true time, the times indicated by the two clocks differ by uAB/c^2, in conformity with Lorentz's expression of the local time. (Darrigol 2005, 10)[25]

That is, according to Poincaré, Lorentz's local time does not only explain optical experiments designed to measure ether-wind, but also why this *time measuring* effect occurs. The latter phenomenon was not envisioned in the scope of Lorentz's mathematical interpretation of local time.

One last reference to Poincaré's reception of Lorentz's theory that I will address has to do with yet another criticism he put forward. Since the theory conceives a purely electromagnetic ether that affects ponderable matter in electrodynamic terms, but which in turn is not affected by the latter—the most apparent example being that its motion has no consequences at all on the ether—it implies a violation of Newton's third law, the principle of action and reaction. Lorentz had seen that point, but unlike Poincaré, and based on the empirical success and the very wide scope of his theory, he simply concluded that the principle had to be considered in a more modest way:

> It is true that this conception [the immobile ether] would violate the principle of the equality of action and reaction –because we do not have grounds for saying that the ether *exerts* forces on ponderable matter– but nothing, as far as I can see, forces us to elevate that principle to the rank of a fundamental law of nature. (from Lorentz's *Versuch*, quoted in Janssen 2003, 34)

Poincaré rejected this attitude towards the principle because its violation gets associated with violations of other important and central mechanical laws, namely, the conservation of momentum and the center-of-mass theorem[26]. He illustrated his point by means of the following example:

> Imagine, for example, a Hertzian oscillator, like those used in wireless telegraphy; it sends out energy in every direction; but we can provide it with a parabolic mirror, as Hertz did with his smallest oscillators, so as to send all the energy produced in a single direction. What happens then according to the theory? The apparatus recoils as if it were a cannon and the projected energy a ball; and that is contrary to the principle of Newton since our projectile here has no mass it is not matter, it is energy. (Poincaré 1958, 101)[27]

If the recoil of the Hertzian antenna is not accompanied by a reaction in the ether, Newton's third law is violated. If this reaction does not happen, there is a loss of momentum in the system that is not compensated. Finally, the recoil of the antenna also implies that the center of mass of the system gets accelerated. Poincaré solved the three problems by one single theoretical stroke: the introduction of a 'fictitious fluid' in the ether. He assigned this fluid an electromagnetic momentum given by $\frac{1}{4\pi c}\int E \times B\, dV$, where V stands for the volume occupied by the fluid. This momentum played the role of the reaction for the recoil of the antenna, so that $\frac{d}{dt}\left[m\mathbf{v} + \frac{1}{4\pi c}\int E \times B\, dV\right] = 0$. Both Newton's third law and momentum conservation are satisfied if the fictitious fluid is considered.

The center of mass theorem could also be saved. The momentum expression for the fictitious fluid implicitly states that it possesses a mass density given by E/c^2, where E stands for the electromagnetic energy carried by the fluid. If no electromagnetic energy is created or destroyed, the mass density of the fictitious fluid is conserved and the global center of mass of matter *and* fluid moves along a straight line and with constant velocity. If electromagnetic energy was destroyed and transformed into another form

[25] u is of course the velocity of A and B with respect to the ether.

[26] This theorem affirms that in an isolated system, a system in which no external forces act, no process can alter the state of motion of its center of mass. It is quite obvious that it is closely connected to Newton's third and first laws.

[27] Originally from *L'État Actuel…* (1904). His original treatment of this issue appeared in *La Théorie de Lorentz et le Principe de Réaction*, included in a collective volume celebrating the 25th anniversary of Lorentz doctorate, and published in 1900. In this work Poincaré also referred to his interpretation of local time and to his criticism of the structure of Lorentz's theory.

of energy, Poincaré had to assume that the fluid associated to the destroyed energy was not really destroyed, but brought to rest and 'stored' at the place of its destruction. In the case of electromagnetic energy created, the required assumption was that 'latent' fluid available at the place of that creation got liberated.

Two things are noteworthy regarding Poincaré's fictitious fluid. First, it implicitly included the energy-mass relation $E = mc^2$, but Poincaré did not have the theoretical framework required to recognize it in its full meaning. He did not interpret the former as stating that energy, in and by itself, has inertia; or that the inertial mass of a material body can vary according to its energy content. Actually, Einstein's (1906) second derivation of his famous equation was based on the validity of the center-of-mass theorem, and he explicitly stated that his treatment of the issue was similar to the one undertaken by Poincaré that I just sketched. Regarding this, Janssen shows that in order to make the theories fully equivalent, Lorentz's must *borrow* $E = mc^2$ from SR. This is certainly right, especially from a historical point of view. However, from a conceptual standpoint, it must be acknowledged that the famous equation was there, in Lorentz's theory, 'waiting to be discovered'. It is true that it was not, and that Lorentz only saw it after Einstein's work; but all the conceptual machinery needed to formulate it was already present in Lorentz's theory. I will return to this issue below.

Second, the physical underpinning of the fictitious fluid was rather strange. Poincaré introduced this quantity with the intention of clarifying how Lorentz's theory violated the Newtonian principles, and as a sort of desperate maneuver involving physical fictions in order to save them. Actually, Lorentz rejected the fictitious fluid insofar as it implied a kind of motion in the ether, which he believed to be totally immobile. However, this unease got softened when Max Abraham, in 1902, (re)introduced the concept of electromagnetic momentum as carried not by a fluid in the ether, but by the electromagnetic fields themselves. Electromagnetic momentum would later be a crucial feature of Lorentz's model of the electron in the definitive formulation of the ether theory.

Summarizing, ca. 1900, Poincaré got crucially involved in the development of Lorentz's theory. He criticized its structure, and as I will now show, Lorentz's reaction greatly improved its foundations. On the other hand, he was able to see that there was a physical meaning contained in 'local time', and he also noticed that the theory had implications on momentum conservation and the center-of-mass theorem that, if properly considered, would lead to the energy-mass relation equation. Both these features are crucial in a case for the predictive equivalence of the theories.

2.2.3 Second stage: 1899-1904

Lorentz's definitive formulation of his theory was presented in his *Simplified Theory of Electrical and Optical Phenomena in Moving Bodies* (1899), and in *Electromagnetic Phenomena in Systems Moving with any Velocity Less than that of Light* (1904). In the first work he gave a unified and exact formulation of the theorem of corresponding states that gets closely connected to the hypothesis of length contraction. In the second, he added his very important and famous model of the electron. I will now offer an exposition of both issues in turn.

The definitive formulation of the corresponding states theorem was given by a modification of the coordinate transformations he had introduced in 1895. The new transformations are:

$$x' = l\gamma x, \qquad y' = ly, z' = lz, \qquad t' = l[t/\gamma - \gamma(v/c^2)x],$$

with $= \frac{1}{\sqrt{1-v^2/c^2}}$. The term l can differ from 1 only by an amount in the order of v^2/c^2 and Lorentz left it undetermined in 1899, but in 1904 set it to 1—for reasons that I will refer to below. In any case, the presentation of the final theorem gets harmlessly simpler if it is set to 1 right away.

It is important to remember that Lorentz is using his two-steps method, so the coordinates in S_0 at rest in the ether convert to the coordinates of S, in motion with respect to the ether, by means of the Galilean transformations. Finally, the S coordinates convert in the coordinates in the auxiliary frame S' through the Lorentz transformations. Therefore, combining the Galilean transformations from S_0 to S with the transformations from S to S', we have that the transformations from S_0 to S', with $l = 1$, are:

$$x' = \gamma(x_0 - vt_0), \qquad y' = y_0, \qquad z' = z_0, t' = \gamma[t_0 - (v/c^2)x_0]^{28}$$

By means of these new transformations, Maxwell's field equations become invariant without neglecting terms of any order of v/c. That is, the auxiliary fields in the system S', considered as functions of the auxiliary coordinates in S', satisfy the same equations as the real fields considered as functions of the real coordinates in S_0. Thus Lorentz's new formulation of the theorem of corresponding states consists in that

> If there is a solution of the source free Maxwell equations in which the real fields **E** and **B** are certain functions of $\mathbf{x}_0$ and t_0, the coordinates of S_0 and the real Newtonian time, then there is another solution of the source free Maxwell equations in which the fictitious fields **E'** and **B'** are those exact functions of **x'** and t', the coordinates of S and the local time in S. (Janssen 1995, section 3.3.3)

As I mentioned above, one must be careful and not to conclude right away that the factor γ expresses the length contraction factor. Actually, in 1892 he used it as a mere mathematical tool, as I showed above. Janssen is very clear that the same precaution must be taken here. This final formulation of the theorem of corresponding states does not logically entail the contraction; this is a further physical assumption. However, it is quite clear that one of the main motivations underlying Lorentz's work of 1899 and on was to merge the theorem with the contraction, in order to provide a unified and general explanation for why none of the optic experiments set out to find ether-wind effects had obtained positive results –that is, to fulfill Poincaré's demand. The accomplishment of this goal was given by a specific interpretation of the theorem under what Janssen dubs the *generalized contraction hypothesis*:

> If a material system, i.e., a configuration of particles, with a charge distribution that generates a particular electromagnetic field configuration in S_0, a frame at rest in the ether, is given the velocity **v** of a Galilean frame S in uniform motion through the ether, *it will rearrange itself* so as to produce the configuration of particles with a charge distribution that generates the electromagnetic field configuration in S that is the corresponding state of the original electromagnetic configuration in S_0. (Janssen 1995, section 3.3.3, my emphasis)

To clearly see the difference between the theorem of corresponding states, as a mathematical tool, and the generalized contraction hypothesis, as a physical assumption, it is enough to pay attention to the fact that the former establishes a relation between –on the one hand– two *real* frames S_0 and S, and—on the other hand—an *auxiliary* frame S' through the application of the coordinate transformations; whereas the physical length contraction assumption establishes a relation between the *real* frames S_0 and S. That is, the field configuration in S_0 *physically* transforms in its corresponding state configuration in the frame S. Notice also that the generalized contraction hypothesis can be understood as a generalization of the plausibility argument that Lorentz offered for the contraction hypothesis in 1892-5.

Janssen, in order to clarify this point, shows that the theorem can be applied to obtain an explanation of the Michelson-Morley experiment *with* and *without* the generalized contraction hypothesis. Without it, the corresponding state of the interferometer in S is a 'stretched out' interferometer in S_0, so that the latter

[28] The derivation of the time coordinate transformation goes as follows:
$t' = t/\gamma - \gamma(v/c^2)x = t_0/\gamma - \gamma(v/c^2)(x_0 - vt_0) = \gamma(t_0[1/\gamma^2 + v^2c^2] - (v/c^2)x_0)$; and since $(1/\gamma^2 + v^2c^2) = 1$, then $t' = \gamma[t_0 - (v/c^2)x_0]$.

> will change its shape as the moving interferometer is rotated. This means that whether it is dark or light at *P'* (and thereby at *P*) will depend on the orientation of the moving interferometer. Without the contraction hypothesis, Lorentz's theory therefore predicts a positive result in the Michelson-Morley experiment.

On the other hand, with the generalized contraction hypothesis,

> The corresponding state of the moving *contracted* interferometer is simply the uncontracted interferometer at rest in the ether. So, the shape of the corresponding state will not depend on the orientation of the moving interferometer with respect to its velocity. As a consequence, we now expect negative results. (ibid)[29]

Let us remind the general explanation of the optic experiments given by the 1895 version of the theorem. It affirmed that the *structure* of a pattern of light and darkness in S_0 is the same as the one in S. This time it must be added that the pattern of light and darkness in S differs from the one in S_0 in that the former is contracted by a factor γ^{-1}. This *real* difference yields that the experiments will not have positive results even for the second order of v/c.[30]

Summarizing, we have that the theorem and the physical hypothesis are logically independent. However, Lorentz's view of the subject consisted in that the theorem had to be interpreted and understood under the physical assumption. He explicitly stated this view in his 1899 work:

> We shall not only suppose that the system S_0 *may* be changed in this way into an imaginary system S[31], but that, as soon as the translation is given to it, the transformation *really* takes place, of itself, i.e., by the action of the forces acting between the particles of the system, and the aether. Thus, after all, S will be the *same* material system as S [this clearly should be S_0].
> The transformation of which I have spoken, is precisely such a one as is required in my explication of Michelson's experiment. (from Lorentz's *Simplified Theory...* (1899), quoted in Janssen 1995, section 3.3.4)

This important subtlety has been many times unnoticed and has led to some confusion, for it can suggest a wrong interpretation of Lorentz's theory –in the sense that it is understood in a *modern relativistic* way that is not justified. For example, Janssen (ibid) quotes a passage of Pais' famous scientific biography of Einstein in which it is affirmed that 'the reduction of the Lorentz-Fitzgerald contraction to a consequence of Lorentz transformation is a product of the nineteenth century'. I think that Janssen is quite clear and right that there is no such reduction, and in that Lorentz was conscious of it.

The revised and improved version of the theorem of corresponding states, understood under the generalized contraction hypothesis, entails a very surprising result from the point of view of classical physics. In the ether-rest frame, Newton's second law has the form $\mathbf{F} = m\mathbf{a}$, whereas in the frame S' in motion along the x-axis of S_0, it is $\mathbf{F}' = m\mathbf{a}'$; for according to Newtonian physics the inertial mass of a body is an

[29] The explanation *without* the contraction hypothesis would correspond to Lorentz's conception of γ in 1892 that I explained above.

[30] Yet another concrete case in which is clear that the theorem and the physical assumption are logically independent, though they get merged by Lorentz in 1899, is mentioned by Janssen and Stachel: 'What prompted Lorentz's new more general theory was in fact a variant of the Michelson-Morley experiment proposed in 1898 by Alfred Liénard. Liénard wanted to repeat the Michelson-Morley experiment with some transparent medium in the arms of the interferometer. In that case, the Lorentz-Fitzgerald contraction would no longer ensure that the travel time in an arm of an interferometer is independent of whether the arm is parallel or perpendicular to the ether drift. Liénard did not actually perform the experiment, but both he and Lorentz strongly suspected that the outcome, as the outcome of so many experiments before, would be negative. As Lorentz emphasized in his 1899 paper, his new theory could account for such a negative result' (2004, 28).

[31] Janssen claims that this system S referred by Lorentz does correspond to the real system S, and that use of the adjective 'imaginary' is simply based on the fact that the state of the system S is described by the fictitious *imaginary* primed coordinates. It might be so, but I think that the meaning of the passage gets even clearer if one substitutes S' for S. I completely agree in that in the second case he is clearly referring to S_0.

absolute invariant quantity. This cannot be the case in Lorentz's theory. The line of thought that led Lorentz to this result was based on the case of an oscillating electron in the ether-rest frame which generates an electromagnetic wave, wave whose oscillation satisfies Newton's law. Then he considered that same electron in a frame moving through the ether with velocity v. Its motion in the corresponding state, determined by the auxiliary quantities in the coordinate transformations, is the same as its motion in the ether-rest frame. This implies that, in terms of the *real quantities*, Newton's law holds only if the mass of the electron depends on its velocity.

Remember that Lorentz assumed that all forces transform as electromagnetic ones do, so that in a frame in motion with respect to the ether the force **F**' is given by

$$\mathbf{F}'_{x'} = \mathbf{F}_x \,, \quad \mathbf{F}'_{y'} = \gamma \mathbf{F}_y = \frac{\mathbf{F}_y}{\sqrt{1-v^2/c^2}}, \qquad \mathbf{F}'_{z'} = \gamma \mathbf{F}_z = \frac{\mathbf{F}_z}{\sqrt{1-v^2/c^2}} \,.$$

On the other hand, the relation between the acceleration **a** in the ether-rest frame and the acceleration **a**' in the auxiliary frame—if the velocity and amplitude of the electron's oscillation are small enough as to be neglected—is given by:

$$\mathbf{a}'_{x'} = \gamma^3 \mathbf{a}_x, \qquad \mathbf{a}'_{y'} = \gamma^2 \mathbf{a}_y, \qquad \mathbf{a}'_{z'} = \gamma^2 \mathbf{a}_z \,.$$

With these expressions, Newton's second law can be formulated, in terms of the *real quantities*, as

$$\mathbf{F}_x = m\gamma^3 \mathbf{a}_x \,, \qquad \mathbf{F}_y = m\gamma \mathbf{a}_y \,, \qquad \mathbf{F}_z = m\gamma \mathbf{a}_z$$ [32].

From these expressions it follows that if Newton's law of motion is to hold in the moving frame, then it cannot be the case that inertial mass is an absolute quantity. In this theory mass becomes a velocity-dependent property. To state the precise formula for this dependence, it must be noticed that the first of the expressions just above holds for the acceleration in the direction of motion –assuming that the motion occurs along the x-axis, of course–, whereas the other two hold for the accelerations perpendicular to the direction of motion. Considering this we have that in Lorentz's theory inertial mass is given by

$$m_L = \gamma^3 m_0 \,, \qquad m_T = \gamma m_0,$$

where m_L is the *longitudinal mass*, m_T is the *transverse mass*, and m_0 is the *ether-rest mass*.

This is a very surprising result from the point of view of Newtonian Mechanics, so Lorentz somehow felt a necessity to provide an explanation of it. He did so in his 1904 work, from the point of view of his model of the electron, model that was closely connected to what is commonly known as the *electromagnetic view of nature*. I now turn to a brief exposition of both subjects[33].

All the difficulties that classical mechanics had faced at the end of the 19th century—the ether problem and Lorentz's solution for it, for instance—led to the formulation of a new program for physical science. In 1900 Wilhelm Wien published a sort of *manifesto* whose main tenets were the assumption of a basic ontology determined by negative and positive charged particles that constitute all ponderable matter, and the view that the inertial mass of those particles is of electromagnetic origin. That is, the inertial mass of any object is the outcome of the interaction between the charged particles, their electromagnetic fields, and the ether. The explanation that Lorentz provided for his derivation of the velocity dependence of mass was grounded on these tenets. He actually claimed that the intrinsic relation between mass and

[32] For $\mathbf{F}'_{x'} = \mathbf{F}_x$; and $\mathbf{F}'_{z'} = \gamma \mathbf{F}_z = m\gamma^2 \mathbf{a}_z$, so that $\mathbf{F}_z = m\gamma \mathbf{a}_z$. The last result also holds, *mutatis mutandis*, for $\mathbf{F}'_{y'}$.

[33] This brief outline of the electromagnetic view of nature is based on the classic paper on the issue: (McCormmach 1970b). (Kragh 1999, chapter 8), and (Harmann 1982, chapters 4 and 6) also deal with this subject. On Lorentz's electron model, I closely follow (Janssen & Mecklenburg 2007) and (Janssen 1995, section 3.4). (Miller 1998, 62-80) is also useful.

velocity-across-the-ether was not that surprising after all, for some years before scientists like Thomson, Heaviside and Searle had already shown that the *effective* mass of a charged particle was a function of its velocity with respect to the ether: the effective mass of a particle was given by the sum of its *Newtonian* mass and its *electromagnetic* mass, where the latter was velocity-dependent. The electromagnetic view simply took one further step and asserted that *all* the inertial mass was of electromagnetic origin and velocity dependent. This view provided Lorentz with the basis for an explanation of his surprising result. If all inertial mass is a function of the interaction among charged particles and the ether, then its velocity dependence becomes quite a natural and expected feature.

Lorentz's specific position on this issue was formulated in his 1904 *Electromagnetic Phenomena in Systems Moving with any Velocity less than that of Light*, and it was grounded on five main assumptions: *i*) that all forces transform in the same way as electromagnetic forces do; *ii*) that a spherical electron, when moving across the ether, undergoes a physical deformation expressed by the coordinate transformations and becomes ellipsoid, i.e., the Lorentz contraction holds also for electrons themselves; *iii*) that the origin of all of the inertial mass of an electron is of electromagnetic; *iv*) that the deformation occurs only in the longitudinal direction with respect to the motion of the electron; and *v*) that the masses of all bodies, charged or not, vary with motion just as the mass of electrons does. Under these assumptions Lorentz was able to derive, for the electron itself, expressions for the velocity dependence of its mass that were equivalent to the ones he obtained in 1899 without considering a specific model of the electron. Notice Assumption *iv*) entails that the factor l I mentioned above gets definitively set to 1. Lorentz's reason for this choice was that 1 is the only value for l in which the velocity dependence of mass is consistent with the corresponding states theorem and the generalized contraction hypothesis, for any other value would make it possible to measure some ether-wind effect.

More specifically, if we take Newton's second law $\mathbf{F} = m\mathbf{a}$ under the assumption that all inertial mass is of electromagnetic origin—and therefore all mechanical mass supervenes on electromagnetic interactions—the law reduces to $\mathbf{F} = 0$. Expressed in terms of momentum variation through time, the law becomes $\frac{d\mathbf{P}_{tot}}{dt} = 0$ (the subscript simply indicates that momentum need not be only mechanical), an expression for momentum conservation. By defining $\mathbf{F}_{ext}$ as the Lorentz-force acting on the electron coming from the external field, and $\mathbf{F}_{self}$ as the Lorentz-force coming from the electron self-field, the equation of motion for an electron in an external field that Lorenz stated was $\mathbf{F}_{ext} + \mathbf{F}_{self} = 0$. The Lorentz-force on the electron from its self-field can be written as minus the derivative of electromagnetic momentum with respect to time, i.e., $\mathbf{F}_{self} = -\frac{d\mathbf{P}_{em}}{dt}$ [34], and given the equation of motion for the electron in an external field, it follows that $\mathbf{F}_{ext} = \frac{d\mathbf{P}_{em}}{dt}$.

Assuming that the momentum is in the direction of motion, its formula can be written as $\mathbf{P}_{em} = \left(\frac{P_{em}}{v}\right)\mathbf{v} = P_{em}\frac{\mathbf{v}}{v}$. Differentiating this expression with respect to time we get $\frac{d\mathbf{P}_{em}}{dt} = \frac{dP_{em}}{dt}\frac{\mathbf{v}}{v} + P_{em}\frac{d}{dt}\left(\frac{\mathbf{v}}{v}\right)$. Since we can consider P_{em} as a function of v, the first term in the right hand side of the last equation can be expressed as $\frac{dP_{em}}{dv}\frac{dv}{dt}\frac{\mathbf{v}}{v} = \frac{dP_{em}}{dv}\mathbf{a}_L$, where $\mathbf{a}_L$ stands for the *longitudinal* acceleration with respect to the direction of motion. The second term can be written as $P_{em}\frac{d}{dt}\left(\frac{\mathbf{v}}{v}\right) = \frac{P_{em}}{v}\mathbf{a}_T$, where $\mathbf{a}_T$ stands for the *transverse* acceleration with respect to the direction of motion. Finally, since $\mathbf{F} = \frac{d\mathbf{p}}{dt} = m\mathbf{a}$, we have that $\frac{d\mathbf{P}_{em}}{dt} =$

[34] 'The Lorentz force that an electron moving through the ether at velocity $\mathbf{v}$ experiences from its self-field can be written as minus the time derivative of the quantity that Abraham proposed to call the electromagnetic momentum:

$$F_{self} = \int \rho(\mathbf{E} + \mathbf{v} \times \mathbf{B}) d^3x = -\frac{d\mathbf{P}_{em}}{dt}.$$

In this expression ρ is the density of the electron's charge distribution, and $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic field produced by this charge distribution. The electromagnetic momentum of these fields is defined as

$$P_{em} \equiv \int \varepsilon_0(\mathbf{E} \times \mathbf{B}) d^3x$$

and doubles as the electromagnetic momentum of the electron itself'. (Janssen and Mecklenburg 2007, p. 8).

$m_L\mathbf{a}_L + m_T\mathbf{a}_T$, with the *longitudinal* mass of the electron $m_L = \frac{dP_{em}}{dv}$, and the *transverse* mass of the electron $m_T = \frac{P_{em}}{v}$.

It can be shown that $\mathbf{P}_{em} = \frac{4}{3}\gamma l\left(\frac{U'_{em}}{c^2}\right)\mathbf{v}$[35]—where U'_{em} stands for the energy of the electron in the frame S' in motion with respect to the ether, c is the velocity of light, and l stands for the transverse deformation factor mentioned above. Plugging the right hand side of this equation in the expressions for the electron masses, in the case of longitudinal mass we get $m_L = \frac{d(\gamma lv)}{dv}\frac{4}{3}\frac{U'_{em}}{c^2}$, and since $\frac{d}{dv}(\gamma v) = \gamma^3$[36], then $m_L = y^3 l\frac{4}{3}\frac{U'_{em}}{c^2}$. In the case of the transverse mass, we get $m_T = \gamma l\frac{4}{3}\frac{U'_{em}}{c^2}$. Since the expressions that Lorentz derived in 1899, from the theorem of corresponding states and the generalized contraction hypothesis applied to $\mathbf{F} = m\mathbf{a}$, are $m_L = y^3 lm_0$ and $m_T = \gamma lm_0$, it turns out that for the 1904 results to be consistent with the 1899 expressions, the rest-mass must be defined as $m_0 = \frac{4}{3}\frac{U'_{em}}{c^2}$, and the factor l must be set to 1[37].

Supporters of the electromagnetic worldview took the equation $\mathbf{F}_{ext} = \frac{d\mathbf{P}_{em}}{dt} = m_L\mathbf{a}_L + m_T\mathbf{a}_T$ as the fundamental equation of motion, and interpreted Newton's second law as special cases of it. For $v = 0$, that is, for an electron at rest in the ether, it holds that $m_L = m_T = m_0$, so that for the ether-rest electron $\mathbf{F}_{ext} = m_0\mathbf{a}$. For a moving electron, the values of m_L and m_T differ from m_0 only by an amount of the order v^2/c^2. Therefore, for $v \ll c$, $\mathbf{F}_{ext} \approx m_0\mathbf{a}$. This reduction of a central law of mechanics to electrodynamics was, of course, considered as a promising result of the electromagnetic program. Furthermore, as we just saw, from his model of the electron Lorentz obtained an expression of the velocity-dependence of inertial mass that was consistent with the expression he found in 1899. That is, he naturally interpreted that he had found a model of the elementary charged-particle that explained the results he obtained in 1899 without specific dynamical assumptions. In other words, the 1904 model of the electron provided strong support for the universal scope of the generalized contraction hypothesis. If all laws are, bottom-line, electromagnetic, then the Lorentz-invariance of all laws of physics is a natural consequence.

In order to understand more deeply the meaning of Lorentz's electron model, it is useful to make a brief comparison with its rivals. By 1905 there were two other alternatives available. Both Abraham's and Bucherer-Langevin's models—just as Lorentz's—assumed a full electromagnetic origin for the inertial mass of charged particles, and that charged particles were the ultimate constituents of ponderable matter. The differences were that Abraham's electron was rigid and not affected by a contraction when set in motion through the ether; and that the Bucherer-Langevin electron, along with a longitudinal contraction, also suffered a transverse expansion due to motion, the value of the contraction being $\gamma^{2/3}$ and the value of the dilation being $\gamma^{1/3}$—so that the factor l is equal to $\gamma^{-1/3}$. In other words, Abraham's model denied Lorentz's assumptions *i*) and *ii*), and assumption *iv*) becomes irrelevant; whereas Bucherer-Langevin's denies assumption *iv*); but all three models share assumptions *iii*) and *v*), which constitute the basic tenets of the electromagnetic view of nature.

The line of scientific development from the problem of the ether to the formulation of Lorentz's theory settled the basic groundings for the electromagnetic world view. By the early 1900s, the three mentioned models committed to this view were contending, so the choice to be made was understood as a matter of empirical tests. The interpretation of the data obtained from Kaufmann's experiments performed during

---

[35] See (Janssen and Mecklenburg 2007, p. 19).

[36] $\frac{d}{dv}(\gamma v) = \gamma + v\frac{d\gamma}{dv} = \gamma + y^3\beta^2$, with $\beta = \frac{v}{c}$. Since $\gamma = \gamma^3(1-\beta^2)$, then $\frac{d}{dv}(\gamma v) = \gamma^3(1-\beta^2+\beta^2) = \gamma^3$. From this result is easy to infer that $\frac{d\gamma}{dv} = \gamma^3\frac{v}{c^2}$, this expression will be useful below.

[37] Since $\frac{d}{dv}(\gamma v) = \gamma^3$ and $\frac{d\gamma}{dv} = \gamma^3\frac{v}{c^2}$, it follows that $m_L = \frac{d(\gamma lv)}{dv}\frac{4}{3}\frac{U'_{em}}{c^2} = \left(y^3 l + \gamma v\frac{dl}{dv}\right)\frac{4}{3}\frac{U'_{em}}{c^2}$. For this expression to be equivalent to $m_L = y^3 lm_0$, it must be the case that $\frac{dl}{dv} = 0$, and since l is a function of v and can differ from 1 only by an amount of the order v^2/c^2, then $l = 1$. This is the dynamical reason why Lorentz established that the deformation of the moving-across-the-ether electron is only longitudinal. With $l = 1$, the coordinate transformations in Lorentz's theory are identical to the ones that Einstein obtained in 1905.

1901-3, in which β-radiation was used in order to measure the precise value for the velocity dependence of the inertia of particles, was the empirical battlefield on which the models—and also SR—competed. It turned out that the technology available was not enough in order to set the experiment in a way that the resulting data could be considered as totally reliable. However, the relevant point is that ca. 1905 the electromagnetic view of nature was hold as a very promising and unifying program for the development of physics, and as a program in that classical mechanics became reduced to electrodynamics.

In spite of the promising path that Lorentz's theory was opening—within the context of the electromagnetic view—, some problems quickly came up. The most relevant one was that of an ambiguity, or even an inconsistency, in the formulation of the longitudinal mass. This problem was first posed by Abraham in 1905: it turned out that the expression of the longitudinal mass of the electron in terms of its momentum was incompatible with the expression in terms of its energy.

Consider an electron moving in the x-direction and assume the absence of an external field. The force the electron experiences from its self-field does an amount of work that can be expressed in terms of the energy of the electron, $dU_{em} = -dW = -\mathbf{F}_{self} \cdot d\mathbf{x}$. Substituting $-\mathbf{F}_{self}$ for $\frac{d\mathbf{P}_{em}}{dt}$ we get $dU_{em} = \frac{d\mathbf{P}_{em}}{dt} \cdot d\mathbf{x} = m_L a_L \cdot d\mathbf{x} = m_L \frac{dv}{dt} dx = m_L v dv$. From this last equation it follows that $m_L = \frac{1}{v}\frac{dU_{em}}{dv}$.

Now, the relation between the energy of an electron in an ether-rest frame and its energy in a moving frame is $U_{em} = l\left(\frac{4\gamma}{3} - \frac{1}{3\gamma}\right) U'_{em}$[38]. If we set l to 1 and plug the right hand side of this equation in $m_L = \frac{1}{v}\frac{dU_{em}}{dv}$ we get $m_L = \frac{1}{v}\frac{d}{dv}\left(\frac{4\gamma}{3} - \frac{1}{3\gamma}\right) U'_{em} = \frac{1}{v}\frac{4}{3}\frac{d\gamma}{dv} U'_{em} - \frac{1}{3v}\frac{d}{dv}\left(\frac{1}{\gamma}\right) U'_{em}$, and since $\frac{d\gamma}{dv} = \gamma^3 \frac{v}{c^2}$, we finally obtain $m_L = \gamma^3 \frac{4}{3}\frac{U'_{em}}{c^2} - \frac{1}{3v}\frac{d}{dv}\left(\frac{1}{\gamma}\right) U'_{em}$. Notice that the first term of the right hand expression in this equation is equal to the value of the longitudinal mass of the electron obtained from its momentum. The presence of the second term is thus the root of the problem: longitudinal mass in terms of momentum is not equal to longitudinal mass in terms of energy. In Lorentz's model of the electron, we have that $m_L = \frac{dP_{em}}{dv}$, that $m_L = \frac{1}{v}\frac{dU_{em}}{dv}$, and that $\frac{dP_{em}}{dv} \neq \frac{1}{v}\frac{dU_{em}}{dv}$. From a modern-relativistic point of view, to define the rest mass as $\frac{4}{3}\frac{U'_{em}}{c^2}$ is rather awkward, for in that case the mass-energy relation becomes $\frac{4}{3}E = mc^2$, instead of $E = mc^2$. This issue is commonly known as the "4/3 puzzle" and it is involved in the problem of the ambiguity of the expression for the longitudinal mass I am considering[39].

Abraham pointed the inconsistency problem in his 1905 *Theory of Electricity: electromagnetic theory of radiation*—published in German. The interpretation he made of this issue was that the disagreement was grounded in the fact that the entire energy of Lorentz's electron could not be accounted for by electromagnetic forces alone. The solution consisted then in including an extra force to account for the total energy. The problem was that Abraham noticed that the introduction of such a force would threaten the purity of the electromagnetic view that Lorentz's theory was endorsing, for the compensating force could not be an electromagnetic one. Moreover, the compensating non-electromagnetic force was necessary to provide stability to Lorentz's electron; otherwise its own Coulomb repulsive forces would make it to explode[40]. A complete solution that followed Abraham's line of thought was introduced by Henri Poincaré in 1906, along with other interpretations and remarks about Lorentz's theory that are essential to really obtain its predictive equivalence with respect to SR.

---

[38] See (Janssen and Mecklenburg 2007, p. 18).

[39] For a deeper treatment of this issue see (Miller 1998, 72-80); (Miller 1986, 266-301, 309-18); (Janssen 1995, section 2.2); (Janssen & Mecklenburg 2007, 19-50); and (Janssen 2003). I will say some more words about it in the next section.

[40] Strictly speaking, the problem of the electron's stability held for the three models. However, Abraham's assumption of the rigid electron was 'axiomatic', so that the counterbalance of the Coulomb forces was not a force, but some sort of 'rigid constraints'. Therefore, he could tackle the problem of stability from within the electromagnetic world picture, that is, without introducing non-electromagnetic forces or energy. See (Janssen 1995, section 3.4.3); and (Janssen & Mecklenburg 2007, 22-5).

2.2.4 Poincaré, once again

In his 1906 *On the Dynamics of the Electron*—published in French[41]—Poincaré introduced a non-electromagnetic quantity in order to solve the problem of stability of the electron and the inconsistency of the expressions for its longitudinal mass. The quantity is commonly known as *Poincaré-pressure*. The formula for this pressure is $P_{Poincare} = -\frac{1}{3}\frac{U'_{em}}{V_0}$, where V_0 is the volume of the electron at rest[42]. Abraham had pointed out that the total energy of the electron could not be accounted for only by means of its electromagnetic energy. Poincaré-pressure thus contributes with the missing energy $\frac{1}{3}\frac{U'_{em}}{\gamma}$, which is minus the product of the P_p and the volume of the moving electron V_0/γ. By adding this missing energy to $U_{em} = l\left(\frac{4\gamma}{3} - \frac{1}{3\gamma}\right)U'_{em}$, we have that the total energy is $U_{tot} = \frac{4}{3}\gamma U'_{em}$. Defining the longitudinal mass in terms of energy as $\frac{1}{v}\frac{dU_{tot}}{dv}$, then $m_L = \frac{1}{v}\frac{dU_{tot}}{dv} = \frac{1}{v}\frac{d\gamma}{dv}\frac{4}{3}U'_{em} = \gamma^3\frac{4}{3}\frac{U'_{em}}{c^2} = \gamma^3 m_0 = \frac{dP_{em}}{dv}$, so the inconsistency is solved.

The compensating pressure Poincaré introduced made Lorentz's electron stable, for it counterbalanced the repulsive Coulomb forces, that, without assuming P_p, would make the moving electron to explode—this is why P_p is a negative quantity. Besides, P_p provided a physical foundation for the Lorentz contraction affecting the electron itself. The pressure is exerted only on the surface of the electron, so the forces involved make the electron to contract as it moves through the ether:

> These forces serve two purposes. First they prevent the electron's surface charge distribution from flying apart under the influence of the Coulomb repulsion between its parts. Second, as the region where $P_{Poincaré}(\mathbf{x})$ is non-vanishing always coincides with the ellipsoid-shaped region occupied by the moving electron[43], these forces make the electron contract by a factor γ in the direction of motion. (Janssen & Mecklenburg 2007, 31)

Yet another nicety of P_p is that, from a modern perspective, it solves the '$\frac{4}{3}$ puzzle' involved in the energy-mass relation implicit in Lorentz's theory. As we will see, This feature is crucial for the case of Lorentz vs. Einstein. If the total energy of the electron were given by the electromagnetic energy, $m_0 = \frac{4}{3}\frac{U'_{em}}{c^2}$ would be the definitive formula for inertial mass, so Lorentz's theory would not be empirically equivalent to special relativity and it would be falsified by the outcome of non-optical ether-drift experiments where the relativistic formula $E = mc^2$ is involved. However, if we consider the missing energy contributed by P_p, the energy of the electron is $U_{tot} = \frac{4}{3}\gamma U'_{em}$, as we saw above, and therefore, $m_0 = \frac{U_{tot}}{c^2}$[44].

Just as Abraham stated, Poincaré's solution entails that the goal of the electromagnetic world view was not completely fulfilled by Lorentz's theory. However, Poincaré did not hesitate about endorsing this result because Lorentz's theory and its model of the electron was the only available theoretical approach that respected his principle of relativity, and so precluded a positive outcome for any ether-wind experiment. As Janssen poses it:

---

[41] A shorter version of the paper appeared in 1905, with the same title.

[42] For an analysis of Poincaré's derivation of this expression, see (Janssen & Mecklenburg 2007, 26-31). For a detailed examination of (Poincaré 1906), see (Miller 1986, 29-150).

[43] The Poincaré pressure can be written, for an electron moving along the x-direction through the ether, as $P_{Poincare}(\boldsymbol{x}) = -\frac{1}{3}\frac{U_{oem}}{V_0}\vartheta\left(R - \sqrt{\gamma^2x^2 + y^2 + z^2}\right)$, where ϑ is a function defined as $\vartheta(x) = 0$ for $x < 0$,and $\vartheta(x) = 1$ for $x \geq 0$, and where R is the radius of the electron at rest. This holds for a co-moving frame related to an ether-rest frame by means of the Galilean transformation.

[44] These remarks on the $\frac{4}{3}$ puzzle rest on a reconstruction made with the benefit of hindsight, and their aim is to make a case for the predictive equivalence between Lorentz's theory, as amended by Poincaré, and SR. From a historical point of view, things are not so modern, for, as Miller points out: 'contrary to what sometimes is attributed to this paper, Poincaré never computed the counter term necessary to cancel the second term on the right hand of (48) [the equation for work that leads to the problematic equation of energy], nor did he reduce the factor of 4/3 in [P_{em}] to unity. Rather, he proved the necessity of introducing mechanical stresses into Lorentz's theory to account for the inertia of a deformable electron in a manner consonant with the principle of relativity' (1986, 70).

> There [was] no electron model that [was] both compatible with the electromagnetic view of nature and compatible with the general experimental indication that we will never be able to detect ether drift, and therefore with Einstein's relativity principle. The Lorentz electron [was] incompatible with the electromagnetic world view. The Abraham and Bucherer-Langevin are incompatible with the absence of any signs of ether drift. (Janssen 1995, section 3.4.1)

Both Lorentz and Poincaré accepted to sacrifice the purity of the promising new program for the unification of physics. As we have seen, the solution to the problem of the ether, the quest for the invariance of Maxwell equations, and Poincaré's synthesis of all the related issues in his principle of relativity were highly valuable achievements. However, Poincaré explicitly remained committed to the view that all the inertial mass of the electron is of electromagnetic origin: 'If the inertia of matter is exclusively of electromagnetic origin, as is generally admitted since Kaufmann's experiment, and all forces are of electromagnetic origin (apart from this constant pressure I just mentioned), the postulate of relativity may be established with perfect rigor'[45]. If it also considered that the basic ontology of the theory remained the same after Poincaré's amendments –charges, fields and the ether–, it is clear that Lorentz's theory stayed committed to the main tenets of the electromagnetic world view.

Another important amendment that the French scientist introduced in Lorentz's theory had to do with a correction in the expressions for the velocity and charge density transformations. The problem with this issue in Lorentz's theory was rooted, Poincaré noticed, in the two-steps method that I mentioned above. Let us recall that this method consisted in that Lorentz first connected a system S_0 at rest in the ether with a system S which moves with respect to S_0 with velocity v by means of the Galilean transformations $x = x_0 - vt_0$, $y = y_0$, $z = z_0$, $t = t_0$. Then the system S is connected to the auxiliary frame S' by means of the Lorentz transformations

$$x' = l\gamma x, \qquad y' = ly, z' = lz, \qquad t' = l[t/\gamma - \gamma(v/c^2)x].$$

The velocity transformation from $u_x = \frac{dx}{dt}$ to $u'_x = \frac{dx'}{dt'}$ that Lorentz obtained was $u'_x = \gamma^2 u_x$. Its derivation goes like this: from the Lorentz transformations from S to S' we have that $dx' = \gamma l dx$, and that $dt' = l\left[\frac{dt}{\gamma} - \gamma(v/c^2)x\right] = \frac{l}{\gamma}dt(1 - \gamma^2 v u_x)$[46]. By plugging the last two equations in $u'_x = \frac{dx'}{dt'}$ one finally obtains $u'_x = \frac{\gamma^2 u_x}{(1-\gamma^2 v u_x)}$[47]. This formula reduces to $u'_x = \gamma^2 u_x$ only if $u_x \ll 1$. That is, it holds if the velocity of the object moving in the frame S is much smaller than c, but if u_x increases the velocity transformation that Lorentz obtained is no longer correct.

Poincaré corrected this problem by means of a maneuver which has important consequences for the meaning of the Lorentz transformations. He simply avoided the two-step method and directly connected S_0 with S' through the Lorentz transformation expressed in their modern way:

$$x' = \gamma(x_0 - vt_0), \qquad y' = y_0, \qquad z' = z_0, t' = \gamma[t_0 - (v/c^2)x_0].$$

[45] From Poincaré *On the Dynamics of the Electron*, quoted in Janssen & Mecklenburg 2007, 34. This passage clearly indicates that Poincaré did not see that $U_{tot} = mc^2$, or at least that he did not interpret it in the modern way.

[46] By choosing the suitable units such that $c = 1$.

[47] $u'_x = \frac{\gamma l dx}{\frac{l}{\gamma}dt(1-\gamma^2 v u_x)}$, and since $u_x = \frac{dx}{dt}$; then $u'_x = \frac{\gamma l u_x}{\frac{l(1-\gamma^2 v u_x)}{\gamma}} = \frac{\gamma^2 u_x}{(1-\gamma^2 v u_x)}$.

By so doing, the expression for the velocity transformation is simply $u'_x = \frac{\gamma d(x-vt)}{\gamma d(t-vx)} = \frac{dx-vdt}{dt-vdx} = \frac{u_x-v}{1-vu_x}$ [48]. This is of course the modern relativistic velocity transformation, which implies that c is the maximum possible velocity in any inertial frame[49].

The deeply important consequence of Poincaré's maneuver that I mentioned consists in that, by directly relating S_0 and S', the relevant velocity v which operates in the transformation is, in the end, the *relative* velocity between the frames, not the velocity of S with respect to the *ether*—this is the reason why in the last paragraph I used simply x and t in the transformation that Poincaré obtained, instead of x_0 and t_0. That is, the amendment introduced by Poincaré is a step towards an interpretation of the Lorentz transformations in terms of relative velocities between the frames involved.

This feature becomes even more apparent by considering yet another improvement that Poincaré made on Lorentz's theory. He also showed that the Lorentz transformations form a *group*. A transformation group is a collection of transformations such that *i*) the transformation obtained through the successive application of two transformations of the collection is also a transformation of the collection; *ii*) the transformations are associative, that is, the transformation obtained by the composition of (AB) and C is equal to the transformation obtained from the composition of A and (BC), where A, B, C are transformations of the collection; *iii*) there is an identity transformation D such that $DA = A$; and *iv*) there exists an inverse transformation.

By proving that the Lorentz transformations comply with requirements *i*) and *iv*), Poincaré took yet another step towards an interpretation only in terms of relative velocities between frames. Consider the transformations from a frame S at rest in the ether to a frame S' moving with velocity v with respect to it:

$$x' = \gamma l(x - vt), \qquad y' = ly, z' = lz, \qquad t' = \gamma l(t - vx),$$

and then consider a second set of Lorentz transformations in terms of γ', l' and v'; but this time connecting S' to a different frame S'' moving with velocity v' with respect to S':

$$x'' = \gamma' l'(x' - v't'), \quad y'' = l'y', \qquad z'' = l'z', \qquad t'' = \gamma' l'(t' - v'x'),$$

where $\gamma = 1/\sqrt{1 - v^2}$, and $\gamma' = 1/\sqrt{1 - v'^2}$. Through the composition of the two transformations one obtains

$$x'' = \gamma'' l''(x - v''t), \; y'' = l''y, \qquad z'' = l''z, \qquad t'' = \gamma'' l''(t - v''x),$$

where $v'' = \frac{v-v'}{1-vv'}$, $\gamma'' = \gamma\gamma'(1 - vv') = 1/\sqrt{1 - v''^2}$, and $l'' = l'l$. These transformations, which connect the frame S with the frame S'', are also Lorentz transformations, so that the first requirement is satisfied. Notice that in the step which goes from S' to S'', the Lorentz transformation applied considers a velocity v' which is simply the relative velocity between the frames—the velocity with respect to the ether is not involved[50].

---

[48] Once again, under the assumption that $c = 1$.

[49] For brevity and simplicity, I only referred to the case of the x-velocity component. The expression for the other two components, in the case of both Lorentz's and Poincaré's derivations, follows quite analogously. For the transformation for the charge density, which is corrected by Poincaré by means of the same maneuver, see (Miller 1986, 47-7, 72-4).

[50] 'The crucial point is that if one considers three coordinate systems S, S', S'' where S' and S'' move along a common axis with uniform speeds ε with respect to S and ε' with respect to S', respectively, then the two Lorentz transformations from S to S' and from S' to S'' can be replaced by a single Lorentz transformation from S to S'' with the relative speed of S'' with respect to S of $\varepsilon'' = \frac{\varepsilon-\varepsilon'}{1-\varepsilon\varepsilon'}$' (Miller 1986, 84).

Yet another important result in Poincaré's proof of the group property of the Lorentz transformations consists in his reasoning leading to $l = 1$ and the symmetry[51] between the Lorentz transformation and its inverse transformation in the group. He considered the case in which the systems S and S' get rotated in 180° about their y-axes. The transformations between the frames so rotated, which must also belong to the group, are

$$x' = \gamma l(x + vt), \qquad y' = ly, z' = lz, \qquad t' = \gamma l(t + vx),$$

and it is assumed that the dependence of the factor l on the velocity v is not at all affected by replacing v with $-v$.

Then he considered the case of the inverse transformations:

$$x' = \frac{\gamma}{l}(x + vt), \qquad y' = \frac{y}{l}, \qquad z' = \frac{z}{l}, \qquad t' = \frac{y}{l}(t + vx),$$

and noticed that the only way in which they can be a part of the group is by establishing that $l = 1$ [52]. In turn, it is clear that if l has this value, the inverse transformations are identical to the transformations considered in the y-axis rotation case, and if this is so, the Lorentz-transformations are symmetrical. Once again, the velocity that is relevant for the transformations and their inverses is simply the *relative* velocity (v or $-v$) between the frames, not the 'absolute' one with respect to the ether.

A physical consequence of this symmetry is that the length-contraction effect becomes also symmetric. If rod B that moves in the ether-rest frame S gets contracted with respect to the rod A at rest in S, in a frame S' that moves with respect to the ether and in which B is at rest and A in motion, A gets contracted with respect to B—we assume that when at rest with respect to each other A and B have the same length. This is also the case in special relativity, so the symmetry involved in the group-proof that Poincaré obtained is yet another crucial amendment to make the theories predictively equivalent. However, the explanations for this symmetry that the theories provide are different. In Einstein's it is just a consequence of the frame-dependence of simultaneity and the resultant frame-dependence of length[53]. In Lorentz's theory the rod B gets *really* contracted, whereas the contraction of A measured in the frame S' is a consequence of the 'wrong' synchronization of clocks in it. That is, since the clocks in S' measure the 'local time', the simultaneous events that determine the length of A in that frame are not *really* simultaneous, and then the length measured is not the real one[54].

Poincaré *explicitly* showed that the correct interpretation of the transformations is symmetrical, i.e., that the relevant velocities involved are the relative ones. This feature was not originally noticed by Lorentz, he interpreted them as asymmetric—his view of the transformations was such that the transformation to *go back* to the ether-rest frame delivers *un*contracted lengths, for example, whereas Poincaré

---

[51] 'Symmetric' could be a confusing term here, it is used with a different sense in contemporary theoretical physics, for example. However, I prefer it to 'reciprocal' because 'symmetric' is the term that is normally used in the literature. The property at issue is simply that the inverse transformation of a Lorentz transformation has the exact same form but substituting $+v$ for $-v$. Both Miller (1973) and Janssen (1995) use 'symmetric' to denote this property.

[52] Notice the different way in which Lorentz and Poincaré determined that the value of l has to be 1. Lorentz obtained it via dynamical considerations, for it was necessary for the expressions of the v-dependence of inertial mass obtained from the correspondence states theorem and the general contraction hypothesis to be equivalent to the one obtained from his model of the electron. Poincaré, on the other hand, determined it by means of simple *mathematical* features of the transformations.

[53] That is, in special relativity the length of a rod is not an 'intrinsic' property of the object, it is the distance between the simultaneous events that coincide with the endpoints of the rod. Thus, if simultaneity is frame-relative, it follows that length is also so.

[54] See (Dorling 1968). For the observer in the system S' to be able to measure the real length of the rod A, the reading of the clocks in S' should be corrected. In order to do so, the velocity of S' with respect to the ether must be known, but this is impossible given the theorem of corresponding states and the generalized contraction hypothesis. If the observer uses the method that Poincaré considers she will inexorably measure the 'local time'.

showed that *both* systems S and S' determine that the lengths of bodies in (relative) motion get contracted. Curiously, as Janssen points out, Lorentz only saw this via Einstein, not via Poincaré, even though he was certainly aware of the Frenchman's work—Lorentz openly accepted the introduction of the *Poincaré-pressure*, for instance[55].

As we will see, the symmetry of the coordinate-transformations in Lorentz's ether theory entails two non-empirical features that can be invoked to argue special relativity's superiority. First, it involves one of the asymmetries that do not belong to the phenomena that Einstein deprecated in the introduction of his 1905 paper. The explanation for the same physical effect, length-contraction, is different whether we consider the frame as at rest or as moving with respect to the ether. Second, the symmetry of the transformations entails that the velocity-with-respect-to-the-ether is not necessary for the derivation of length-contraction, clock-retardation or the velocity-dependence of inertial mass. The velocity involved is simply the *relative* velocity between the frames involved—this is clear if we consider the transformations from S' to S in the example, in that case v is the velocity of the ether-rest frame S with respect to S', not the velocity of S with respect to the ether, of course. This makes the ether not only undetectable, but also empirically superfluous.

One final contribution to the ether theory that Poincaré made in 1906 was that he noticed that the Lorentz transformations leave the quantity $ct^2 - x^2 + y^2 + z^2$ invariant. Moreover, he noted that by defining a four-dimensional space with axes x, y, z and it, the transformations represent rotations of such a space around a fixed origin, and he also discovered that some physical quantities, e.g. electric charge and current density, can be combined in four-component, Lorentz-invariant quantities. That is, Poincaré anticipated the seminal work of Hermann Minkowski on the four-dimensional formulation of special relativity. However, unlike relativity in four-dimensional space-time, in the ether theory these properties represent mere *mathematical niceties* that do not have a physical meaning. As we will see below, Poincaré's glimpse at four-dimensional space-time is a relevant feature to assess one of the reasons that have been typically invoked to favor Einstein's theory over Lorentz's, namely, mathematical elegance.

Before finishing this section and turning to Einstein's theory, I will make a brief and general remark about a fascinating and important issue: did Poincaré independently discover SR? Some authors state that he did—Giedymin, Zahar and Whittaker[56]. The main arguments for this conception are Poincaré's analysis of the measurement of time and the determination of distant simultaneity, his amendments on the meaning of the Lorentz transformations, and the fact that he derived some *mathematical* results from the transformations that clearly prefigure Minkowski's space-time.

In spite of the many results that Poincaré obtained, and in spite of the many epistemological considerations which resemble some of the ones that Einstein also did, I think that Poincaré did not discover SR. Even though he made a step towards it with his right foot, his left foot and his whole body stayed in the core of classical mechanics. His *relativistic glimpses* are only spots in a non-relativistic backdrop. To prove that, allow me to quote at length a passage that shows his commitment to an ether that stands still with respect to the motion of bodies in a Euclidian-space-through-absolute-time:

> [in order to account for the negative result of the ether-wind experiments] The most ingenious [hypothesis] was that of local time. Imagine two observers who wish to adjust their timepieces by optical signals; they exchange signals, but as they know that the transmission of light is not instantaneous, they are careful to cross them. When station B perceives the signal from station A its clock should not mark the same hour as that of station A at the moment of sending the signal, but this hour augmented by a constant representing the duration of the transmission. Suppose, for example, that station A sends its signal when its clock marks the hour 0, and that station B perceives it when its clock marks the hour t. The clocks are adjusted if the slowness equal to t represents the duration of the transmission, and to verify it, station B sends in its turn

[55] See (Janssen 1995, section 3.5)

[56] (Zahar 2001, Ch. 4); (Whittaker 1953, Ch. 2); (Giedymin 1982, Ch. 5). I thank Roberto Torretti for showing me how subtle this subject is and for his generous suggestions and corrections on my views about it.

> a signal when its clock marks 0; then station should perceive it when its clock marks t. The timepieces are then adjusted.
> And in fact they mark the same hour at the same physical instant, but on the one condition, that the two stations are fixed. Otherwise the duration of the transmission will not be the same in the two senses, since the station A, for example, moves forward to meet the optical perturbation emanating from B, whereas the station B flees before the perturbation emanating from A. The watches adjusted in that way will not mark, therefore, the true time; they will mark what may be called the *local time*, so that one of them will gain on the other. It matters little, since we have no means to perceive it. All the phenomena which happen at A, for example, will be late, but all will be equally so, and the observer will not perceive it, since his watch is slow; so, as the principle of relativity would have it, he will have no means of knowing whether he is at rest or in absolute motion.
> Unhappily, that does not suffice, and complementary hypotheses are necessary; it is necessary to admit that bodies in motion undergo a uniform contraction in the sense of motion. One of the diameters of the earth, for example, is shrunk by one two-hundred-millionth in consequence of our planet's motion, while the other diameter retains its normal length. (Poincaré 1958, 99-100)

This passage clearly shows that, according to Poincaré, the exchange of light signals between the systems can be used to synchronize clocks, with the travel time *expressed as* L/c for each trip, only if the clocks are at rest with respect to the ether. If the procedure is used between clocks in relative rest, but which move with velocity v with respect to the ether, it synchronizes the clocks with respect to 'local time', but not with respect to the real time. If the moving-with-respect-to-the-ether observers in A and B want to achieve *real* synchronization, the procedure must be such that if the station A sends the signal at time 0 in its clock, the clock in station B must set its clock to the time $\frac{L\left(\sqrt{1-v^2/c^2}\right)}{c+v}$, and when the signal returns to A its clock should read $\frac{L\left(\sqrt{1-v^2/c^2}\right)}{c-v}$. However, this procedure cannot be applied, for the length contraction effect on the measurement of distance and the effect of local time on the readings of the clocks—which taken together determine that the velocity they measure for the light signals is always c—entail that the results obtained will be the same as if they were at rest in the ether. It is clear that this line of reasoning is not relativistic at all, for it presupposes a Euclidian-space-through-absolute-time, and an immobile ether which determines certain dynamical effects that 'deceive' observers in motion with respect to it.

2.3 SPECIAL RELATIVITY

Now that the essentials of Lorentz's ether theory have been presented, we can consider the other theory involved in the first case of EE we will address. In this section I offer an exposition of the special theory of relativity as presented by Einstein in his two famous papers of 1905. First I will present what were the motivations and context underlying Einstein's work. Then I will examine his revolutionary reformulation of the concept of simultaneity and the physical effects that result from it. In the third subsection I will describe how the Lorentz transformations follow from the two principles and the new concept of simultaneity. Finally, I will shortly deal with the derivation of Einstein's famous equation $E = mc^2$.

2.3.1 Motivation and the two principles

As I mentioned at the beginning of this chapter, there was a time when an essential connection was stated between the Michelson-Morley experiment and SR. That is, the former was thought to have been a direct motivation for Einstein to develop his theory. After the work of historians like Hirosige, Holton, Stachel, Miller and others, this view has finally been shown to be wrong. Einstein's motivation was not

to solve the problem of the ether. What really took him to create a radically new theory were some foundational issues that he, unlike the rest of the scientific community—of which he was not a part ca. 1905—understood as problematic.

More specifically, the main problematic feature he found out was that

> It is known that Maxwell's electrodynamics –as usually understood at the present time– when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between these two cases in which either the one or the other of these bodies is in motion. (Einstein 1905b, 37)

If a conductor is considered as at rest in the ether, and a magnet moves with respect to it, then an electric current is induced; and if the magnet is considered as at rest in the ether and the conductor is in motion, exactly the same result obtains—an electric current is observed. However, electromagnetic theory provides a different explanation in each case. In the first one, the motion of the magnet creates an electric field which in turn causes the current; whereas in the second case, the current is the result of an electromotive force in the conductor produced by its motion with respect to the magnet, but no electric field is involved. The problematic aspect, for Einstein, was that in his view the only relevant feature for the phenomenon produced—the electric current—was simply the relative motion of the bodies involved. Such motion is totally symmetrical, of course; nevertheless, the theoretical explanation that electrodynamics provided was asymmetric.

After this example of the theoretical asymmetries that do not correspond to the observed phenomena, Einstein briefly refers to 'the unsuccessful attempts to discover any motion of the earth relatively to the light medium', and states that these failed attempts 'suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest' (ibid). These two observations are the only reasons that Einstein mentions in his 1905 paper as leading him to the formulation of his *relativity principle*. That the only specific experiment he mentions is the one of electromagnetic induction and his generic reference to the failed attempts to detect ether-wind effects clearly point out that the Michelson-Morley experiment had no special relevance for the formulation of the theory. After all, this experiment is only one more among the unsuccessful attempts to discover any motion of the earth relatively to the light medium. This does not mean that the experiment was totally irrelevant for Einstein, of course; it only means that the problematic issue that really motivated him to create a new theory was a much more general one, a foundational flaw –of which the Michelson-Morley experiment was yet another instance–, rather than a specific empirical problem. Actually, the explanation that Lorentz's theory provided for its null result can easily be conceived in analogous terms with respect to the magnet-conductor case: in the ether rest frame, the pattern of interference obtained depends only on the velocity of light and the length of the arms of the interferometer; whereas in the moving frame, the very same interference pattern depends also on the length contraction and on local time, but the only observable difference is the relative motion of the interferometers[57].

Einstein, as a conclusion of his analysis of this foundational problematic issue, introduces the first of the two basic principles of the theory, the *relativity principle*: 'the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good' (ibid, 37-8). That is, all the laws of physics have the same form in all inertial frames. In classical mechanics this principle

[57] In spite of some comments that Einstein made many years later, claiming that he did not know about the Michelson-Morley experiment by 1905, there are good reasons to think that he actually read Lorentz's 1899 work. In that case, it is clear that he knew about it. Moreover, there is a recently discovered document—notes taken from a talk that Einstein gave in 1921 at the Parker school in Chicago—which strongly indicates that, some time around 1899, he knew the experiment (see van Dongen 2009). However, it is also clear that he did not assign any special importance to it. For a historical survey of this issue, see (Holton 1988), (Stachel 1982), (Stachel 2002) and (van Dongen 2009).

had been assumed and tested without any problems, but in electrodynamics, the fact that Maxwell's equations were interpreted as valid for the ether-rest frame implied that they should change their form in a frame in motion through the ether. In the previous section I showed how Lorentz's theory was able to cope with the complete failure of detecting any effects of the Earth's motion through the ether, and how Poincaré related this achievement to *his* principle of relativity The difference between Einstein's and Poincaré's relativity principles consists in that, for the latter, it was the result of empirical tests and their theoretical explanation; that is, it was the outcome of dynamical features like the Lorentz-Fitzgerald contraction and local time. In the case of Einstein, the principle, even though empirically suggested, takes the form of a constrictive axiom of the theory, i.e., it is not an assumption to be explained. I will return to this difference below.

After these considerations Einstein—rather abruptly and without any further justification—introduces the second principle of the theory, the *light principle*: 'light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body' (ibid, 38). The lack of further comments by Einstein on his motivations to introduce this principle might suggest that the failed experiments on ether-wind effects also counted as the drive behind it, or that the first principle itself led him to the second. However, Stachel (1982; 2002) has compellingly shown that this is not the case. The real motivation for the formulation of the principle was grounded in Einstein's thoughts on the electromagnetic nature of light. His famous *light-rider* thought experiment suggested him that a light ray looks the same to any observer, regardless of his velocity with respect to the emitting source. In a 1912 letter to Paul Ehrenfest, Einstein wrote:

> I well knew that the principle of the constancy of the velocity of light was something quite independent of the relativity principle; and I weighed which was more probable: the principle of the constancy of c, as required by Maxwell's equations, or the constancy of c exclusively for an observer at rest with respect to the source of light. I decided for the former, because I was convinced that any light is completely defined by frequency and intensity, quite independent of whether it comes from a moving light source or one at rest. It did not even enter my mind to imagine a deflected radiation propagated through a point could behave differently than radiation newly emitted at the point in question. Such complications seem to me much more unreasonable than those which the new concept of time involves. (Quoted in Stachel 1982, 51)

The second principle, just like the first one, is also a constraining axiom. Even though empirically justified –Einstein did not so consider it, but the negative results of ether-wind experiments in optics could be interpreted as empirical support for it– it is not an assumption that requires an explanation. Once again, this differentiates Einstein's approach from Lorentz's and Poincaré's. In their theory, the *measured* velocity of light is c in all inertial frames; but in the ones which move with respect to the ether this measurment is a 'deception', for it is the outcome of compensating-conspiring dynamical effects. The velocity of light is *really* c only in the ether-rest frame.

2.3.2 Relative simultaneity

Taken in isolation, both principles are rather natural even from the point of view of classical mechanics and electrodynamics. The first one is part of the very core of classical mechanics; and in electrodynamics, as Lorentz and Poincaré showed, it obtained as the outcome of compensating dynamical effects. The second principle is totally consistent with electromagnetic theory, but only for one special inertial frame, the ether-rest frame; only in it light is emitted with constant velocity c regardless of the state of motion of the source. In any other inertial frame, the speed of light must respect the Galilean law of addition of velocities: it is the sum of c as defined in the ether-rest frame and the velocity of the moving frame with respect to the ether. But if the two principles are taken together it follows that the speed of light is c in

any inertial frame—and therefore, there is no privileged inertial frame any more—regardless of their relative velocity and regardless of the state of motion of the source. This is a blatant violation of the classical expression for the addition of velocities that I just mentioned.

However, Einstein noticed that the contradiction is only apparent; there is no real contradiction at all. The law for the addition of velocities, Einstein found out, rests upon a certain definition of distant simultaneity that is reflected in the methods to determine it, and that presupposes physical assumptions:

> If we wish to describe the motion of a material point, we give the values of its co-ordinates as functions of the time. Now we must bear carefully in mind that a mathematical description of this kind has no physical meaning unless we are quite clear as to what we understand by "time". We have to take into account that all our judgments in which time plays a part are always judgment of *simultaneous events*. If, for instance, I say, "That train arrives here at 7 o'clock", I mean something like this: "The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous".
> […] In fact such a definition is satisfactory when we are concerned with defining a time exclusively for the place where the watch is located; but is no longer satisfactory when we have to connect in time series of events occurring at different places, or—what comes to the same thing—to evaluate the times of events occurring at remote places from the watch. (Einstein 1905b, 39)

This passage shows that Einstein realized that the determination of the simultaneity between two *distant* events is an *inference*, rather than an *a priori* relation defined in terms of absolute time[58]. In order to provide a sound method according to which simultaneity between distant events can be determined—and that does not require any knowledge of the distance between the events or between the clocks—Einstein proposes his famous method to synchronize distant clocks:

> If at the point A of space there is a clock, an observer at A can determine the time values of events in the immediate proximity of A by finding the positions of the hands which are simultaneous with these events. If there is at the point B of space another clock in all respects resembling the one at A, it is possible for an observer at B to determine the time values of events in the immediate neighborhood of B. But it is not possible to compare, in respect of time, an event at A with an event at B. We have so far defined only an "A time" and a "B time". We have not defined a common "time" for A and B, for the latter cannot be defined at all unless we establish *by definition* that the "time" required by light to travel from A to B equals the "time" it requires to travel from B to A [this *definition* is exactly what the second postulate allows]. Let a ray of light start at the "A time" t_a from A towards B, let it at the "B time" t_b be reflected at B in the direction of A, and arrive again at A at the "A time" t'_a.
> In accordance with the definition the two clocks synchronize if $t_b - t_a = t'_a - t_b$. We assume that this definition of synchronism is free from contradictions, and possible for any number of points. (ibid, 39-40)

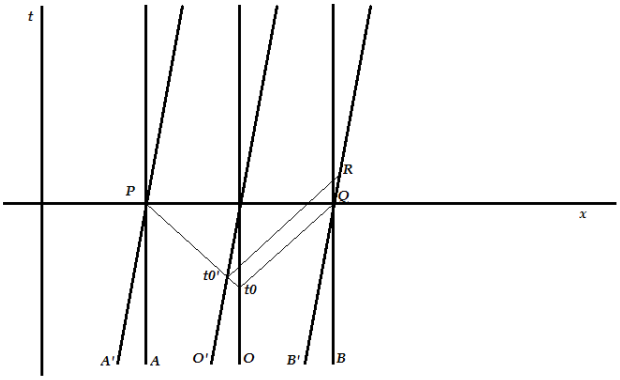
This method for the synchronization of distant clocks, and to determine distant simultaneity, is the key to dissolve the 'contradiction' between the postulates. The method is *by definition* consistent with the two principles, for the value of the velocity of light is assumed to be c in *all* inertial frames[59]. Moreover, as we will see, the kinematics this definition of simultaneity entails –along with the two principles– contain a law for the *composition* of velocities which is not the Galilean one –though when v is much smaller than c

[58] With respect to Einstein's motivations and philosophical background for this insight, Stachel comments: 'Here, I believe, Einstein was really helped by his philosophical readings. He undoubtedly got some help from his readings of Mach and Poincaré, but we know that he was engaged in a careful reading of Hume at about this time; and his later reminiscences attribute great significance to his reading of Hume's *Treatise on Human Nature*. What could he have gotten from Hume? I think it was a relational –as opposed to an absolute– concept of time and space. This is the view that time and space are not to be regarded as self-subsistent entities; rather one should speak of the temporal and spatial aspects of physical processes; 'The doctrine,' as Hume puts it, 'that time is nothing but the manner in which some real object exists'. I believe the adoption of such a relational concept of time was a crucial step in freeing Einstein's outlook, enabling him to consider critically the tacit assumptions about time going into the usual arguments for the 'obvious' velocity addition law' (Stachel 2002)

[59] 'In agreement with experience [and with the second postulate] we further assume the quantity $\frac{2AB}{t'_a - t_a} = c$, to be a universal constant –the velocity of light in empty space' (ibid, 40).

the relativistic law reduces to the Galilean formula– and that is totally consistent with the principles of the theory.

A very important consequence that follows from the two postulates and the synchronization method just presented is the *relativity of simultaneity*. That is, a statement like *two events are simultaneous* becomes meaningful only *with respect to a specific inertial frame*. If we consider two inertial frames in relative motion, two events that are simultaneous in one of them will not be so in the other one. This result can be better visualized in the following figure:



A, O and B are three observers at rest with respect to each other, the distance from O to A is the same as the distance from O to B; and the three of them carry synchronized clocks according to the method just described. A', O' and B' are three other observers at rest with respect to each other, and A' and B' are equidistant with respect to O'. As the diagram shows, these three observers are in inertial motion with respect to A, O and B; and they are also equipped with clocks synchronized by the same method. At t_0, O emits a light ray towards A and another towards B. At $t_0{}'$ the light ray from O to A reaches O', and at that same instant O' emits a light ray towards A' and another towards B'. Since A and B are equidistant with respect to O, the events which coincide with the arrival of the light rays sent from O at t_0 are simultaneous, i.e., when the light ray reaches A, its clock marks the same time as the time that the clock in B marks when it receives the other light ray. On the other hand, since A' and B' are equidistant with respect to O', the events which coincide with the arrival of the light rays sent from O' at $t_0{}'$ are simultaneous; when the light ray reaches A' its clock marks the same time as the time that the clock in B' marks when it receives its corresponding light ray. The events at which the light rays towards A and A' arrive are the same, namely, P. However, the events Q in B and R in B' that coincide with the arrival of their light rays are not the same. The situation is then that according to the clocks in A, in O and in B, P and Q are simultaneous events; but according to the clocks in A', in O' and in B', the events P and R are simultaneous. The explanation for this discrepancy is that simultaneity is relative. Simultaneous events for observers A, O and B are not so for the observers A', O' and B'—and the other way around—because they are in relative inertial motion[60].

The relativity of simultaneity has consequences also for how the lengths of bodies are conceived in SR. The length of an object, say, a rod, is defined as the spatial distance between two simultaneous events: the ones that coincide with its end points. But we just saw that inertial motion between two frames determines that the events which are simultaneous in one of them will not be so in the other one. Therefore, the events which define the length of an object in one of the frames are not the same events which define

[60] Or more precisely, since the inertial rest frame of A, O and B is in a state of inertial motion with respect to the inertial rest frame of A', B' and O'; simultaneous events in one of them are not so in the other. This statement also precludes any misunderstanding about the subjectivity of the observers as being involved in this issue. Observers could be totally dispensed of and the result would be the same.

the length of the same object in the other frame. This means that the relativity of simultaneity implies the relativity of length.

2.3.3 Lorentz transformations derived

These considerations are mainly qualitative. Einstein's next step was to obtain the specific quantitative transformations that relate coordinates of events in one inertial frame with their corresponding coordinates in a different frame. The first postulate can be stated as saying that any experiments whatsoever will have the same results in any inertial frame. This implies that result of an experiment will be the same even if its initial conditions differ only in terms of a translation and/or rotation in some inertial frame. Moreover, identical experiments carried out in inertial frames at different times will also yield the same outcomes. These two remarks mean that the first postulate implies the homogeneity and isotropy of space and time:

> It will also turn out, as a direct consequence of the relativity principle, that all inertial frames are spatially homogeneous and isotropic, not only in their assumed Euclidean geometry but for the performance of all physical experiments. By this we mean that the outcome of an experiment is the same whenever its initial conditions differ only by a translation (homogeneity) and rotation (isotropy) in some inertial frame. [...] Again, as a consequence of the relativity principle, it will presently turn out that inertial frames are temporally homogeneous, i.e., that identical experiments (relative to a given inertial frame) performed at different time yield identical results. (Rindler 1991, 6-7)

The spatiotemporal homogeneity and isotropy implied by the relativity postulate lead to a constraint in the nature of the coordinate transformations connecting events in different inertial frames: the transformations must be linear[61]. If they were not so, then, for example, the coordinates of a freely moving particle in one inertial frame, when transformed to a different one, would not yield an inertial description of its state of motion in the second frame; and this would be a blatant violation of the first principle of the theory[62].

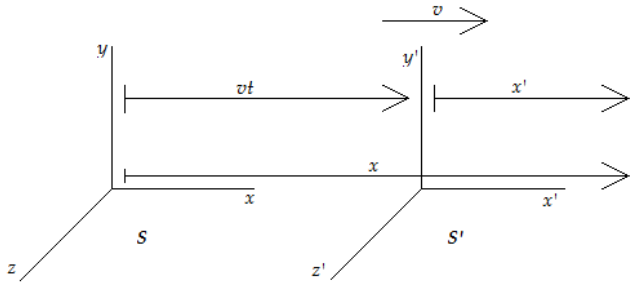
To obtain the transformations in a simple way a harmless assumption can be made: the frames to be related by the transformations can be considered as in *standard configuration*, that is, they satisfy the following conditions: *i)* the motion of the frame S' occurs only along the x-axis of the frame S, and that the planes defined by the $y = 0$ and $z = 0$ coordinates of the frames coincide with the planes defined by $y' = 0$ and by $z' = 0$; *ii)* when the origins of the two frames coincide the time coordinate of that event in both frames is 0; *iii)* that the coordinate plane in the moving frame S' defined by $x' = 0$ coincides with the plane in S defined by $x = vt$—where v is the velocity of S' with respect to S; and *iv)* the transformations are invariant under a reversal of the x and z-axes in both frames and an interchange of primed and unprimed coordinates—the same holds for reversal around x and y-axes. In simple words, the frames considered for the derivation of the transformations are two identical frames S and S'—with their origin and

---

[61] Einstein's statement that the linearity follows from the homogeneity and isotropy of space and the homogeneity and isotropy of time, taken separately, has been challenged. According to Torretti (1996, § 3.6), for instance, the correct view is that the homogeneity and isotropy of *space-time* is what really entails the linearity of the transformations. For example, it can be shown that the coordinate transformations between a resting and a uniformly rotating system are *not* spatiotemporally homogeneous, even though they are spatially homogeneous and temporally homogeneous—and that the transformations are not spatiotemporally homogeneous is rather natural in this case, for the second system is accelerated. Spatiotemporal homogeneity requires that a space-time location-independent variation applied to the coordinates of one system leads to a space-time location-independent variation in the coordinates of another. This is stronger requirement than the conjunction of spatial homogeneity (that a spatial location-independent variation in the coordinates of one system leads to a spatial-location-independent variation in another) and temporal homogeneity (that a time-independent variation in the coordinates of one system leads to a time-independent variation in another).

[62] For a formal proof that the transformations have to be linear, see (Rindler 1991, 11).

their three axes coinciding—but such that at the instant 0 of both frames the frame S' is set in uniform motion with velocity v with respect to S along its x-axis:



If we begin for the transformation for the y to y' coordinates, the linearity condition states that $y' = Ax + By + Cz + Dt + E$, where the coefficients are v-dependent constants. The first assumption above implies that if $y = 0$, then $y' = 0$; therefore $y' = By$. Applying the x-z invariant reversal we have that $y = By'$, and then $B = \pm 1$; but if the velocity v tends to 0 the transformation must lead to an identity transformation, so that B can only be 1. The resulting transformation is then $y' = y$; and completely analogous reasoning leads to $z' = z$.

Turning now to the x and x' coordinates, we have that because of linearity, $x' = Ax + By + Cz + Dt + E$. Assumption *iii)* above implies that if $x = vt$, then $x' = 0$; and therefore the transformation reduces to $x' = A(x - vt)$ and an x-y reversal yields $x = A(x' + vt')$. In a classical mechanics scenario, $t' = t$, and that would lead to $A = 1$, which in turn yields the Galilean transformations. However, Einstein's second postulate entails that the coordinates of the same light ray in S and S' are given by $x = ct$ and $x' = ct'$ respectively. These expressions can be plugged in the transformations so that one obtains $ct' = At(c - v)$ and $ct = At'(c + v)$. Then, the following equation can be set, $ct \cdot ct' = At'(c + v) \cdot At(c - v)$, and solving for A it follows that $A = \frac{1}{\sqrt{1-v^2/c^2}}$. It is quite clear that the coefficient A is identical to the factor γ in the Lorentz transformations. Finally, t' can be determined by replacing ct', ct, and x/c, for x', x, and t respectively in $x' = A(x - vt)$, so that $ct' = A(ct - v\frac{x}{c})$; and solving for t' it follows that $t' = A(t - \frac{vx}{c^2})$.

Summarizing, and adapting the notation, from Einstein's two postulates—plus the assumptions about the configuration of the frames—a set of coordinate transformations follow which are mathematically equivalent to the Lorentz transformations:

$$x' = \gamma(x - vt), \qquad y' = y, \qquad z' = z, \qquad t' = \gamma(t - vx/c^2);$$

and assumption *iv)* above entails, just as Poincaré showed, that the transformations are symmetric, i.e.:

$$x = \gamma(x' + vt'), \qquad y = y', \qquad z = z', \qquad t = \gamma(t' + vx'/c^2).$$[63]

In spite of their mathematical equivalence, the Lorentz transformations as conceived by Lorentz-Poincaré and as conceived by Einstein are rather different in their physical meaning. The main difference is that for the Dutch and the French they are the *outcome* of a set of dynamical effects, the Lorentz-Fitzgerald contraction and local time; whereas Einstein did not introduce any dynamical grounds for his *derivation* of the transformations. In a word, the Lorentz transformations, for Lorentz and Poincaré, are *dynamically*

[63] The derivation of the transformations I just presented is not the one that Einstein performed in § 3 of his 1905 paper, but a simpler one given in (Rindler 1982, 11-6). For a detailed analysis and commentary of Einstein's own derivation, see (Miller 1998, 195-205); and (Torretti 1996, § 3.4).

grounded; whereas for Einstein they are *kinematically* grounded. For the Dutch and the French, the interaction between matter and the ether underlies them. Einstein does not make any assumption about the ultimate nature of matter, the two postulates are enough; and therefore, the ether becomes superfluous[64]. This point of view allows Einstein, unlike Lorentz and Poincaré, to conceive that the only physically relevant velocity involved in the transformations is the *relative one* between the frames. As we saw above, even though Poincaré showed that the transformations are mathematically symmetric, the ether was still underlying their physical meaning. On this respect Miller comments:

> In the (1895) or (1904), Lorentz's plausibility argument for the Lorentz contraction hypothesis involved cross-multiplying the quantities $\sqrt{1-v^2/c^2}$ in the spatial portions of the S'' and S_r transformations, respectively. But since in special relativity K and k were equivalent, then Einstein could move between k and K by changing v to $-v$, and interchanging Greek and Roman letters. For Lorentz in 1904 this interchange had no physical meaning because the system K was fixed in the ether. In 1905 Poincaré attributed only a mathematical interpretation of the reciprocity property of the Lorentz transformations—that is, reciprocity corresponded to a rotation of K and k by 180° about their common y-axes. (Miller 1998, 204)[65]

With the coordinate transformations already presented, the meaning of the relativity of simultaneity and the relativity of length becomes much more precise and quantitative. Consider first the relativity of length. Frames S and S' are configured in the same way I assumed above. A rod of length $\Delta x'$ lies at rest in the x'-axis of S'. We want to find out what is its length Δx in the S frame. This length is given by the spatial distance between two simultaneous events in S which coincide with the end-points of the rod in S. The formula we need is the Δ-form of the transformation from x' to x, that is, $\Delta x' = \gamma(\Delta x - v\Delta t)$. The requirement of the simultaneity in S of the events which determine the length of the rod implies that $\Delta t = 0$. Therefore, the length of the rod Δx in S is equal to the length of the rod $\Delta x'$ in S' divided by γ. In other words, $L_s = \sqrt{1-v^2/c^2}\, L_{s'}$. This means that observers in S, for whom the rod is in motion, will measure it as contracted by the factor $\sqrt{1-v^2/c^2}$ with respect to its rest-length or *proper length* in S'—the frame in which the velocity of the body is 0 measures the largest possible length, and this largest possible length is called the *proper length* of a body. It is clear that the grounding of this contraction is the relativity of simultaneity, not dynamical effects such as the Lorentz-Fitzgerald contraction which results of the motion of objects across the ether. This can be noticed by considering that if the rod were at rest in S its length in S' would be contracted with respect to its proper length in S.

Now we consider the case of time in two different inertial frames in relative motion. Once again let us assume that S and S' are configured as before. Suppose that a clock w is fixed in S and that two events at that clock—their spatial coordinates are the same—are separated by Δt according to that clock. We want to find out what is the $\Delta t'$ between those two events as marked by a clock w' stationary in S'. We use the Δ-form of the transformation for the time coordinate $\Delta t' = \gamma(\Delta t - v\Delta x/c^2)$, and since $\Delta x = 0$ then $\Delta t' = \gamma\Delta t$. This means that for the observer in S for which the clock w' is in motion, such clock measures a time interval between the two events which is dilated by a factor $\sqrt{1-v^2/c^2}$ with respect to the *proper time* that her stationary clock w measures between the same two events—the frame in which the spatial distance between the events is 0 measures the shortest possible time-interval between them, i.e., their *proper time*. The formal expression for this *time dilation* effect is thus $T_{s'} = \frac{T_s}{\sqrt{1-v^2/c^2}}$. More generally, clocks that move in an inertial frame go slower than clocks at rest in that same frame. Just like length-contraction,

---

[64] In Einstein's own words: 'These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory of stationary bodies. The introduction of a 'luminiferous ether' will prove to be superfluous inasmuch as the view here to be developed will not require an 'absolute stationary space' provided with special properties' (1905b, 38).

[65] The frames S'' and S_r correspond to the frames S and S' in Lorentz's two-step method involving S_0, S and S', respectively. Einstein's K and k frames correspond to S and S' in the derivation I just presented, and the interchange between Greek and Roman letters corresponds to the interchange between primed and unprimed coordinates.

this is a *kinematically* grounded effect. There are no *dynamical* processes affecting the rates of the clocks. This can be easily seen by shifting the roles between the frames. In S the moving clock w' goes slower with respect to the stationary clock w, but in S' the moving clock w goes slower with respect to the stationary clock w'.

I will now mention some other consequences of the coordinate transformations that Einstein obtained, in order to make a case for the predictive equivalence of his theory with respect to Lorentz's. First, the expression for the composition of velocities: assume frames S and S' to be configured in the standard way, so that the x-velocity of a particle is given in S by $u_x = \frac{dx}{dt}$ and in S' by $u'_x = \frac{dx'}{dt'}$. From the coordinate transformations we have that $dx' = \gamma(dx - vdt)$, and that $dt' = \gamma(dt - vdx/c^2)$. Plugging the right hand of these equations in the expression for u'_x, then $u'_x = \frac{dx - vdt}{dt - v\frac{dx}{c^2}}$ obtains. Comparing this result with the expression for u_x, it follows that $u'_x = \frac{u_x - v}{1 - u_x v/c^2}$. If we remind that in his derivation Poincaré set c to 1, it is obvious that the velocity composition law he obtained is identical to Einstein's—which is rather natural since both were carried out from identical coordinate transformations.

This last result was used in 1907 by Laue to derive the Fresnel coefficient in a very simple way. Suppose a container filled with a transparent medium with a refractive index $n = c/u'$, where $u' = c/n$ is the velocity of light in the medium when it is considered at rest in the ether. Recall that Fresnel found out that if the medium is set in motion—with respect to the ether—with a velocity v, the velocity u of the light ray in the refractive medium is given by $u = u' + v(1 - 1/n^2)$, where the term in brackets is Fresnel's drag coefficient. What Laue did was to show that this last formula, in the context of SR, is nothing but a consequence of the velocity composition law. The derivation is quite simple and very meaningful. First, the ether-rest frame does not play any role, v is simply the relative velocity between frames S and S'. From this perspective, u is the velocity of light across the refractive medium as measured in S—for which the container moves with velocity v—and u' is the velocity of light in the refractive medium as measured in S'—for which the container is at rest. Therefore, $u = \frac{u' + v}{1 + u'v/c^2}$[66]. This last formula can be approximated, neglecting terms of second order of v/c, to $u = (u' + v)(1 - \frac{u'v}{c^2})$ and then to $u = (u' + v)(1 - \frac{u'^2}{c^2})$. Comparing the second factor with the formula for n, it turns out that $u = u' + v(1 - 1/n^2)$, in agreement with Fresnel's coefficient. The fact that this derivation was carried out only from the law of composition of velocities indicates that, in the context of SR, Fresnel's drag is a kinematically grounded effect that does not need any dynamical explanation in terms of the interaction between light and the medium on the one hand, and the ether on the other[67].

Now I will dedicate a few words for the velocity dependence of inertial mass in SR. Einstein's first step in his derivation was to consider a frame S in which an electron is at rest at a time t_0, but in motion at the *next instant of time* t_1—that is, both times differing by an infinitesimal amount. The motion of this electron in the frame S is described by Newton's formula $\mathbf{F} = m\mathbf{a}$, and the net force $\mathbf{F}$ is given by the influence of an external electric field $\mathbf{E}$ on the electron charge e. Therefore, the equations of motion for the electron in the frame S are

$$m_0 d^2x/dt^2 = e\mathbf{E}_x \qquad m_0 d^2y/dt^2 = e\mathbf{E}_y \qquad m_0 d^2z/dt^2 = e\mathbf{E}_z;$$

[66] This is, of course, the inverse velocity transformation, in which $-v$ replaces v, and in which the primed and unprimed quantities have shifted roles.

[67] Remember that though Lorentz's 1895 derivation of Fresnel's coefficient was rather general and it did not presuppose any electromagnetic features, in 1886 he had provided a derivation in that the drag of light was dynamically explained in terms of the interaction of the constituting molecules of the transparent medium and the ether. Anyhow, his 1895 derivation cannot be a rejection of this explanation, for it was crucial for his assumption of an immobile ether—the electromagnetic explanation he provided for Fresnel's coefficient was not based on an ether-partial-drag.

with the proviso that the motion of the electron is slow (the reason for this condition is to neglect any change that the relativity of simultaneity could produce in the formulation of the Newtonian law), and where m_0 is the inertial mass m in Newton's law.

The next step was to suppose that at the instant t_0 the electron is moving with velocity v with respect to the frame S, and that the frame S' is such that at that same instant the electron is at rest. According to the relativity postulate, at the *next instant of time* in S', the equations of motion for the electron will have the same form as the equations above—expressed in terms of x', y', z', and t'; of course. By applying the coordinate transformations to the equations of motion in S' it follows that in the frame S,

$$\frac{d^2x}{dt^2} = \frac{e}{m_0\gamma^3}\mathbf{E}_x \qquad \frac{d^2y}{dt^2} = \frac{e}{m_0\gamma}(\mathbf{E}_y - \frac{v}{c}\mathbf{B}_z) \qquad \frac{d^2z}{dt^2} = \frac{e}{m_0\gamma}(\mathbf{E}_z - \frac{v}{c}\mathbf{B}_y).$$

In order to obtain the expressions for the transverse and longitudinal mass of the electron, Einstein formulated the last equations in the following way:

$$m_0\gamma^3\frac{d^2x}{dt^2} = e\mathbf{E}_x = e\mathbf{E}_{x'}$$

$$m_0\gamma^2\frac{d^2y}{dt^2} = e\gamma(\mathbf{E}_y - \frac{v}{c}\mathbf{B}_z) = e\mathbf{E}_{y'}$$

$$m_0\gamma^2\frac{d^2z}{dt^2} = e\gamma(\mathbf{E}_z - \frac{v}{c}\mathbf{B}_y) = e\mathbf{E}_{z'}$$

The terms in the left hand of these expressions are the product of mass and acceleration, so that the formulas for the longitudinal and transverse mass become $m_L = m_0\gamma^3$ and $m_T = m_0\gamma^2$. If one compares the value of the *transverse* mass that Lorentz obtained with the one that Einstein derived it turns out that they are not the same. But the difference is only apparent. The middle and right-hand terms in the equations above represent the net force in the frame S and in the frame S' respectively, so that $\mathbf{F} = \mathbf{F}'$. As we saw above, in the case of Lorentz's theory, the equivalence between the forces in the frames only holds in the case of the x and x'-axes—and that is the reason why the *longitudinal* masses are equivalent in both theories. However, in Lorentz's theory the force equality does not hold in the two other cases. Therefore, the discrepancy between Lorentz's and Einstein's transverse masses is only a matter of a different definition of force. This is obvious when one considers that the equations above, for the y and z-axes, can be written as

$$m_0\gamma\frac{d^2y}{dt^2} = e(\mathbf{E}_y - \frac{v}{c}\mathbf{B}_z) \qquad m_0\gamma\frac{d^2y}{dt^2} = e(\mathbf{E}_y - \frac{v}{c}\mathbf{B}_z).$$

In this case, the value of the transverse mass, just as in Lorentz theory, becomes $m_0\gamma$; and the equality between $\mathbf{F}_y$ and $\mathbf{F}'_{y'}$ is no longer the case—the same holds for the equality corresponding to z and z'. But this is not a problem, for the forces in the two frames remain connected via the coordinate transformations. Therefore, the velocity dependence of the inertial mass is the same both in Lorentz's and Einstein's theories. However, notice that, unlike Lorentz, Einstein did not speculate about the ultimate nature of the electron mass, its rest-mass m_0 is just taken as a given.

2.3.4 Mass and energy

Finally, I will turn to the relation between mass and energy contained in SR. Einstein published this result in September 1905 in a paper entitled *Does the Inertia of a Body Depend upon its Energy Content?*, three

months after the publication of *On the Electrodynamics of Moving Bodies*. The derivation of the relation at issue was grounded on a result he had already obtained in the latter article. Suppose that a plane wave of light possesses the energy l in a frame S; and that the direction of the ray makes an angle φ with respect to the x-axis of S. Consider now a frame S' moving with velocity v with respect to S—and they are configured in the standard way. The value of the energy of the system of plane waves in S' is given by $l' = \frac{l(1-\frac{v}{c}\cos\varphi)}{\sqrt{1-v^2/c^2}} = \gamma l(1-\frac{v}{c}\cos\varphi)$.

Now Einstein proposes the following thought experiment: a body at rest in S simultaneously emits a light wave of energy $L/2$ at an angle φ with respect to the x-axis, and another light wave with the same energy in the opposite direction, so that after the emission the body remains at rest. E_0 is the energy of the body before the emission and E_1 is its energy after it. Considering energy conservation, then $E_0 = E_1 + L$. Turning to the description of this same situation in the frame S', and applying the formula for the transformation of energy of the light wave he had already obtained, it follows that in the S' frame $E'_0 = E'_1 + \frac{1}{2}L\gamma\left(1-\frac{v}{c}\cos\varphi\right) + \frac{1}{2}L\gamma\left(1+\frac{v}{c}\cos\varphi\right) = E'_1 + \gamma L$, where E'_0 and E'_1 are the energies of the body before and after the emission, respectively.

The difference between E'_0 and E_0, that is, $E'_0 - E_0 = E'_1 - E_1 + (\gamma - 1)L$, can be related to the kinetic energy of the body before and after the emission in S'. According to Einstein, the difference $E' - E$ is equal to the kinetic energy of the body in S', for the latter is defined as the difference between the body's energy when in motion and when at rest, which in this case can be represented as E' and E, respectively[68]; so that $K_0 = E'_0 - E_0$, and $K_1 = E'_1 - E_1$, where K_0 and K_1 are the kinetic energies of the body before and after the emission, respectively. Therefore, $K_0 - K_1 = L(\gamma - 1)$; which means that the kinetic energy of the body in S' diminishes as the outcome of the emission of the light wave. The connection of this result with the mass of the body follows from the fact that, neglecting terms of fourth and higher orders of *v/c*, $K_0 - K_1 = \frac{1}{2}\frac{v^2}{c^2}L$; and from this equation it follows that '*if a body gives off the energy L in the form of radiation, its mass diminishes by* L/c^2. The fact that the energy withdrawn from the body becomes energy of radiation evidently makes no difference, so that we are led to the more general conclusion that the mass of a body is a measure of its energy content' (Einstein 1905c, 71)[69].

2.4 COMPARING THE THEORIES

Now that both theories have been presented, I turn to an evaluative comparison between them. First I will deal with some subtleties that need to be considered if they are going to be understood as predictively equivalent. Secondly, I will outline why the two theories are *rival*s.

[68] The rationale for the possibility of considering E—which is the rest-energy of the body *in the frame* S—as the rest energy of the body *in the frame* S', relies on the relativity postulate: the energy of a body at rest must be the same in *any* frame.

[69] Einstein is assuming that, according to the binomial theorem, $\gamma = (1 - v^2/c^2)^{-1/2} = 1 + \frac{1}{2}(v/c)^2 + \frac{3}{8}(v/c)^4 + \cdots$; and that –with $v \ll c$– $K = \frac{1}{2}mv^2$; so that $K_0 - K_1 = \frac{1}{2}\frac{v^2}{c^2}L$ reduces to $m_0 - m_1 = L/c^2$ –where m_0 and m_1 are the masses of the body before and after the emission, respectively. It is important to underscore that the low-velocity assumption does not imply that mass-energy relation is approximate, for 'all the quantities in equation (7.9) [$m_0 - m_1 = L/c^2$] are measured in the body's rest frame. The relation between them cannot depend on the velocity of the auxiliary frame S' in which the kinetic energy is expressed. We are free to assign to S' any velocity we please. Hence the result is rigorously true' (Sartori 1996, 205).

2.4.1 Empirically equivalent

At first sight it looks pretty clear that Lorentz's ether theory and Einstein's SR are predictively identical. As I showed above, both predict an equal velocity dependence of inertial mass, both allow to derive Fresnel's drag coefficient, and they both entail a longitudinal length contraction and effects on the readings of clocks as the outcome of the state of motion of bodies. Even though these features have a different physical meaning in each theory, their mathematical forms and values are equal.

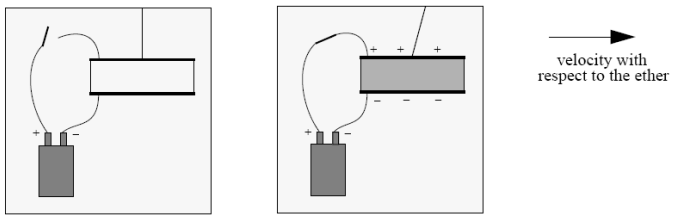
The root of this issue lies of course in the identical coordinate transformations that the theories include. Since the empirical predictions of both theories are entailed by these transformations it is quite natural that if the transformations are identical then the predictions will also be so. Moreover, if we consider that the very core of SR is given by its two postulates, we can find a sort of *Lorentzian* version of them in the other theory. As I showed in the sections dedicated to Poincaré's contributions, the French scientist explicitly formulated a *principle of relativity* that he regarded as a requirement that Lorentz's theory should accomplish in order to be fully satisfactory. That is, Poincaré thought that Lorentz's theory had to be formulated in a way such that the outcome of *all* experiments set out to measure any kind of ether-wind effect should be negative: Poincaré demanded a full Lorentz-invariance for the laws of physics. We saw that the main difference between Poincaré's and Einstein's relativity principle was that for the former it was the result of a set of dynamical-compensating effects that precluded the possibility to detect any sort of alteration in the laws of physics; whereas for Einstein it was simply an axiomatic constraint of the theory which did not need any kind of dynamical explanation.

Something similar holds for the light postulate. Einstein simply took the principle that light has a constant velocity c in any direction independently of the state of motion of the source as a point of departure. Together with the first postulate, it follows that the velocity of light is c in all directions in *any* inertial frame, not only in the ether-rest frame. I also showed that the commitment to these two principles requires a radical reconsideration of the nature of simultaneity and time. In the case of Lorentz's theory, the analogous feature with respect to the light postulate is that the same compensating dynamical effects that resulted in the Poincaréan version of the relativity principle, also determined that in any frame moving inertially with respect to the ether the *measured* value of the velocity of light is c. Its real velocity is $c + v$, but the dynamical effects inexorably *deceive* the observer, he has no way to find out its real value. One of the compensating effects that participate in this conspiracy is local time. Analogously to what happens in SR, in Lorentz's theory the events that observers in a frame that moves with respect to the ether describe as simultaneous are not so for observers in a different frame with a different velocity with respect to the ether. However, and unlike Einstein's theory, Lorentz's theory describes this effect as a *deception*, for the *real* time measurements can be carried out only in the ether-rest frame.

So far, so good, but what about $E = mc^2$? The relationship between mass and energy was not formulated as an explicit result in Lorentz's theory, and therefore this result must be carefully considered if the empirical equivalence of the theories is to be argued. Actually, if one considers only Lorentz's view of his own theory, the absence of the famous formula seems to imply that, after all, there is a crucial experiment to ground an empirically based decision—namely, the Trouton experiment.

In 1900 Frederick Trouton, based upon an original idea by George F. Fitzgerald, designed an ingenious experiment in order to look for an ether-wind effect. The interesting feature of this particular experiment was that, unlike many of the attempts to detect an effect of the motion of the Earth across the ether, it was not based on optics: according to Fitzgerald, if a capacitor in motion through the ether is charged or discharged it should suffer an impulse. Trouton's experiment was designed to measure this effect. In a terrestrial laboratory—which, of course, moves with respect to the ether—a hanging capacitor at rest is connected to a battery. When the battery is switched on an electromagnetic field is produced between the plates of the capacitor. If it were at rest in the ether, only an electric field would be induced, but its motion through the ether adds the generation of a magnetic field. Fitzgerald reasoned that the extra energy needed to produce the magnetic field had to come from the kinetic energy of the capacitor, and the

loss of kinetic energy must result in a jolt in the direction opposite to the motion across the ether. The actual display of the experiment was such that the capacitor was a part of a torsion pendulum, and the charges and discharges were done by a clock-work at time intervals corresponding to the free period of swing of the pendulum, so that the jolt effect would accumulate with its natural oscillation and so become more easily observable. A simple sketch of the experiment is depicted in the following figure (taken from Janssen 2003, 31):



Just as in every ether-wind experiment performed, the result of Trouton's was negative, no impulse in the capacitor was observed. In any possible outcome, though, this experiment posed a deep theoretical challenge. Let us recall that Newton's third law is closely associated with the law of conservation of momentum and with the center of mass theorem. If the outcome of the Trouton experiment is positive, the jolt of the capacitor constitutes a clear violation of the center of mass theorem[70]; but if the result is negative, then the law of conservation of momentum is violated. The dilemma consists in that it seems that a theoretical explanation of the experiment necessarily implies the abandonment of one of these two core tenets of classical physics.

As I showed above, Poincaré had already found a similar problem. In 1900 he introduced a fictitious fluid in the ether carrying energy and momentum that allowed him to save the momentum conservation law and also the center of mass theorem—and with this maneuver he got very close to a formulation of the energy-mass relation. However, Lorentz rejected Poincaré's solution. The fact that a fictitious ether-fluid carries momentum means that the ether is in motion under certain conditions, and this would be at odds with the central assumption of an immobile ether in his theory:

> Lorentz's opinion of Poincaré's valiant attempt at saving the principle of action and reaction was, in his words: "I must claim to you that it is impossible for me to modify the theory in such a way that the difficulty that you cited disappears." Lorentz went on to emphasize several times that his ether acted on bodies but that there was no reaction on the ether. He explained that the "phenomena of aberration," that is, fist order effects, had "forced him" to assume a motionless ether […]. Lorentz continued in his letter: "I deny therefore the principle of reaction in these elementary actions." In mechanics, Lorentz continued, action and reaction were instantaneous because disturbances were not mediated by an ether; however, in electromagnetic theory the reaction of an emitter of radiation was not compensated simultaneously by the action on the absorber. Poincaré had avoided this problem by attempting to satisfy the principle of reaction separately by emitter and absorber. Consistent with his desire to maintain an absolutely immobile ether, Lorentz protested Poincaré naming the quantity in Eq. [$G = \frac{1}{4\pi c}\int E \times B\, dV$; the momentum of the *fictitious fluid*] which Lorentz compared to Poynting's vector, to be an electromagnetic momentum. To Lorentz the term momentum, of course, connoted motion. Lorentz was willing to concede only that Poincaré's electromagnetic momentum was formally "'equivalent' to a momentum". Thus, in the 1901 letter, Lorentz informed Poincaré of his own sensitivity toward adding further hypothesis to an already overburdened theory, especially theses invented solely to save a principle whose violation permitted the theory's formulation in the first place. (Miller 1986, 5-7)[71]

[70] This theorem states that no process in an isolated system can change the state of motion of the system's center of mass: for any system with no external forces, the center of mass moves with constant velocity.

[71] Miller's article contains a full reproduction of the letter of Lorentz to Poincaré, dated January 20th, 1901.

Accordingly, Lorentz's attitude towards the Trouton experiment was simply to hold one of the horns of the dilemma. His account of the experiment did consider an electromagnetic momentum, but interpreted as Abraham did in his 1903 *Principles of the Dynamics of the Electrons*. Abraham's electromagnetic momentum was carried by the electromagnetic field, and not by an ether-fluid, so that Lorentz's immobile ether was not threatened. Lorentz's explanation of the Trouton experiment was that the capacitor's loss of kinetic energy and momentum caused by the production of the magnetic field was compensated by the electromagnetic momentum carried by the magnetic field; so that the total momentum of the system as a whole remains constant. This explanation implies that the impulse in the capacitor does happen, and consequently, that the center of mass theorem does not hold. However, since Lorentz did not hesitate in abandoning Newton's third law in the name of all the achievements of his theory, this was not a problem for him:

> I take this opportunity for mentioning an experiment that has been made by Trouton at the suggestion of Fitzgerald, and in which it was tried to observe the existence of a sudden impulse acting on a condenser at the moment of charging or discharging; for this purpose the condenser was suspended by a torsion balance, with its plates parallel to the earth's motion. For forming an estimate of the effect that may be experienced, it will suffice to consider a condenser with ether as dielectricum. Now if the apparatus is charged there will be an electromagnetic momentum $\mathfrak{C} = \frac{2U}{c^2}w$ [where U is the energy of the charged condenser at rest and w is its velocity with respect to the ether] (terms of the third and higher orders are here neglected). This momentum being produced at the moment of charging and disappearing at that of discharging, the condenser must experience in the first case an impulse $-\mathfrak{C}$ and at the second an impulse $+\mathfrak{C}$. However Trouton has not been able to observe these jerks.
> I believe it may be shown (though his calculations have led him to a different conclusion) that the sensibility of the apparatus was far from sufficient for the object Trouton had in view. (From Lorentz's *Weiterbildung der Maxwellschen Theorie Elektronentheorie*, quoted in Janssen 2003, 37)

Besides Lorentz's abandonment of the action and reaction law—and of the theorem of the center of mass[72]—it must also be noticed that his account of the Trouton experiment would imply a violation of the principle of relativity as conceived by Poincaré, and hence, a breakdown of the EE with respect to SR. By 1904, Lorentz did not conceive his theory as an expression of the *full* invariance of the laws of physics under his coordinate transformations. In the same article in which he introduced his explanation of the Trouton experiment, he wrote that his goal was to show that: 'by means of certain fundamental assumptions, and without neglecting terms of one order of magnitude or another, that *many* electromagnetic actions are entirely independent of the motion of the system' (ibid, 38)[73]. In other words, Lorentz did not think that any way to detect ether-wind effects would fail from the outset. Therefore, in order to make a case for the predictive equivalence of the theories at issue an account of the Trouton experiment that respects the principle of relativity must be offered in the context of Lorentz's theory.

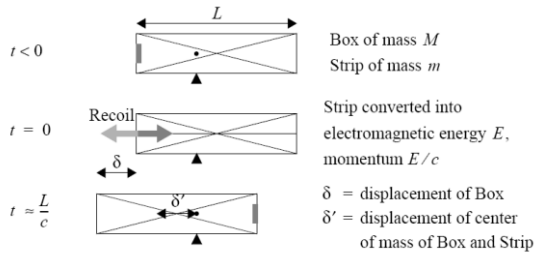
This explanation can be carried out only by considering $E = mc^2$. If energy has mass, a transfer of energy from the battery to the capacitor is also a transfer of mass from the former to the latter; and in a frame in which they are both moving, a transfer of momentum is involved as well. Therefore, by charging the capacitor, it gains an amount energy, mass and momentum, while the battery loses the same amount of these quantities. Total momentum is then conserved. However, if energy has mass, the momentum circulation within the system does not imply a change in the velocity of its parts, for the increase of the capacitor momentum is given by a *mass*-gaining, not by a change in its *velocity*. On the other hand, and from the point of view of Fitzgerald's interpretation of the experiment, that the extra energy needed to produce the magnetic field came from the kinetic energy of the capacitor, not from the battery, meant

[72] The action and reaction law is violated in Lorentz's theory because it includes an ether that affects ponderable matter but which is never affected by it. The center-of-mass theorem is violated because the jolt produced by the emission implies that the center of mass of the system gets accelerated, even though no external forces have been applied on the system.
[73] As I mentioned above, later on Lorentz—under the influence of Einstein, not of Poincaré—ended up interpreting his theory as in full predictive equivalence with respect to SR. See (Janssen 1995, section 3.5).

that the kinetic energy lost by the capacitor and which is taken away by the magnetic field is indeed accompanied by a loss of momentum of the capacitor which is compensated by the electromagnetic momentum of the field—just as Lorentz said. Nevertheless, the loss of momentum of the capacitor means that it loses mass, not velocity, so that the center of mass theorem still holds.

The connection between the Trouton experiment and the energy-mass relation becomes even clearer when one considers Einstein's second derivation of $E = mc^2$. In his 1906 *The Principle of the Conservation of Motion of the Center of Gravity and the Inertia of Energy*, Einstein established, by means of a thought experiment, that if the center of mass theorem is to hold in systems in which electromagnetic processes take place along with mechanical ones, it is necessary to consider the mass-energy relation. He considered a box of mass M and length L. In the left wall of the box an amount of energy E is stored. At time $t = 0$ the energy is converted into electromagnetic radiation and travels to the other side of the box, where it is absorbed and converted to its original form and stored in the other wall. The emission of the electromagnetic radiation produces a recoil in the box in the opposite direction of the motion of the wave; and when it is received in the opposite wall another recoil of the same magnitude occurs, so that the box is brought back to rest. Electromagnetic theory says that the momentum of the radiation is E/c. Conservation of momentum requires that the recoil experienced by the box has that same amount momentum. The thought experiment shows that a closed system can move itself. If the energy involved has no mass –as classical physics says–, then we are facing a violation of the center of mass theorem.

The way out of the problem is simply $E = mc^2$. If the energy E stored in the left wall has a mass $m \ll M$, then, before its emission, the center of mass of the system formed by the box plus E lies slightly to the left of the middle of the box. When the emission of E occurs, a mass m travels to the opposite wall, and when it is received and stored, the center of mass of the system has moved to somewhere slightly to the right of the center of the box. If δ' is the displacement of the center of mass, its value can be calculated from the condition which determines where a wedge supporting the system should have to be placed in order to keep it perfectly balanced, namely, $M\left(\frac{\delta'}{2}\right) = m\left(\frac{L-\delta'}{2}\right)$. The value of the center of mass displacement is then $\delta' \approx \left(\frac{m}{M}\right)L$. The center of mass theorem is respected only if δ' is exactly compensated by δ, the displacement of the box in the opposite direction between its two recoils. The time during which the box displaces is given by L/c, and the velocity of the displacement is given by its momentum divided by its mass, that is, $\frac{E/c}{M}$. Hence, $\delta \approx \left(\frac{E/c^2}{M}\right)L$. Then, $\delta = \delta'$ if and only if $(E/c^2) = m$. In other words, the center of mass theorem holds only if $E = mc^2$. The Trouton experiment is clearly an empirical instance of Einstein's thought experiment. Actually, in the former, a tiny recoil occurs when the capacitor is charged, but it is compensated by a displacement of the center of mass, and the theorem is respected. The following figure is a good depiction of these remarks (Janssen 2003, 40):[74]



[74] A full and precise explanation of the Trouton experiment in the context of Lorentz's theory that includes the energy-mass relation must consider some further subtleties. The definition of electromagnetic momentum involved is not Lorentz invariant, insofar as it includes a special reference to the ether-rest frame. The *hyperplane of simultaneity* that is the base for the integration over space that defines the electromagnetic momentum is the one that corresponds to the ether-rest frame. Accordingly, the explanation must also consider *Laue's effect*, namely, that stresses in the capacitor considered at rest give rise to momentum in a frame in which it moves. This momentum must be added to the electromagnetic one in order to explain Trouton's experiment. See (Janssen 2003, 44-9).

These statements clearly show that if Lorentz's theory is to be proved predictively equivalent to SR, then the energy-mass relation has to be considered. Accordingly, it must also be shown that this relation can be derived from Lorentz's theory. Otherwise, the implications of $E = mc^2$ could be used as a crucial experiment. Herbert Ives has shown that the famous equation is indeed contained in Lorentz's theory, and that Poincaré was very close to formulate it. Actually, in his 1906 paper, Einstein himself acknowledges that his derivation is quite similar to Poincaré's result in his 1900 criticism of Lorentz violation of Newton's third law. Ives offers an analysis of Poincaré's introduction of his fictitious fluid that underscores the mentioned similarity. Departing from the fact that the French scientist derived the expression $M = S/c^2$ for the fictitious fluid, where *M* is the momentum of the radiation, and *S* is the *flux* of radiation; Ives points out:

> Consider how Poincaré got his numerical result. He was using his formula, derived in his article, for the momentum radiation, $M = S/c^2$, and he was putting down the expression for the conservation of momentum in the recoil process. Putting μ for the mass of the recoiling body, and v for its velocity, his working equation is then $\mu v = S/c^2$. For *S*, the energy flux, he put the energy *E* times *c*. He then has $\mu v = S/c^2 = Ec/c^2 = E/c^2 \cdot c$.
> The significant thing for our present study is that Poincaré in his calculation used E/c^2 for the coefficient of *c* in stating the momentum of the radiation, that is E/c^2 plays the role of mass. The relation $E = m_R c^2$ was thus contained in his relation $M = S/c^2$. (Ives 1952, 540)

What Ives means by m_R is the mass equivalent for free radiation. Since in his 1900 paper Poincaré, as we saw above, established that the fictitious fluid was not indestructible in the sense that it could not be entirely transferred in the emission or absorption of energy, and hence it always had to appear as energy in other guises; then Poincaré precluded the possibility of interpreting m_R also as m_M. By m_M Ives designates the mass of matter, in opposition to the mass equivalent of free radiation. In other words, Poincaré was not in a position to interpret the expression $E = m_R c^2$ underlying his formula for the fictitious fluid momentum—even if he had actually seen that it was there—as an expression referred to the gain or loss of mass that matter experiences when it emits radiation.

However, Ives also points out that, had Poincaré tackled the issue from the point of view of *his* relativity principle, and even within the context of an ether-theory ontology, he had all the tools needed to obtain the mass-energy relation as referring both to m_R and to m_M:

> Consider a body suspended loosely, as by a nonconducting chord, in the interior of an enclosure, the whole system being stationary with respect to the radiation transmitting medium [the ether]. Let the body emit symmetrically in the 'fore' and 'aft' directions the amount of energy $\frac{1}{2}E$. The momenta of the two oppositely directed pulses cancel each other, the body does not move, and no information can be obtained as to its change of state.
> Now let the whole system of enclosure and suspended particle be set in uniform motion with respect to the radiation transmitting medium with the velocity *v*. the body now possesses the momentum $mv[1 - (v^2/c^2)]^{1/2}$ and the problem is to determine the effect on this momentum of the two emitted wave trains. Now the energy contents of the two wave trains emitted for the same (measured) period of emission, taking into account the change of frequency of the source and the lengths of the trains, are
>
> $$\frac{E}{2}\frac{[1+(v/c)]}{[1-(v^2/c^2)]^{1/2}} \quad \text{and} \quad \frac{E}{2}\frac{[1-(v/c)]}{[1-(v^2/c^2)]^{1/2}}.$$
>
> The accompanying momenta, from Poincaré's formula, are
>
> $$\frac{E}{2c^2}\frac{[1+(v/c)]}{[1-(v^2/c^2)]^{1/2}}c \quad \text{and} \quad \frac{E}{2c^2}\frac{[1+(v/c)]}{[1-(v^2/c^2)]^{1/2}}c.$$
>
> These being oppositely directed, the net imparted momentum is $Ev/c^2[1 - (v^2/c^2)]^{1/2}$. Forming the equation for the conservation of momentum we have
>
> $$\frac{mv}{[1-(v^2/c^2)]^{1/2}} = \frac{m'v'}{[1-(v^2/c^2)]^{1/2}} + \frac{Ev}{c^2[1-(v^2/c^2)]^{1/2}},$$

where v' is the velocity of the body after the emission of the radiation.
Now according to Poincaré's principle of relativity, the body must behave in the moving system just as in the stationary system first considered, that is, it does not change its position or velocity with respect to the enclosure, hence $v' = v$, and we get

$$\frac{(m-m')v}{[1-(v^2/c^2)]^{1/2}} = \frac{Ev}{c^2[1-(v^2/c^2)]^{1/2}},$$

giving exactly $(m - m') = E/c^2$, a relation independent of v, and so holding for the stationary system. The radiating body losses mass E/c^2 when radiating mass E. This is the relation $E = m_M c^2$.(ibid, 541)[75]

In simple words, what Ives' analysis shows is that—just as Einstein himself acknowledged—in the context of Lorentz's theory it is possible to carry out a derivation of $E = mc^2$ that is analogous to the one that Einstein performed in 1906—with the required provisos, namely, that electromagnetic momentum and the ether must be considered. Moreover, a much more simple derivation is available once Poincaré's stress in considered and plugged into Lorentz's expression for electromagnetic momentum, as I showed above.

I think that these remarks are enough in order to see that the mass-energy relation derived by Einstein in the context of SR is mathematically and physically contained in Lorentz's theory. It is true, though, that from a historical point of view the famous equation was not directly discovered by Lorentz or Poincaré, but it had to be *borrowed* from Einstein's results. However, in the context of this research, the fact that $E = mc^2$ was *conceptually* contained in Lorentz's theory is enough to make a complete case for the predictive equivalence of the theories at issue. On the other hand, the fact that many of the amendments, extensions and interpretations that Poincaré introduced are crucial to argue for the empirical equivalence, it seems more correct to state that the theories that are really predictively identical are SR and something like a *Lorentz-Poincaré theory*, rather than Lorentz's. It must be remarked that what we can dub the Lorentz-Poincaré theory (LPT) is a *conceptual* reconstruction which is only possible with the benefit of hindsight. This tag I propose is not meant to refer to a theory which existed from a *historical* standpoint. No textbook about it was ever written, for example.

Before turning to a treatment of the differences between the theories which ground their rivalry, I will briefly mention one argument that has been put forward in order to deny their empirical equivalence. Arthur Miller (1986, 232), for example, claims that Lorentz's theory cannot yield the relativistic Doppler effect. However, as Janssen points out (1995, section 3.3.5), if Poincaré's contributions are considered, that is, if the factor l is set to unity and if the transformations are understood as symmetric; then the relativistic expression for the Doppler effect does follow from the ether theory[76]. This remark is yet another reason to consider the *LPT* as empirically equivalent with respect to SR.

2.4.2 The rivalry

The first difference between Einstein's SR and the ether theory of Lorentz and Poincaré that I will address was already mentioned above. In this section I will simply state it in more precise terms. In a newspaper article he wrote in 1919, Einstein introduced a distinction between two kinds of scientific theories, namely, between *constructive* theories and theories *of principle*:

[75] One could object that this derivation rests upon the momentum $M = S/c^2$ carried by the fictitious fluid; and we saw the reasons that Lorentz gave for his rejection of this concept. However, the generality of Ives' analysis allows that the electromagnetic momentum *as understood by Abraham* could be used as well, i.e., as the momentum carried by the *electromagnetic field*—which added to the momentum of the recoil of the emitter gives the total momentum that is conserved.
[76] Janssen also mentions that Miller acknowledged this point.

> We can distinguish between various kinds of theories in physics. Most of them are constructive. They attempt to build up a picture of the more complex phenomena out of the materials of a relatively simple formal scheme from which they start out. Thus the kinetic theory of gases seeks to reduce mechanical, thermal, and diffusional processes to movements of molecules, i.e., to build them up out of the hypothesis of molecular motion. When we say that they have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question.
> Along with this most important class of theories there exists a second, which I will call 'principle-theories'. These employ the analytic, not the synthetic, method. The elements which form their basis and starting point are not hypothetically constructed but empirically discovered ones, general characteristics of natural processes, principles that give rise to mathematically formulated criteria which the separate processes or their theoretical representations of them have to satisfy. Thus the science of thermodynamics seeks by analytical means to deduce necessary conditions, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible.
> The advantages of the constructive theories are completeness, adaptability, and clearness, those of the principle theory are logical perfection and security of the foundations.
> The theory of relativity belongs to the latter class. In order to grasp its nature, one needs first of all to become acquainted with the principles on which it is based. (From Einstein's *My Theory*, quoted in Dieks 2009, 2)

From what has been said so far it is quite clear that the *principles* of SR are the relativity postulate and the constancy of light postulate. Just as Einstein's definition of a theory of principle states, these principles do not presuppose any conceptions about the ultimate nature of the processes they refer to; rather, they are general features—empirically suggested—that work as constraints the physical processes must satisfy. On the other hand, it is also clear that the LPT is a constructive theory, for some particular electrodynamical descriptions of matter and physical processes are the features that determine the picture of the world the theory provides.

Even though this is a very important difference between the theories, it is not sufficient to establish their rivalry. It is just a formal or schematic dissimilarity which does not determine that the LPT and SR are contending opponents. For instance –and using Einstein's own example–, though thermodynamics and the kinetic theory of gases are different in the sense at issue, and even though they offer an account of the same phenomena, they are not rivals, but complementary.

Actually, the early reception of SR and of Lorentz's theory was such that the principle-constructive difference between them was somewhat noticed; but most of the scientific community understood that the generalization that Einstein had introduced with respect to the achievements of Lorentz did not imply that they were contenders. This fact is quite apparent when one looks at the way in which the Kaufmann experiments were interpreted. In 1901 Walter Kaufmann started a series of experiments with β-radiation –electrons produced in radioactive decay and emitted with a velocity close to c– set out to test the v-dependence of the inertial mass of the β-particles. These experiments were considered as a way to empirically decide between three alternative theories: Abraham's theory, Bucherer and Langevin's, and the *Lorentz-Einstein* theory. That is, scientists, until around 1911, considered that the difference between the theories at issue was quite similar to the difference between thermodynamics and the kinetic theory of gases. This curious feature becomes understandable if one considers that since the reception of the theories occurred in the context of the Kauffmann's experiments, the scientific community paid more attention to the electrodynamical part of Einstein's paper than to its kinematical part[77].

The radically new approach to physics that Einstein's paper contained with respect to the notions of space and time were crucially developed and clarified by the work of Hermann Minkowski. In his famous paper of 1909 entitled *Space and Time*, he elucidated that Einstein's theory implied a revolutionary reformation of the meaning of these concepts by the introduction of a four-dimensional manifold that we now call *space-time*.

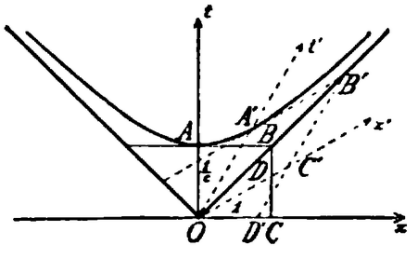
Minkowski's paper departs from a critical consideration of the way in which the geometrical and kinematical transformations of coordinates valid for Newtonian mechanics were usually regarded:

[77] See (Miller 1998, section 7.4).

> The equations of Newton's mechanics exhibit a two-fold invariance. Their form remains unaltered, firstly, if we subject the underlying system of spatial-coordinates to any arbitrary *change of position*; secondly, if we change its state of motion, namely, by imparting to it any *uniform translator motion*. […] Each of them by itself signifies, for the differential equations of mechanics, a certain group of transformations. The existence of the first group is looked upon as a fundamental characteristic of space. The second group is preferably treated with disdain, so that. […] Thus the two groups, side by side, lead their entirely apart. Their utterly heterogeneous character may have discouraged any attempt to compound them. But it is precisely when they are compounded that the complete group, as a whole, gives us to think. (Minkowski 1909, 75-6)

Minkowski's contribution was precisely to compound the transformation groups from the point of view of Einstein's theory. In simple words, he undertook a geometrical approach to kinematics. To achieve his goal, Minkowski proposed us to consider a point in space and in time—a *world-point*—in terms of a set of *Cartesian-like* coordinates x, y, z, t; so that 'the multiplicity of all thinkable x, y, z, t systems of values we will christen the *world*' (ibid). Minkowski's world is then a *four-dimensional space-time*. Then he pays attention to the path of one particular world-point. The differential of its coordinates that define that path determine his curve in the world, its *world-line*.

After these preliminary considerations, he introduces his crucial idea: 'The whole universe is seen to resolve itself into similar world-lines, and I would fain anticipate myself by saying that in my opinion physical laws might find their most perfect expression as reciprocal relations between these world-lines' (ibid). In order to materialize the idea he anticipates, he tells us to consider the positive parameter c and the graphical representation of the hyperbola $c^2t^2 - x^2 - y^2 - z^2 = 1$—or more precisely, he considers the sheet of the hyperbola in the region $t > 0$. He then considers 'those homogeneous linear transformations of x, y, z, t into four new variables x', y', z', t', for which the expression for this sheet in the new variables is of the same form'. The rotational and translational transformations, just as in Newtonian mechanics, are the ones that entail that $x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2$—provided that $t = 0$ determines the same instant in both systems. In the kinematic case, the transformations looked for—presupposing a standard configuration of the systems—are the ones that entail that $c^2t^2 - x^2 = c^2t'^2 - x'^2$. Minkowski concludes that the group of the rotational and translational transformations plus the kinematical transformations—the group G_c—for which the expression of the hyperbola retains its form depends on the parameter c. In order to illustrate his line of thought, Minkowski depicted this famous diagram:[78]



[78] "We consider the sheet in the region $t > 0$, and now take those homogeneous linear transformations of x, y, z, t into four new variables x', y', z', t', for which the expression for this sheet takes the same form. It is evident that the rotations of space about the origin pertain to these transformations. Thus we gain full comprehension of the rest of the transformations simply by taking into consideration one among them, such that y and z remain unchanged. We draw the section of this sheet by the plane of the axes of x and t—the upper branch of the hyperbola $c^2t^2 - x^2 = 1$, with its asymptotes. From the origin O we draw any radius vector OA' of this branch of the hyperbola; draw the tangent of the hyperbola at A' to cut the asymptote on the right at B'; complete the parallelogram $OA'B'C'$ […]. Now if we take OC' and OA' as axes of oblique coordinates x', t', with the measures $OC' = 1, OA' = 1/c$, then that branch of the hyperbola again acquires the expression $c^2t'^2 - x'^2 = 1$, $t' > 0$, and the transition from x, y, z, t to x', y', z', t' is one of the transformations in question. With these transformations we now associate the arbitrary displacements of the zero point of space and time, and thereby constitute a group of transformations, which is also, evidently, dependent on the parameter c. this group I denote by G_c". Lorentz *et al*. 1952, 77-8.

These kinematical-geometrical considerations acquire their *Einsteinean* physical meaning simply by identifying c with the velocity of the propagation of light. In that case, the kinematical transformations of G_c are, of course, the Lorentz transformations. Minkowski's own evaluation of the deep physical significance implied by his geometrical contribution was the following:

> The existence of the invariance of natural laws for the relevant group G_c would have to be taken, then, in this way:
> From the totality of natural phenomena it is possible, by successively enhanced approximations, to derive more and more exactly a system of reference x, y, z, t, by means of which these phenomena then represent themselves in agreement with definite laws. But when this is done, this system of reference is by no means unequivocally determined by the phenomena. *It is still possible to make any change in the system of reference that is in conformity with the transformations of the group G_c, and leave the expression of the laws of nature unaltered.*(ibid, 79)

The revolutionary conception of space and time implied by SR that Minkowski's contribution clarifies becomes quite clear by comparing it with the Newtonian framework. In the latter, the *place* in which two successive events occur is relative to a specific inertial framework, but the *instants* in which those events occur is an absolute feature. In Einstein's theory the Lorentz transformations entail that the time at which successive events occur is also frame-relative. Minkowski's space-time contributes to make clear that this relativity is inherently associated to the geometric-kinematic properties of a four-dimensional space-time whose invariants are determined by G_c.

A second important difference lies on the *invariant* interval between events of Minkowski's space-time. Whereas in the Newtonian world the geometrical and kinematic transformations leave intact the distance $\Delta x^2 + \Delta y^2 + \Delta z^2$ and leave intact the time interval Δt between to events; Minkowski shows that in a four-dimensional space-time governed by G_c the invariant interval between events is given by $c\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$, which is normally denoted by $\Delta \mathbf{s}^2$. Moreover, the Newtonian distance between E_1 and E_2 is greater than 0, or equal to 0 only if $E_1 = E_2$; but the Minkowskian 'distance' can be greater, equal or less than 0. If the $\Delta \mathbf{s}^2$ between two events is greater than 0, the 'distance' between them is called *time-like*, if it is equal to 0 it is called *null*; and if it is less than 0 the interval is *space-like*. Therefore, and with respect to a given event E, Minkowskian space-time gets divided in three regions: the set of events whose interval is time-like; the set of events whose interval is null, and the set of events whose intervals is space-like. The set of events whose 'distance' from E is null form E's light cone, the time-like events with respect to E are situated within its cone of light, and the space-like ones are outside the cone. This arrangement of events in turn reflects the causal structure of space-time. Only the events in the surface or inside the cone of light of E can be causally connected with it; in the former case the only possible connection is given by a light ray between them, in the later there is always the possibility of an observer passing between two time-like events. Events outside the cone of E cannot be causally connected with it.

This classification can also be applied to the 'lengths' of curves connecting events and to define the Minkowskian geodesics or straight lines. If a curve C that connects two events in space-time is such that its length $\int_C d\boldsymbol{s}$ is larger than the length of any other curve connecting the same two events, then C is a geodesic—if it is a time-like, null or space-like geodesic depends on the value of the corresponding $\Delta \mathbf{s}^2$, of course. In turn, geodesics in space-time allow a geometric account of inertia. Consider the world-line of a particle p. If p is a freely moving particle, then its inertial motion is represented by a time-like geodesic; and if the word-line of p is *not* a time-like geodesic, then its motion is accelerated.

After this brief revision of the main tenets of Minkowski's space-time, it is possible to assert that, in Einstein's theory, the Lorentz transformations express the structure and the metric properties of a Minkowskian space-time. This is a simple way to understand the main difference which underlies its rivalry with respect to the LPT. It is true that the LPT can be expressed in terms of a 'Minkowskian description'—

recall that Poincaré discovered the invariant interval and other four-dimensional features—but that description only express some *mathematical niceties* of the theory. The LPT does *not* describe the world as characterized by a Minkowskian structure and metric. The metric and geometry of the 'Lorentzian space-time' is still Newtonian[79]. The Lorentz transformations do not express geometric and kinematical features of that world, but the *dynamical* features which govern the behavior of objects inhabiting a Newtonian space-time. SR and the LPT are not rivals because the former offers a top-down explanation and the latter a bottom-up one. The rivalry is grounded in the fact that the constructive approach of the LPT lies upon a conception of the nature of space-time that is inconsistent with the corresponding conception contained in the principles of SR.

This way to characterize the difference and rivalry between the theories at issue allows a straightforward rejection of an interesting argument proposed by László Szabó (2011) that aims to show that SR and the LPT are the very same theory. He claims that

> According to [the] widespread view, special relativity was, first of all a new theory about space and time. A theory about space and time describes a *certain group of objective features* of physical reality, which we call (the structure of) space-time. Consider claims like these:
>
> - According to classical physics, the geometry of space-time is $\mathbb{E}^3 \times \mathbb{E}^1$, where $\mathbb{E}^3$ is a three-dimensional Euclidean space for space and $\mathbb{E}^1$ is a one-dimensional Euclidean space for time, with two independent invariant metrics corresponding to the space and time intervals.
> - In contrast, SR claims that the geometry of space-time is different: it is a Minkowski geometry $\mathbb{M}^4$.
>
> The two statements are usually understood as telling *different things about the same* objective features of physical reality. One can express this change by the following logical schema: Earlier we believed in $G_1(\widetilde{M})$, where $\widetilde{M}$ stands for (the objective features of physical reality called) space-time and G_1 denotes some predicate (like "of type $\mathbb{E}^3 \times \mathbb{E}^1$"). Then we discovered that $\neg G_1(\widehat{M})$ but $G_2(\widetilde{M})$, where G_2 denotes a predicate different from G_1 (something like "of type $\mathbb{M}^4$").
> *This is however not the case*. Our analysis will show that the correct logical schema is this: Earlier we believed in $G_1(\widehat{M})$. Then we discovered for some *other* features of physical reality $\widehat{M} \neq \widetilde{M}$ that $\neg G_1(\widetilde{M})$ but $G_2(\widetilde{M})$. Consequently, it still may (and it actually does) hold that $G_1(\widehat{M})$. In other words, in comparison with the pre-relativistic Galileo-invariant conceptions, special relativity tells us nothing new about the geometry of space-time. It simply calls something else "space-time", and this something else has different properties. We will also show that all statements of special relativity about those features of reality that correspond to the original meaning of the terms "space" and "time" are identical with the corresponding traditional pre-relativistic statements. Thus the only new factor in the special relativistic account of space-time is the terminological decision to designate something else "space-time".
> So the real novelty in special relativity is some $G_2(\widetilde{M})$. It will be also argued, however, that $G_2(\widetilde{M})$ does not contradict to what Lorentz claims. Both, the Lorentz theory and special relativity claim that $G_1(\widetilde{M})\&G_2(\widetilde{M})$. In other words: *SR and the Lorentz theory are identical theories about space and time in all sense of the words.* (Szabó 2011, 1-2)

It might be true that $G_1(\widehat{M})$ holds both in Einstein's and in Lorentz's theories. But according to what I have said above, I do not agree with the second part of Szabó's argument. It is not true that both theories assert $G_2(\widetilde{M})$. If by this statement we understand something like 'the physical world is characterized by the properties of the Minkowskian space-time', I just explained that the LPT does not claim that. When Szabó claims that 'then we discovered for some *other* features of physical reality $\widehat{M} \neq \widetilde{M}$ that $\neg G_1(\widetilde{M})$ but

[79] Minkowski's four-dimensional approach allows to construct a four-dimensional *Newtonian* space-time: 'If we now allow c increase to infinity, and $1/c$ therefore to converge towards zero, we see form the figure that the branch of the hyperbola bends more and more towards the axis of x, the angle of the asymptotes becomes more and more obtuse, and that in the limit this special transformation changes into one in which the axis of t' may have any upward direction whatever, while x' approaches more and more exactly to x. In view of this is clear that group G_c when $c = \infty$, that is the group G_∞, becomes no other than the group which is appropriate to Newtonian mechanics' (Minkowski 1909, 78-9).

$G_2(\widetilde{M})'$, I think that the right interpretation of this view is that the LPT claimed that $G_1(\widetilde{M})$—understanding this last statement as something like 'the physical world is a *Newtonian* space-time in which some dynamical features expressed by the Lorentz transformations govern the behavior of objects[80].

2.5 ON THE REASONS TO CHOSE

Now we are in position to deal with the epistemological problem involved. We have two physical theories which have the same empirical consequences. However, they are mutually inconsistent. Then, and under a thorough *hypothetical-deductive* conception of the confirmation of theories, the choice to be done between them is deeply underdetermined regarding empirical evidence. In this section I will propose an evaluative analysis of the possible reasons to make a choice. I will considered the reasons that have been considered in the received view concerning the case of Einstein vs. Lorentz, a much more recent argument introduced by Michel Janssen, and I will argue that an analysis of some historical facts in the development of early 20th century physics yield some reasons to choose that, to my knowledge, have not been addressed by philosophers and historians.

2.5.1 The Lorentz-Fitzgerald contraction and *ad-hoc*ness

In *The Logic of Scientific Discovery*, Karl Popper offered a definition of *ad hoc* hypotheses within the context of his falsificationist framework. A hypothesis is called *ad hoc* if it is unfalsifiable, that is, if the hypothesis does not entail any predictions that could put it 'at risk'. He mentioned the Lorentz-Fitzgerald length-contraction as a paradigmatic example of an *ad hoc* hypothesis—and this judgment became very influential regarding the epistemological assessment of Lorentz's theory. As we saw above, the Lorentz-Fitzgerald contraction was originally introduced with the specific goal of providing an account for the negative result of the Michelson-Morley experiment. If this were the only empirical prediction for which the hypothesis is logically relevant, then it would clearly qualify as *ad hoc* in Popper's sense. Since Popper stated as a methodological principle that the introduction of new hypotheses in a given theory is allowed only if such hypotheses increase the degree of falsifiability of the theory, then the *ad-hoc*ness of length-contraction is reason enough, according to Popper, to dismiss Lorentz's theory and favor special relativity:

> An example of an unsatisfactory auxiliary hypothesis would be the contraction hypothesis of Fitzgerald and Lorentz which had no falsifiable consequences but merely served to restore the agreement between theory and experiment—mainly the findings of Michelson and Morley. An advance was here achieved only by the theory of relativity which predicted new consequences, new physical effects, and thereby opened up new possibilities for testing, and for falsifying, the theory. (Popper 2002, 62-3)

Adolf Grünbaum (1959) showed that this view is mistaken. The Kennedy-Thorndike experiment does provide a potentially falsifying prediction if the Lorentz-Fitzgerald hypothesis is assumed. The difference

[80] A possible reply could be that if $\widetilde{M}$ is defined in terms of *measurements results*, then it could be said that both the theories assert that $G_2(\widetilde{M})$. However, at least in the Minkowski formulation of SR, $\widetilde{M}$ is not defined in that way. Moreover, this maneuver would be quite close to a logical-positivistic verificationist criterion of meaning, and my criticism to Szabó's argument can be taken as yet another example to show that, contrary to the claim of logical positivists, the meaning of a scientific theory does not reduce to its empirical consequences; these are only *a part* of that meaning. Actually, a logical positivistic-like semantic criterion seems to underlie Szabó's argument: the title of his paper is *Lorenztian theories vs. Einsteinian SR—a logico-empiricist reconstruction*. Moreover, when he refers to his methodological principles, he states that 'it is to be noted that our analysis is based on the following very weak operationalist/verificationist premise: physical terms, assigned to *measurable physical quantities*, have different meaning if they have different empirical definitions' (ibid, 20).

between the Kennedy-Thorndike and the Michelson-Morley experiment is simply that the two arms of the interferometer have different lengths, L and l. Grünbaum showed that both assuming and not assuming the length-contraction hypothesis, the expected outcome is a difference in the travel time for light rays along the different paths in the interferometer. That is, in both cases a positive result, testifying the effect of the motion of the earth through the ether, was expected. Grünbaum's main point is that the value for the difference in travel time is given by $\frac{2}{\sqrt{v^2-c^2}}(L-l)$ if length-contraction is assumed; but if it is not, the time-difference is given by $\frac{2}{\sqrt{v^2-c^2}}\left(L-\frac{l}{\sqrt{1-v^2/c^2}}\right)$. In other words, in the case of the Kennedy-Thorndike experiment, the Lorentz-Fitzgerald contraction hypothesis did entail a risky prediction.

Popper acknowledged this point, but he did not change his conclusion. He simply adjusted his view in terms of *degrees* of *ad-hoc*ness and concluded that Lorentz's theory was more *ad hoc* than Einstein's:

> Professor Grünbaum's correction shows that this hypothesis was testable and thus not *ad hoc* to the degree I believed. Accordingly *it was an advance*. But it was, of course, more *ad hoc* than special relativity. In other words, we have here an excellent example of 'degrees of *ad hocness*' and of one of the main theses of my book—that *degrees* of *ad hocness* are related (inversely) to degrees of testability and significance. (Popper 1959, 50)

But this whole line of reasoning is misconceived. It settles the issue by considering the length-contraction hypothesis in isolation with respect to the theory it belongs to. As mentioned above, the risky prediction that Grünbaum refers to is a *positive* result for the Kennedy-Thorndike experiment, but it turns out that one of the explicit goals of Lorentz's theory was to provide an explanation for the *negative* results obtained in all optical experiments to measure the effects of the earth's velocity with respect to the ether. That is, within Lorentz's theory the contraction hypothesis *must* entail a negative outcome for the Kennedy-Thorndike experiment—otherwise its observed result would remain unexplained. Such a prediction can be obtained only if the 'clock-retardation' effect is also considered, and this effect, in turn, results from the concept of 'local time'[81].

Grünbaum was aware of this point[82], but he simply concluded that 'only a version of the aether theory incorporating *both* the Lorentz-Fitzgerald contraction *and* the Lorentz-Larmor-Poincaré time dilation is vulnerable to the *ad hoc* charge, which tradition and Professor Popper have unjustly leveled against the first of these two auxiliary hypotheses alone' (Grünbaum 1959, 50). But if it is the ether theory including both the hypotheses (length-contraction and clock-retardation) what is subject to the *ad hoc* accusation, then the choice favoring special relativity holds, in spite of Grünbaum's argument.

It is not difficult, though, to demonstrate that the length-contraction hypothesis is not *ad hoc* in Popper's sense. As we saw above, in the 1895 version of Lorentz's theory the contraction hypothesis was indeed disconnected from the rest of the theoretical core of the theory. Thus, it was relevant only for the explanation of the negative result of the Michelson-Morley experiment, or, at best, for the explanation of the negative results of experiments of the same kind, as Grünbaum showed. Nevertheless, as also mentioned above, in the mature 1899 and 1904 versions of the theory the length-contraction hypothesis got generalized and connected to the theorem of corresponding states *via* the generalized contraction hy-

[81] In the 1895 version of the theory, the coordinate transformations that included the auxiliary quantity of 'local time' and the length-contraction hypothesis were two disconnected parts of the theory, so that in that version, a negative result for the Kennedy-Thorndike experiment was not possible. However, with the exact 1899 version of the theorem of corresponding states, in connection with the generalized contraction hypothesis, the clock-retardation effect required to explain the negative result of the Kennedy-Thorndike got interlocked with the length-contraction effect.

[82] 'Lorentz himself was quite clear that the contraction hypothesis would not suffice to make the body of experimental findings known at the time conform to the expectations of the aether theory. He therefore invoked the *further* auxiliary hypothesis known as the Lorentz-Larmor-Poincaré time dilation, construed as issuing a *spurious* 'local' time in all systems moving relatively to the ether' (Grünbaum 1959, p. 50).

pothesis—more precisely, the length-contraction hypothesis becomes a special case of the latter hypothesis. Thus, the range of experiments that could test the length-contraction hypothesis was no longer restricted to the Michelson-Morley type. Janssen underscores this feature of Lorentz's mature theory in order to dismiss *ad-hoc*ness accusations:

> the possibilities of testing the contraction hypothesis in its original form are thus very limited. This is not true of the generalized contraction hypothesis, which is part of Lorentz's mature theory based on the exact theorem of corresponding states. This theory gives definite predictions for any experiment used to test it. There is, of course, something peculiar about these predictions. The prediction is always that no effect of the earth's presumed motion through the ether will be detected. This is a problematic feature of the theory. But it has nothing to do with lack of testability. All that matter on that score is that the theory predicts a definite result. The generalized contraction hypothesis therefore passes Popper's falsifiability test with flying colors, whereas the original contraction hypothesis earned only a marginally passing grade courtesy of Roy Kennedy and Edward Thorndike' (Janssen 2002a, 433).

There is a very important exception, though, regarding Janssen's observation that all the predictions for which the length-contraction hypothesis is relevant consist in negative results in ether-wind experiments. As mentioned above, the theorem of corresponding states and the generalized contraction hypothesis are responsible for the derivation of the velocity-dependence of mass. Therefore, the hypothesis of length-contraction—which in the 1895 version of the theory could indeed be accused of *ad-hoc*ness in Popper's sense—becomes testable and falsifiable in a substantial way in the definitive version of Lorentz's theory. Actually, the hypothesis is relevant for the derivation of risky (and successful) predictions for the results of Kaufmann's experiments on the velocity-dependence of mass. Therefore, it is not justified to invoke *ad-hoc*ness of the length-contraction hypothesis in order to dismiss Lorentz's theory.

Besides Popper's, though, there remain two senses in which the length-contraction hypothesis might be accused of *ad-hoc*ness. First, it was *cooked up* with the only and specific goal of solving one single experimental difficulty. This is true, but as we just saw, further development of the theory connected the hypothesis with unexpected empirical results. In this sense, the length-contraction hypothesis is on an analogous stand with Planck's quantum of energy hypothesis—which was introduced with the only and specific goal of providing an account for the observed spectrum of black-body radiation—and in neither of the two cases this sense of *ad-hoc*ness could be invoked to reject the corresponding hypothesis, of course. Moreover, one could ask what is wrong with *ad hoc* hypotheses—in the sense of cooked up hypotheses—*in themselves*. Even if a hypothesis is helpful in providing an explanation for one single experimental result, it does contribute to enlarge the scope and fruitfulness of the corresponding theory[83].

The second remaining sense of *ad-hoc*ness that could be attributed to the length-contraction hypothesis is that it was a sort of *rabbit from the hat* maneuver, that is, that the contraction was postulated without a justified physical underpinning. In this case, the *ad-hoc*ness accusation boils down to an accusation of

---

[83] As Larry Laudan states it: '*a theory is ad hoc if it is believed to figure essentially in the solution of all and only those empirical problems which were solved by, or refuting instances for an earlier theory* […] Assuming that adhocness is understood in this way, we are entitled to ask: *what is objectionable about it*? If some theory T_2 has solved more empirical problems than its predecessor—*even just one more*—then T_2 is clearly preferable to T_1, and, *ceteris paribus*, represents cognitive progress with respect to T_1. […] In urging that adhocness (so defined) is a cognitive virtue rather than a vice, I am clearly not implying that ad hoc theories are invariably better than non-ad hoc ones. My claim, rather, is that an ad hoc theory is preferable to its non-ad hoc predecessor (which was confronted with known anomalies). […] But it might be argued that I have missed the point of the critics of adhocness. They might say yes, of course, T_2 is better than its *refuted* predecessor T_1; but the relevant comparison is between T_2 and some other theory T_n which is not ad hoc but still solves as many problems as T_2'. Einstein's special theory of relativity might exemplify T_n while the Lorentz modified aether theory was T_2. The obvious reply to such criticism is to ask why the admittedly ad hoc character of the Lorentz contraction constitutes a decisive handicap against it comparing it with special relativity. If the empirical problem-solving capacities of the two theories are, so far as we can tell, equivalent, then they are (empirically) on a par; defenders of the view that the adhocness of T_2 makes it distinctly inferior to T_n must spell out why, in such cases, the comparable problem-solving abilities and equivalent degree of empirical support can be thrown to the winds simply by stipulating that ad hoc theories are intrinsically otiose' (Laudan 1977, 115-7). See also (Grünbaum 1976, section 4).

implausibility. We saw above that Lorentz did propose a plausibility argument for his length-contraction hypothesis, and the fact that the same argument had been already envisioned by Fitzgerald bestows Lorentz's with good credentials. Moreover, the eventual fulfillment of the main goal of the electromagnetic worldview program—namely, the reduction of all physics to electromagnetism—would have provided the generalized contraction hypothesis with firm physical foundations.

However, as we saw above, in 1906 Poincaré clarified that the required corrections in Lorentz's model of the electron included the postulation of a non-electromagnetic quantity, Poincaré-pressure. This meant that the goal of the program of reducing all physical laws to electrodynamics was not fulfilled. This episode in the historical development of Lorentz's theory has been interpreted by Kenneth Schaffner (1974) as opening a flank for an *ad-hoc*ness accusation of what he calls the 'molecular forces hypothesis' (that strongly resembles Janssen's generalized contraction hypothesis):

> There were serious reservations about the satisfactory applicability of such a generalized M.F.H. to electrons. […] it was shown by Poincaré (1906) that the contractile electron could be considered a stable entity only if a definitively non-electromagnetic counter-pressure were invoked. To extend the M.F.H. to cover *this* type of force would violate the reduction thesis (of the M.F.H to electromagnetic forces) which provided the plausibility (and independent support) for the original M.F.H. (Schaffner 1974, 52).

It must be mentioned that Poincaré did offer a sort of plausibility argument for the non-electromagnetic stress he introduced—an argument that Schaffner does not refer to in his paper. After the mathematical formulation of his amendment to Lorentz's theory, Poincaré states that '*the pressure due to our supplementary potential is proportional to the fourth power of the experimental mass of the electron.* Since the Newtonian attraction is proportional to the experimental mass, one is tempted to infer that there exists a general relation between the causes giving rise to gravitation and those which give rise to the supplementary potential' (From Poincaré's *On the Dynamics of the Electron*, quoted in Miller 1986, 120).

It is true that this plausibility argument is not as good as the one offered by Lorentz regarding the molecular forces. However, it still *is* a plausibility argument. Moreover, in the context of Poincaré's goal—the construction of a theory within the spirit of the electromagnetic worldview—it is an argument that fits into that program. The last sections of Poincaré (1906) were dedicated to an attempt of reducing gravitation theory to electromagnetism. So, if the stress introduced could be shown to be gravitationally determined, it was not madness to think that it was therefore an electromagnetic feature in the end –and that would be shown by a successful electromagnetic account of gravitation. More generally, I think that the quoted passage shows that Poincaré was clear about the need of a justification for the force he introduced. He offered a tentative argument which was more a *promise* of an explanation than an actual explanation.

Anyways, the serious reservations, Schaffner argues, turned into an insurmountable problem with the advent of quantum physics. The conflict between electrodynamics and the quantum hypothesis was so deep that it became impossible to offer a physical-plausibility argument for the M.F.H.:

> If we examine the changing background assumptions of this time [early 1900s] I believe that we should see the weak rationale for some of Lorentz's hypotheses becoming even weaker. Consider the generalized M.F.H. discussed earlier. This hypothesis is ad hoc but it might be said to possess a slight degree of theoretical support on the supposition that Poincare's arguments concerning the non-electromagnetic nature of the force holding the electron together were misguided. If all forces were electromagnetic the generalization of the M.F.H. would not be ad hoc but might follow from a deeper classical theory. The calling into serious question of the fundamental assumptions of the Lorentz electron theory by light quanta made such ad hoc assumptions seriously ad hoc, by introducing in advance, counterarguments to any such deeper classical theory. (ibid, 75).

At this point, I will make only two brief comments about Schaffner's view. First, the problems with the M.F.H. do involve a non-empirical criterion that can be invoked to argue special relativity's superiority.

However, issues concerning implausibility of theoretical features are better assessed in an *explanatory* context rather than in terms of accusations of *ad-hoc*ness[84]. This is the standpoint that Michel Janssen takes on this matter. He has argued that the failure of the electromagnetic program implies that a central feature of Lorentz's theory—the Lorentz-invariance of *all* dynamical laws, electromagnetic and non-electromagnetic—remains an unexplained coincidence. I will undertake a detailed assessment of Janssen's position later on.

Second, the way in which Schaffner evaluates the relevance of the quantum hypothesis in the case of Lorentz vs. Einstein is, I think, misguided. It is true that the conflict between electrodynamics and quantum physics was a crucial factor in the historical course of events, and also that it provides a conceptually justified ground to definitively favor Einstein's theory. However, I disagree in that the friction with quantum physics is to be addressed in terms of *ad-hoc*ness. Since the conflict is related to a problem concerning some empirical tests of the quantum hypothesis, it opens a possibility to evaluate the choice to be made between the LPT and SR from an empirical evidence point of view. I will elaborate on this in the conclusions of this work.

2.5.2 Mathematic-aesthetic features

A second feature that could be used in order to make a choice favoring Einstein's theory is given by the aesthetic-mathematical virtues that characterize SR. Actually, these features did play a historically relevant role in the matter[85]. Consider, for example, the following passage included in Max von Laue's 1911 textbook (the first ever published) on Einstein's theory:

> Though a true experimental decision between the theory of Lorentz and the theory of relativity is indeed not to be gained, and that the former, in spite of this, has receded into the background, is chiefly due to the fact that, close as it comes to the theory of relativity, it still lacks the great simple universal principle, the possession of which lends the theory of relativity an imposing appearance. (Quoted in Schaffner 1974, 74)

Minkowski's introduction in 1908 of the four-dimensional formalism in which the theory can be expressed was a crucial factor regarding the general judgment about the mathematical simplicity of special relativity. Scott Walter (2010) has offered a historical survey of the early reception of Minkowski's work in connection with the acceptance of special relativity. He shows that Laue's specific reasons to embrace Einstein's theory was given—in spite of its difficult visualizibility in intuitive terms—by the mathematical elegance and simplicity that the four-dimensional formulation allowed:

> Laue considered Minkowski space-time as an "almost indispensable resource" for precise mathematical operations in relativity. He expressed reservations, however, about Minkowski's philosophy, in that the geometrical interpretation (or "analogy") of the Lorentz transformation called upon a space of four dimensions: "[A] geometric analogy can exist only in a four-dimensional manifold. That this is inaccessible to our intuition should not frighten us; it deals only with the symbolic presentation of certain analytical relationships between four variables". One could avail oneself of the new four-dimensional formalism, Laue assured his readers, even if one was not blessed with Minkowski's space-time intuition, and without committing oneself to the existence of Minkowski's four-dimensional world' (Walter 2010, 17).[86]

---

[84] As Schaffner points out, his sense of *ad-hoc*ness corresponds to Elie Zahar's concept of ad hoc_3. A theory is ad hoc_3 'if it is obtained from its predecessor through a modification of the auxiliary hypotheses which does not accord with the spirit of the heuristic of the programme' (Zahar 1973, 101). In our case, the new hypothesis which violates the spirit of the original program is, of course, the M.F.H., and the accusation holds if it can be proven that there is no possible plausibility argument for the mentioned hypothesis in the context of the electromagnetic worldview program. Since the concept of ad hoc_3 boils down to a matter of plausibility, I think that accusations of *ad-hoc*ness in this sense get better described in terms of the explanatory features of the theory involved.

[85] See (Brush 1999).

[86] As Walter points out, Emil Wiechert took a similar stance: 'Another physicist, Minkowski's former colleague and director of the Göttingen Institute for Geophysics, Emil Wiechert welcomed Minkowski's space-time theory, but felt there was no

But what are the precise aesthetic virtues that the four-dimensional presentation of the theory contains? In what sense mathematical simplicity is being considered? Peter Galison (1979) has offered a detailed account that we can use as a sample. He states that the virtues are three: symmetry, generality and invariance.

First, the symmetry that Galison refers to is geometric. In Minkowski space-time, provided that an arbitrary event is assigned with the coordinates (0,0,0,0), the set of all possible space-time coordinates for one event is given by the four-dimensional hyperboloid $ct^2 - x^2 - y^2 - z^2 = C$. The Lorentz transformations allow that any point on the hyperboloid can be transformed to lie on the t-axis—and for the corresponding vector, velocity equals zero. Galison's point is that the Lorentz transformation applied to a specific event-point takes the hyperboloid back into itself—and the physical consequence is that absolute (ether) rest becomes undefined:

> Different observers assign different coordinates to a given event. Minkowski reasons that since $t^2 - x^2 - y^2 - z^2$ is Lorentz-invariant, the four-dimensional hyperboloid $t^2 - x^2 - y^2 - z^2 = constant$ represents the set of all possible space-time coordinates of one event. The principle of relativity tells us that "absolute rest corresponds to no properties of the phenomena". Since in four dimensions there is a non-zero vector lying on the hyperboloid and corresponding to zero velocity, any point (x, y, z, t) on the hyperboloid can be transformed to lie on the t-axis. *Such a Lorentz transformation will take the hyperboloid back into itself. This is the geometric symmetry which Minkowski introduces into relativity. Its physical consequence is that no particular measurement of the coordinates of an event can indicate absolute rest.* […]
> The four-dimensional representation places rest and motion on equal graphical footing. *Since any four-vector can be transformed to the "rest-vector", leaving the hyperboloid of the appropriate invariance unchanged, the principle of relativity, i.e., that no phenomena are attached to absolute rest, stands fully exposed*. Such a symmetry is clearly distinct from the physical symmetry of Einstein [Galison here refers to Einstein's *motto* 'theoretical asymmetries that do not reflect in the phenomena'] and the formal group or group symmetries of Poincare' (Galison 1979, 104-5).

Second, the generality that Galison explains consists in that the four-dimensional geometry that the German mathematician introduced allows us to express different groups of invariant transformations—transformations which in turn determine different space-times with different metrics. The transformations in which the velocity of light takes the constant and frame-independent value of c correspond to the group of Lorentz transformations G_C and express the metric of Minkowski space-time. If the transformations do not pose an upper limit to the velocity of light, the corresponding group is G_∞, the Galilean transformations that determine a space-time with Newtonian metric.

Finally, the third aesthetic factor that Galison points out is *invariance*:

> The existence of invariants for the relativistic transformations forms the third aesthetic criterion Minkowski considers in his four-dimensional relativistic theory. "The innermost harmony of these [electrodyamic] equations", he writes, "is their invariance under the transformations of the expression $dx^2 + dy^2 + dz^2 - dt^2$ into itself". In Newtonian space-time the free *t*-axis prevents us from constructing such an invariant expression. Like symmetry and generality, invariance is an aesthetic geometric-criterion which supports the new conception of space-time. (ibid, 111-2)

need to dismiss absolute space. Following a remark made by Minkowski in "Raum und Zeit," Wiechert proposed to recover the notion of direction in Euclidean space with what he called "Schreitung" in space-time, or what amounted to the direction of a four-velocity vector. As for Minkowski's claim that a new intuition of space and time was required, this did not bother Wiechert at all. In a non-technical review of relativity theory, Wiechert wrote that the special relativity theory was "brought by Minkowski to a highly mathematically-finished form." He continued: "It was also Minkowski who, with bold courage, drew the extreme consequences of the theory for a new space-time-intuition […] and contributed so very much to the theory's renown". It was precisely Minkowski's space-time-intuition, or his identification of the extreme consequences of this intuition, that had made the theory of relativity famous in Wiechert's view. For Wiechert, however, all intuitions, including ether, and matter in motion, were but anthropomorphic "images," the reality of which was beyond our ken' (Walter 2010, p. 16).

Mathematical-aesthetic features like these could thus be invoked in order to decide the competition between special relativity and Lorentz's theory. It is true that aesthetic features are problematic from an epistemological point of view. Questions like 'why is beauty desirable in scientific theories?' or 'can a theory be objectively evaluated as more beautiful than a rival?' are open philosophical problems. However, even if we take for granted that aesthetic features can be objectively assessed and that they can ground theory-choice in a justified way, in the case of Lorentz vs. Einstein they cannot be invoked anyway.

Galison's mathematical-aesthetic features are clearly rooted in the four-dimensional mathematical structure that Minkowski introduced, and the quotes above from Laue are consistent with this view. We saw that in 1906 Poincaré derived some mathematical results of Lorentz's theory. He noticed that the expression $ct^2 - \mathbf{x}^2$ is invariant under Lorentz transformations, so that the transformations can be understood not only as rotations around the x and z axes, but also around the x-axis and a fourth axis ict. That is, Poincaré showed that Lorentz's theory can be formulated in the four-dimensional geometric language that Minkoswki developed in 1908. Therefore, all the mathematical-aesthetic virtues that can be predicated of special relativity in terms of its formulation in a four-dimensional geometry can also be assigned to Lorentz's theory. According to Einstein's theory, the Lorentz transformations express the Minkowskian metric of space-time, so that its formulation in terms of a four-dimensional geometry has to be understood in a 'literal' physical way. On the other hand, since Lorentz's theory does not depict a Minkowski space-time, but a Newtonian one, the fact that the theory can be formulated in a four-dimensional geometric language can only be understood as a *mathematical nicety* that does not express its physical content. However, this is enough in order to state that Lorentz's theory can be presented in such a way that the mathematical virtues that are assigned to special relativity can also be assigned to it. Therefore, mathematical-aesthetic features cannot be used in order to make a choice in this case[87].

Furthermore, Galison's own analysis of the import of the mathematical virtues in SR can be understood as providing reasons for the adoption of four-dimensional physics as a language and method, rather than as reasons for the adoption of *Einstein's* four-dimensional physics over *Lorentz's* four-dimensional physics. Galison's analysis is not posed within the context of the competition between these theories, but his conclusion is coherent with what I claim:

> If one grants that Minkowski can pass from good mathematics to productive physics, it remained for him to ground the new physics on mathematics alone. He accomplishes this by comparing Newtonian and relativistic theories on the basis of three criteria of geometrical elegance that emerge from his visual thinking: symmetry, generality and invariance. Together they seem to form the motivation and the justification for Minkowski's adoption of the new physics. (1979, 103)

> From his belief in the "pre-established harmony" [between mathematics and physics] and his discovery of these geometrically satisfying properties, Minkowski concludes that the four-dimensional theory is superior to Newtonian three-dimensional physics. (ibid, 109)

[87] My argument is not analogous to say that a genuinely virtuous person is no better than a thief that dresses just as the virtuous people do. My point is that, in and by itself, the clothes of the virtuous and the thief cannot make a difference with respect to their moral assessment—what makes them virtuous or vicious is not their clothes, they both have the same wardrobe. It might be argued that in the case of the virtuous, his choice of clothes is somehow connected to his morals (he might use the same mental faculty to choose his clothes and to determine his behavior), but that cannot be argued only in terms of the aesthetic features of his outfit. Something analogous occurs in the case of special relativity and Lorentz's theory. The mathematical features of the former reflect physical features. For example, Galison's symmetry, insofar as it is connected to the impossibility of defining *rest* in an absolute sense, is not a *merely* mathematical feature of special relativity, but it is linked to the *physical content* of the theory. Actually, it cannot be strictly applied to Lorentz's theory, for in it rest with respect to the ether is qualitatively distinct to relative rest (though the symmetry that Galison points out could be taken as reflecting the fact that the state of absolute (ether) rest is empirically undetectable in Lorentz's theory). Therefore, Galison's concept of symmetry manifests the *physical simplicity* of special relativity. However, this does not mean that the theories at stake do not stand on an equal foot regarding (strictly) mathematical virtues connected to their four-dimensional formulations. Physical simplicity is not *only* a matter of aesthetics. As I will argue below, in the case of special relativity vs. Lorentz's theory, this feature exhibits the firmer physical foundations of the former.

In these two passages, *the new physics* and the *four-dimensional* physics can be understood as referring to both Einstein and Lorentz's theories, whereas the Newtonian rival is simply classical mechanics. Actually, as Galison clearly shows (1979, 90-5), Minkowski was a supporter of the electromagnetic worldview, and even when he wrote his famous papers on four-dimensional space-time he was thinking in terms of the Lorentz-Einstein theory—but he did not have the time to see the clarification that his work involved, for he died shortly after he published his seminal papers.

2.5.3 Janssen's argument

Michel Janssen has offered a very influential argument regarding the case of Lorentz vs. Einstein (1995, 2002a, 2002b, 2009, and Balashov and Janssen 2003). This argument relies on the different way in which the theories involved draw the line between dynamics and kinematics. Recall that in Lorentz's theory effects like clock-retardation, length-contraction and the velocity-dependence of mass are grounded on a peculiarity of the laws that govern the microstructure of all physical systems, namely, their Lorentz-invariance. Thus, the Lorentz-invariance of all physical laws has a *dynamical* foundation within Lorentz's theory. The ultimate dynamical ground of such invariance could be traced to the interaction between electromagnetic systems and the ether. Therefore, for the supporters of the electromagnetic worldview, Lorentz-invariance was a natural feature of physical laws. In the case of special relativity, the Lorentz-invariance of the laws of physics is not dynamically grounded. The mentioned physical effects are a manifestation of the new and revolutionary kinematics that the theory postulates. This feature of the theory became clear and explicit in the seminal work of Minkwoski: the metrical features of Minkwoski space-time—that rule the normal spatio-temporal behavior of physical objects—determine that the laws governing physical systems are Lorentz-invariant. That is, in special relativity the Lorentz-invariance of physical laws has a *kinematical, chrono-geometrical* foundation.

Janssen argues that special relativity can be preferred over Lorentz's theory because in the former the kinematical explanation of Lorentz-invariance is based on a *common origin inference* (*COI*). In Lorentz's theory, there is no common origin to explain the fact that *all* the laws of physics are Lorentz-invariant. Lorentz developed his theory from the fact that *electromagnetic* laws are Lorentz-invariant—recall his plausibility argument for the length-contraction hypothesis. However, the mature version of his theory required that *all* the laws of physics have the property at issue. If a unified explanation of Lorentz-invariance is not available, then the fact that both electromagnetic and non-electromagnetic laws of nature possess that property becomes a sort of cosmic coincidence. There was a time when the electromagnetic worldview, that intended to reduce all physics to electrodynamics, offered a promising possibility: if Lorentz-invariance is characteristic of electromagnetic laws, and if all laws are electromagnetic, then the Lorentz-invariance of all laws is guaranteed and explained by a common origin. However, the necessary introduction of Poincaré-pressure implied that a full reduction of physics to electrodynamics was not achieved[88].

This argument, of course, requires further epistemological justification. After all, a similar unexplained coincidence in Newton's theory—the equivalence between gravitational and inertial mass—did not challenge its acceptance. It is true that Newton's theory was replaced by general relativity, and a salient feature of the latter is that it contains a common origin explanation for the mentioned equivalence. However, general relativity defeated Newton's theory on the empirical-evidence battleground. In a case of *empirical equivalence*, such as special relativity vs. Lorentz's ether theory, it must be shown that a choice based on

[88] At this point we can see a similarity between Janssen's and Schaffner's arguments. Schaffner states that the M.F.H. became definitively ad hoc_3 with the rise of quantum physics. That is, no possible explanation for the physical underpinning of the Lorentz-invariance of non-electromagnetic dynamical laws could be offered—whereas special relativity was not ad hoc in this sense. Janssen claims that the Lorentz-invariance of all physical remains an unexplained coincidence, whereas in special relativity this feature gets naturally explained by the kinematics of the space-time postulated by the theory.

a specific *COI* is epistemologically justified. That is, it must be explained *why* to have a common origin explanation is better than not having it.

Concerning such a justification, Janssen has developed and subtly modified his argument over time. The first version of the argument (1995, 2002a, 2002b) is of a causal nature, Minkwoski space-time is taken as the *cause* of Lorentz-invariance:

> The contraction of physical systems and the retardation of processes in such systems when the system is set in motion, no matter whether the system is of an electromagnetic or of a non-electromagnetic nature, are effects that seem to have a common cause in special relativity, but that are due to unexplained coincidences in the ether theory. (Janssen 1995, section 4.2.1)

> Lorentz invariance manifests itself in many different phenomena. Ultimately, these phenomena form the input of the common-cause argument. The most obvious examples are length-contraction and time dilation. The length of a system in uniform motion is less than that of an identical system at rest by a factor that depends only on the velocity of the moving system. A process in the moving system takes longer than the corresponding process in the system at rest by that same factor. In the Newtonian space-time of Lorentz's theory this is a consequence of the unexplained coincidence that all of these systems are governed by Lorentz-invariant laws. In the Minkowski space-time of special relativity, it is a consequence of the way in which particular space-time slices are used to define the length of a system or the duration of a process. (Janssen 2002a, 439).

To speak of space-time as a common *cause* is of course a risky maneuver, for it seems to imply a substantivalist position concerning space-time—and the question of the ultimate ontology of space-time is an open, *philosophical* problem. This is why Janssen (1995) softened his position by stating that his argument is of a common cause *type*:

> The reason for calling this a 'common cause'-type argument rather than a common cause argument, is that Minkowski space-time does not seem to be a common cause in quite the same sense that a shrimp cocktail contaminated with the salmonella bacteria is the common cause of the sudden death of half the population of a cheap Dutch old folks home. […]
> Although the status of the 'common cause' obviously needs further philosophical clarification, it is safe to say, I think, that this is a very strong argument for preferring special relativity over an empirically equivalent classical ether theory. (Janssen 1995, section 4.2.1).

Unlike Janssen (1995), I think that it is very unsafe to argue that space-time can be a cause of anything. This position, in spite of the softening clause, seems to be essentially committed to a substantivalist view, and the relationist would surely reply that there is no sense in which space-time can be a cause. That is, the proviso that the argument is of a common cause *type* is not enough to avoid the problematic commitment regarding the ontology of space-time.

In his (2002b), Janssen still argues for a causal argument. Minkowski space-time is again attributed a causal role in the sense that it is responsible for the Lorentz-invariance of physical laws. However, this time Janssen is explicit in that we should not take this as an indication that Minkowski space-time is a substance. To understand Minkowski space-time as a *structure* is enough in order to endow it with causal powers:

> What makes the special-relativistic explanation of Lorentz invariance of physical laws a good explanation? Is it because it provides a unified account or because it provides a causal account? In other words, does the example support Salmon's causal account of explanation or Friedman and Kitcher's unification account?
> If unification were all that matters, it is unclear why Lorentz's theory does not provide a perfectly adequate explanation of length-contraction. Lorentz's claim that all laws are Lorentz invariant has tremendous unifying power. On the basis of this one claim it can account for everything special relativity can account for. Yet, the theory is unsatisfactory from an explanatory point of view. It does not have the resources to explain why physical systems in the Galilean space-time posited by the theory are all governed by Lorentz-invariant laws. In short, Lorentz's theory unifies but does not explain. […]

> At first glance, Salmon's causal account is unsatisfactory as well. Minkowski space-time certainly explains length-contraction, but it hardly qualifies as a causally efficacious substance. This objection can be avoided by broadening Salmon's concept of causation […]. The COI in this case is to a structure rather than a substance. In this way Salmon gets this example right: special relativity explains and Lorentz's theory does not. (Janssen 2002b, 501).

I think that this maneuver does not do the trick either. It is not clear at all what does it mean that Minkowski space-time is a structure with causal powers. Is it a *physical* structure such as a crystal? This position would be very hard to defend. But if it is an *abstract* structure then it is totally unclear in what way such a structure can be the cause of anything. Harvey Brown makes a similar objection. When considering the argument that Minkowski space-time explains the Lorentz-invariance of physical laws he asks 'So how is its influence on these laws supposed to work? How in turn are rods and clocks supposed to know which space-time they are immersed in?'(Brown 2005, 143).

In a subsequent paper, Balashov and Janssen (2003) tried to disentangle the argument from any ontological presuppositions about space-time, while conserving the common origin kinematic explanation for Lorentz-invariance that special relativity offers. They make their point by means of an analogy. They ask us to imagine Cyrano running off Roxanne's house. As Roxanne sees Cyrano, he turns and goes away. As he turns, Roxanne sees his nose to become more and more prominent, then to get smaller and smaller until it vanishes. Balashov and Janssen claim that it is the structure of Euclidean space what explains why the laws that hold Cyrano's nose together are invariant under rotations. They also claim that, analogously, it is the structure of Minkowski space-time what explains that physical laws are Lorentz-invariant. This way to present the matter still retains a substantivalist flavor. However, Balashov and Janssen then state that, even under a relationist ontology, the arrow of the explanation points in the same direction:

> Since the invariant group (Lorentzian or Galilean) of the dynamical laws essentially *is* the space-time structure for a relationist, the (effective) Lorentz invariance of the dynamical laws in a sense does seem to explain for a relationist why the (effective) space-time structure is Minkowskian. Such an explanation, of course, does not amount to an explanation of why the space-time structure is Minkowskian *rather than Newtonian* […]. Nor does it interchange the role of *explanans* and *explanandum* in our Cyrano-and-Roxanne example. The behavior of Cyrano's nose is just an instance of the normal spatio-temporal behavior of objects in Minkowski space-time, no matter whether one is a substantivalist or a relationist about the ontology of space(-time). The explanatory considerations in the text are therefore largely independent of one's stance on the ontology of space(-time). (Balashov and Janssen 2003, 341, footnote 11).

I think that in this passage Balashov and Janssen do approach a formulation of the explanatory nature of special relativity independent of ontological assumptions about space-time. The key sentence is that 'the behavior of Cyrano's nose is just an instance of the normal spatio-temporal behavior of objects in Minkowski space-time', meaning that the Lorentz-invariance of physical laws encodes the universal spatio-temporal behavior of physical objects. The clause 'objects in Minkowski space-time' still retains a substantivalist flavor, but we can understand it—and I think this is what Balashov and Janssen have in mind—as stating that the natural spatio-temporal behavior of physical objects is given by the kinematics encoded in the Lorentz-invariance of physical laws.

Janssen, though, interprets this manner of displaying the explanatory form of special relativity in a way I do not agree with. He claims that the fact that Minkowski space-time explains the Lorentz-invariance of physical laws entails that a *constructive* explanation of the invariance is given by special relativity[89]:

[89] Balashov and Janssen formulate Einstein's distinction in the following way: 'In a theory of principle, one starts from some general, well-confirmed empirical regularities that are raised to the status of postulates (e.g., the impossibility of perpetual motion of the first and second kind, which became the first and second laws of thermodynamics). With such a theory, one explains the phenomena by showing that they necessarily occur in a world in accordance with the postulates.

> For Einstein special relativity was a theory of principle. With the introduction of Minkowski space-time, however, it became a constructive theory. Minkowski space-time is the structure responsible for all the effects derivable from special relativity alone. Special relativity, from this point of view, replaced Newtonian space and time by Minkowski space-time and does not make any claims about the contents of the new space-time other than their spatio-temporal behavior had better accord with Minkowski's new rules. (Janssen 2002b, 506).
>
> Both the space-time and the Neo-Lorentzian interpretation [of special relativity] provide constructive-theory explanations. In the space-time interpretation, the model is Minkowski space-time and length-contraction is explained by showing that two observers who are in relative motion to one another and therefore different sets of space-time axes disagree about which cross-sections of the 'world-tube' of a physical system give the length of the system. (Balashov and Janssen 2003, 331).

This view becomes problematic insofar as the constructive nature of the explanation, according to Janssen, still holds in the case of its ontological commitments-free formulation. In his (2009), Janssen introduces important amendments and remarks that aim to detach his stance from any possible ontological presuppositions about space-time. For example, he states that his argument is no longer of a causal type:

> I claim that Minkowski space-time explains Lorentz-invariance. For this to be a causal explanation, Minkowski space-time would have to be a substance with causal efficacy. Like Brown, I reject this view […]. As I hope to make clear, Minkowski space-time explains by identifying the *kinematical nature* (rather than the cause) of the relevant phenomena.
> Special relativity, *as a physical theory*, is agnostic about the ontology of space-time. (Janssen 2009, 28).[90]

After setting the epistemological framework of his argumentation, that is, after being explicit in that his position is completely independent of specific ontological stances regarding space-time, Janssen still argues that the kinematic explanation of Lorentz-invariance qualifies as a constructive one:

> The Lorentz invariance that can be derived from the postulates (in conjunction with the assumption that space and time are homogeneous and isotropic) finds its natural interpretation in terms of the geometry of Minkowski space-time. On my definition of the principle-constructive distinction[91], this interpretation amounts to the constructive-theory version of special relativity. It says that the space-time component of any acceptable model of a world in accordance with the postulates is Minkowski space-time. (Ibid, 39)
>
> Minkowski (1909) did for special relativity, understood strictly as a principle theory, what Boltzmann had done for the second law of thermodynamics. It turned special relativity into a constructive theory by providing the concrete model for the reality behind the phenomena covered by the principle theory. (Ibid, 40).

I think that the problem with Janssen's position is quite clear. If, as Einstein's and Janssen's definition states, a constructive theory aims to get at the underlying *reality* behind the phenomena, how is it possible that the Minkowskian formulation of special relativity can count as a constructive theory *without in some sense reifying or hypostatizing Minkowski space-time*? I think this is just impossible. The very definition of a constructive explanation precludes it. Even if it is argued that Minkowski space-time—understood not as a thing, but as a collection of *relations*—constructively explains the Lorentz-invariance of dynamical laws, the spatio-temporal relations must be characterized as ontologically independent from the *relata*—

Whereas theories of principle are about the *phenomena*, constructive theories aim to get at the underlying *reality*. In a constructive theory one proposes a (set of) model(s) for some part of physical reality (e.g., the kinetic theory modeling a gas as a swarm of tiny billiard balls bouncing around in a box). One explains the phenomena by showing that the theory provides a model that gives an empirically adequate description of the salient features of reality.' (Balashov and Janssen 2003, p. 331).

[90] Notice that these remarks strengthen the view that when Janssen uses the expression Minkowski space-time we have to understand something like 'the kinematics encoded in the Lorentz-invariance of physical laws'.

[91] For his definition of the principle-constructive distinction, Janssen remits to the definition presented in (Balashov and Janssen 2003) that I quoted in footnote 24.

otherwise it would be the Lorentz-invariance of the laws governing the *relata* (physical objects) what constructively explains the structure of the collection of spatio-temporal relations. That is, space-time would be hypostatized anyway, for it must still be presupposed that space-time exists independently of, or prior to, the existence of physical objects[92]. I think that it is not possible to argue that in special relativity the kinematics *constructively* explain the dynamics without assuming a strong ontological commitment regarding the existence of space-time—a commitment that is not sanctioned by Einstein's theory. As we saw above, Janssen himself states that 'special relativity, *as a physical theory*, is agnostic about the ontology of space-time' (2009, 28).

After this review and critical assessment of Janssen's arguments, we can now consider the best possible formulation of what he regards as a reason to prefer special relativity. Recall that the basic idea is that in Lorentz's theory the fact that all the laws of physics are Lorentz-invariant remains as an unexplained coincidence; whereas in special relativity Lorentz-invariance gets naturally explained by the specific kinematics of space-time. For the reasons mentioned above, this explanation cannot be taken either as causal or as constructive. If it is an explanation at all, it has to be an explanation of principle.

A more formal presentation of this argument is offered in (Balashov and Janssen 2003, 341). This formulation relies on John Earman's *symmetry principles*: *SP1)* any dynamical symmetry of a theory *T* is a space-time symmetry of *T*, and *SP2)* any space-time symmetry of *T* is a dynamical symmetry of *T* (Earman 1989, 46). In Lorentz's theory, the Galilean transformations express the spatio-temporal symmetries, but they do not express the symmetries of the dynamical laws. In turn, the Lorentz transformations express the symmetries of the dynamical laws, but they do not correspond to the spatio-temporal symmetries of the theory. This defect is not present in special relativity, of course. In this case the Lorentz transformations represent both the space-time and the dynamical symmetries. In other words, Lorentz's theory violates Earman's principles, whereas special relativity respects them.

Janssen takes it that in Einstein's theory it is the space-time symmetries of the theory that *explains* the symmetries in the dynamical laws—I have argued that this explanation cannot be either causal or constructive (unless we endorse some form of substantivalism). We are thus left with a principle-kinematic explanation. However, Janssen's stance—that the observance of Earman's principles in special relativity means that the (kinematic) space-time symmetries explain the dynamical symmetries—has been challenged by Harvey Brown and Oliver Pooley. They argue that the arrow of explanation points in the opposite direction: it is the dynamic symmetries that explains the space-time symmetries. Their argument to refute Janssen's view is that

> as a matter of logic alone, if one postulates space-time structure as a self-standing, autonomous element in one's theory, it need have no constraining role on the form of the laws governing the rest of content of the

[92] Janssen's motivation for insisting in characterizing special relativity as a theory of principle in which Minkowski space-time plays the role of 'the reality underlying the phenomena' is that, concerning the comparative explanatory power of theories of principle and constructive theories, he endorses the following view: 'Brown and Pooley (2006, pp. 74-5) correctly point out that, contrary to what Balashov and I suggested, principle theories are not explanatory. Explanations are about the reality behind the phenomena (be it about their causes or about their nature). Principle theories, on the definition used by Balashov and me, are agnostic about that' (Janssen 2009, p. 38, footnote 27). However, I think that the situation is not that bad concerning principle theories and their explanatory power. To say that principle theories *are not* explanatory seems exaggerated. It makes more sense to argue that, in principle, and given a certain realm of phenomena, a constructive explanation is superior in terms of the understanding it provides of those phenomena when *compared* to an explanation of principle. The paradigmatic example of thermodynamics and statistical mechanics is illustrative. Complete understanding of the corresponding phenomena was given by means of the constructive theory, but that does not mean that the principle theory does not explain. Actually, in some contexts an explanation of principle might be superior to a constructive theory. Balashov and Janssen defend the explanatory power of principle theories referring to thermodynamics (2003, pp. 32-3), but Janssen abandoned this view in his (2009), as we just saw. Matthias Frisch (2005, 2011) has offered an interesting comparison between Einstein's principle/constructive distinction and a very similar one introduced by Lorentz in 1900. He also argues that both Einstein and Lorentz (especially the latter) did assign an important explanatory value to the 'principle approach'.

> theory's models[93]. So how is its influence on these laws supposed to work? Unless this question is answered, space-time's Minkowskian structure cannot be taken to explain the Lorentz covariance of dynamical laws. (Brown and Pooley 2006, 84).

Brown and Pooley's point is that, as we saw above, the only way in which it can be argued that Minkowski space-time can constructively explain Lorentz-invariance of the dynamical laws is by reifying it[94]. But even if this step is taken, it is still unclear in what way the very peculiar entity that space-time is can determine the behavior of physical objects[95]. The remaining alternative is not to postulate the space-time structure as autonomous and self-standing, but as determined by the peculiarities of the dynamical laws. But this view, Brown and Pooley argue, boils down to state that the arrow of explanation goes from dynamics to kinematics[96].

My intention here is not to take sides in the debate regarding which is the correct interpretation of special relativity, kinematical or dynamical, neither in the debate concerning the ontological status of space-time. My point is simply that if we follow Janssen's path in interpreting special relativity's compliance with Earman's symmetry principles in terms of a (kinematic) explanation of Lorentz-invariance, then we get involved in an epistemological, *open* dispute: the ontology of space-time and the direction of the arrow of explanation between kinematics and dynamics in special relativity. The upshot would thus be that the reason we are invoking to decide the case of Einstein vs. Lorentz relies on epistemological grounds that can be (and are) attacked. The acceptance of Janssen's argument to select special relativity over Lorentz's theory presupposes that we are ready to follow him in that, in Einstein's theory, kinematics constructively explains the dynamics. In simple words, Janssen's argument to prefer special relativity works for those who think that he is right and Brown is wrong concerning the debate about the correct interpretation of special relativity. This is an instance of what I say below about non-empirical features as reasons to decide cases of empirical equivalence: they cannot ground a fully objective and uniquely determined choice.

We may then look for epistemic justifications of the symmetry principles that do not involve us in open epistemological problems regarding the direction of the arrow of explanation in special relativity and the ontology of Minkowski space-time. An alternative way to defend the symmetry principles is given by considerations of physical and ontological simplicity. The violation of the principles in Lorentz's theory reflects in that the ether—and the corresponding privileged frame of reference the theory postulates—is empirically superfluous and implies asymmetries that do not correspond to the phenomena, as Einstein would put it. The ether and the privileged frame of reference are empirically superfluous in the sense that they are not relevant in the derivation of empirical consequences—I will return to this issue below. On the other hand, the different explanation of the electrodynamic interaction between a magnet and a

---

[93] This is actually the case in Lorentz's theory. The Newtonian space-time structure postulated does not constrain the dynamical laws to express the same symmetries.

[94] Both Brown and Pooley, on one side, and Janssen and Balashov, on the other, are looking for *constructive* explanations in the kinematics vs. dynamical interpretation of special relativity. That is, even if we assume that theories of principle do explain, their debate presupposes that a full and physical explanation is given in constructive terms. If explanations of principle are possible, they only explain from a *formal* point of view, we could say, and in this sense an explanation of principle is *ontologically* subordinated to a constructive one. For an alternative in which the explanatory nature of special relativity is not absolute, but context-dependent, see (Dieks 2009).

[95] Brown (2005, 24) argues that the substantivalist must assume that physical objects have 'space-time feelers' that allow them to react to space-time's 'ruts and grooves'.

[96] In a nutshell, Brown's dynamical interpretation of special relativity (see Brown 2005) consists in that if we take the Lorentz-invariance of dynamical laws as a given that need no further explanation—after all, the laws of science explain the phenomena, but we usually do not demand for explanations of why the laws hold, or for why they have the form they have—, then it *follows* that the kinematic, space-time structure will be Lorentz-invariant as well. This view, of course, is rather sympathetic to a relationist ontology concerning space-time. For an interesting assessment of the debate between Brown-Pooley and Janssen-Balashov, see (Frisch 2011). There the author argues that even though there is a real disagreement concerning the direction of the arrow of explanation between kinematics and dynamics in special relativity, there are more points of agreement than what it seems at first sight.

conductor, depending on whether we consider the former or the latter as at rest in the privileged frame, is a paradigmatic example of how the ether introduces dubious asymmetries that do not belong to the phenomena, as Einstein mentions in his 1905 paper.

The violation of the symmetry principles in Lorentz's theory results in that the ether—in connection with the privileged ether-rest frame and the Newtonian structure of space-time—becomes suspicious of representing nothing physical. The fact that special relativity does not postulate entities or structures that are dubious in this sense makes it a physically and ontologically simpler theory than Lorentz's. That is, the comparative simplicity of Einstein's theory is not a merely pragmatic virtue, but it reflects more solid ontological foundations[97].

To describe special relativity's physical simplicity in terms of Earman's symmetry principles has the advantage of allowing us to make sense of the historical role that this feature played in the acceptance of Einstein's theory. The asymmetries that the ether and a privileged frame of reference entail were one of the motives determining Einstein's formulation of special relativity, and the physical simplicity of the theory was surely one of the reasons why it was accepted (over Lorentz's). In the previous section I argued that *mathematical* simplicity does not work as a criterion to make a choice, but it is very reasonable to think that the *physical* simplicity associated to the formulation of special relativity in terms of Minkowski space-time did not escape the eye of the scientists of the time. So when Laue, for example, refers to the simplicity acquired by the four-dimensional geometric formulation, it is rather likely that he also had in mind the *physical* simplicity involved[98].

These remarks thus show that the best way to reassess Janssen's argument concerning Lorentz-invariance is the following: the fact that special relativity respects the symmetry principles—whereas Lorentz's theory violates them—is a manifestation of a higher degree of physical simplicity. From the point of view of physical and ontological simplicity, Einstein's special relativity is clearly superior to Lorentz's ether theory, and this feature thus offers a good reason in order to prefer the former over the latter.

2.5.4 The superfluous ether

The problematic status of the ether is another feature that has been typically considered as a reason to dismiss Lorentz's theory and to embrace special relativity. Usually, the ether is held to be problematic chiefly because it is undetectable. Since there is no possible observation that directly indicates the reality of the ether, we should simply apply Occam's razor and pick the most economic theory. Adolf Grünbaum, for example, has argued along this line:

> Since the observational consequences of the *aether-theoretic interpretation* of the Lorentz transformations are the same as those of their rival relativistic interpretation, the aether-theoretic interpretation can have no observational consequences which are different from those of the rival special theory of relativity. Hence there can be no observational consequences which would support the doubly amended theory[99] as against

[97] Note that the conceptual justification to pick the observance of Earman's principle as a feature that shows special relativity's superiority over Lorentz's theory is not grounded on simplicity in and by itself, or as a merely aesthetic virtue. As I mentioned above, it is philosophically problematic to assert that simplicity is always desirable property in scientific theories. In this case, the simplicity manifested in the observance of Earman's symmetry principles *reflects firmer ontological foundations*.

[98] Actually, Schaffner (1974, p. 74), in the passage by Laue he quotes, interprets the simplicity invoked in a *physical* way. Janssen (2002b, pp. 502-7) quotes Poincaré, Lorentz, Einstein and Minkowski in order to show that they explicitly discussed the foundations of Lorentz-invariance in connection to the Einstein vs. Lorentz case, and argues that such discussions can be described in terms of a *COI*. I think that we can make sense of all those references by means of *physical simplicity*, with the advantage of not getting involved in the epistemological troubles discussed above.

[99] What Grünbaum calls the doubly amended theory is Lorentz's core theory plus the length-contraction and the clock-retardation hypotheses. As Janssen points out, the doubly amended theory is a sort of toy model that does not really grasp the historical subtleties in the development of Lorentz's theory. However, in this passage we can harmlessly replace 'the doubly amended theory' by 'Lorentz's definitive theory'.

> the new rival special theory of relativity, a theory that refuses to postulate the existence of some one preferred inertial aether frame *when there is no kind of physical foundation for doing so*. (Grünbaum 1973, 724; last emphasis is mine).

Two objections can be leveled against this view. First, it is awkward to state that there was no physical foundation in order to postulate a preferred inertial ether frame. Maxwell's equations were originally understood as being valid in the ether-rest frame, and of course, this is the way in which Lorentz understood them. In his theory the ether-rest frame became physically privileged in order to account for the negative result of ether-drift experiments within the framework of a Newtonian space(-time) and classic electrodynamics. That is, the view that there was no physical foundation for the postulation of a preferred ether-rest frame is incompatible with historical facts. In the long run, it turned out that such postulation was unnecessary (or plainly wrong), but that is a different story.

Second, and more importantly, the automatic rejection of unobservable entities as a normative principle is debatable. Even if an entity postulated by a certain theory is *directly* undetectable, there still might be other observable effects that are connected to that entity, and in this sense they could be interpreted as observable *traces* of its reality. As Janssen puts it,

> as part of his analysis of the doubly-amended theory Grünbaum offers a more accurate diagnosis of the trouble with Lorentz's theory. What makes the theory unsatisfactory are not the elements that are added, but some of the original elements, notably the ether and Newtonian space-time, that are rendered more and more invisible with every amendment. This suggests that Ockham's razor is all that is needed to settle the case of Einstein *versus* Lorentz. The problem with this type of argument is that it derives its force from a blanket rejection of unobservables in scientific theories, whereas it is widely accepted that such elements should not be banned automatically. Rather than condemning unobservables in general, I think it is wiser to demand arguments to put forward on a case-by case basis to show why a particular unobservable is otiose' (Janssen 2002, 438)

This is exactly what happens in Lorentz's theory. It is true that the ether is undetectable in this theory. However, the generalized contraction hypothesis is grounded in the interaction between physical systems and the ether. Therefore, all observable consequences in which this hypothesis is logically relevant for their derivation—such as the negative results of ether-drift experiments and the velocity-dependence of mass—can be understood as traces of the reality of the ether.

These remarks are enough in order to show that the standard arguments regarding the problematic status of the ether in Lorentz's theory are not compelling. However, there is a different perspective from which the ether can be questioned anyway—to my knowledge, this perspective has not been considered in the relevant literature. The real problem with the ether relies on the fact that, after Poincaré showed that the Lorentz coordinate transformations are symmetric, it became not only directly unobservable, but also empirically *superfluous*—a hypothesis being superfluous if it is logically irrelevant for the derivation of the empirical consequences of the theory it belongs to[100].

Recall that Poincaré demonstrated that if Lorentz's theory is to respect the principle of relativity, the Lorentz coordinate transformations must form a group. In turn, if the transformations form a group, in

[100] I take this concept of superfluity from (Norton 2008). Let us recall that this author argues that, given a pair of predictively equivalent theories, if one of them contains superfluous structure, this strongly suggests that the theories are, after all, one and the same—with one of the versions carrying some unnecessary extra-baggage, so that by excising the superfluous structure from the less economic theory we would obtain the empirically equivalent 'rival'. I think this not so in the case of Einstein vs. Lorentz. If we apply Occam's razor and excise the ether from Lorentz's theory, what we obtain is not special relativity. It is still possible to retain the Newtonian space-time plus conspiring dynamical effects by defining *by fiat* a privileged reference frame. For example, the Lorentzian could baldly say that the privileged frame is the one in which the real time is measured, period—even if it is impossible to physically determine which frame is that. That would be a theory with questionable foundations of course, but the point is simply that the main difference between special relativity and Lorentz's theory is based on the incompatible space-times they postulate. This difference remains even if we get rid of the ether.

the *inverse* coordinate transformations the parameter $-v$ is replaced by $+v$, and the primed and unprimed quantities shift roles. If this is so, the velocity parameter in the coordinate transformations refers to the *relative* velocities between the frames involved, not to the velocity with respect to the ether—this is particularly clear if we apply the transformations in a case in which both frames are in motion with respect to the ether. Recall that in the definitive formulation of the theory its empirical consequences are derived from the coordinate transformations interlocked with the generalized contraction hypothesis. Since the velocity parameter in the *symmetric* Lorentz transformations is simply the relative velocity between the frames, reference to the ether is not logically required in order to derive the empirical consequences of the theory—its function becomes only theoretical. The only role that the ether plays is to provide the physical plausibility argument for the generalized contraction hypothesis. For example, in Lorentz's theory rods in motion contract because of their interaction with the ether, but reference to the ether is not necessary to predict that rods in motion get contracted (recall the example of rods A and B in section 2). In this scenario, to interpret the negative result of ether-drift experiments and the velocity-dependence of mass as traces of the reality of the ether becomes a dubious stance.

This unease in Lorentz's theory can thus be used to argue that special relativity is a better theory. By becoming empirically superfluous, the ether acquires a rather metaphysical flavor. As John Norton argues (2008, 35), superfluous structures or entities are highly suspicious of representing nothing physical. Within Einstein's theory, on the other hand, there is no such problem. The problematic status of the ether can thus also be described in terms of physical simplicity: the ontology of special relativity is more economic than the ontology of Lorentz's theory. Just as the violation of the symmetry principles, the superfluous character of the ether is a flaw in Lorentz's theory not (only) because of aesthetic or pragmatic reasons. The comparative ontological simplicity of special relativity is connected to firmer ontological foundations.

So far, our survey of the case of Lorentz vs. Einstein, from the point of view of non-empirical features, has delivered the following results. The accusation of *ad-hoc*ness, in spite of the main role it used to play in the philosophical discussions about Lorentz's theory, has proven ungrounded. In the case of simplicity, the result is more subtle. If we consider simplicity from a *mathematical* point of view, we have that the four-dimensional language in which Minkowski presented Einstein's theory can also be used, *mutatis mutandis*, to formulate Lorentz's theory. Therefore, all the mathematical elegance and simplicity that the four-dimensional formulation of special relativity made explicit can also be assigned to Lorentz's ether theory. *Physical* simplicity, on the other hand, is a non-empirical feature that can certainly work as a criterion according to which special relativity proves superior. Its compliance with Earman's symmetry principles and the absence of the superfluous ether make special relativity a simpler theory from an ontological point of view; and this comparative simplicity reflects that Einstein's is a theory with more solid foundations than its rival. Now I will consider two empirically and evidentially grounded reasons that can be used in order to make a choice between SR and the LPT. The first one is rooted in the inter-theoretical relation between the LPT and quantum physics. The second relies on the logical and conceptual connection between SR and the general theory of relativity.

2.5.5 Lorentz's theory and quantum physics

I showed above that by the last years of the 19th century the electromagnetic worldview arose as a revolutionary program to replace mechanics as the basic and universal framework under which physical science was to be understood. Lorentz and others, up to a certain extent, succeeded in reducing Newtonian mechanics to electromagnetic laws. In 1900 Lorentz also made an attempt to include gravitation into the scope of the electromagnetic view[101], attempt which was further developed by Wilhelm Wien. Even

[101] See (McCormmach 1970b, 476-7).

though it was not a successful endeavor some of its results were received as promising. It is true that the introduction of non-mechanical forces was a drawback for the program, but the real problems started with Planck's work.

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 work, the black-body radiation spectrum had been an intractable problem for classic electrodynamics and thermodynamics. Planck's quantum hypothesis was soon acknowledged to be deeply at odds with the foundations of 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 the model of the electron it included—was 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.

Actually, Hendrik Lorentz himself played a central role in the recognition of the essential conflict between quantum physics and the core of classic electrodynamics (see Kox 2013 and McCormmach 1970, 486-7). Lorentz, between 1900 and 1903, devoted his work to find the expression for density radiation of black-body as a function of its temperature and the wavelength of the radiation. The formula he obtained on the basis of electrodynamics—using his model of the electron—applied to thermodynamics was equivalent to the Rayleigh-Jeans law, and, of course, it worked only in the long-wavelength part of the spectrum of emission[102]. Although he first took this result as very promising, he soon realized that his model of the electron—a cornerstone of his ether theory—and electrodynamics in general were at deep odds with the results of black-body experiments and the quantum hypothesis required in order to explain such results:

> 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.
>
> 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. (McCormmach 1970, 487).

[102] '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.' (McCormmach 1970, pp. 486-7).

This conflict between electrodynamic theory and quantum physics, that Lorentz himself contributed to make explicit, became more and more important in the scientific discussion during the following years. It was eventually realized that Lorentz's electrodynamic model of the electron was essentially incompatible with the new physics of the quantum[103]. This recognition, in turn, worked as one of the main reasons why Lorentz's theory was abandoned, leaving the way open for the prevalence of SR:

> 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. (McCormmach 1970, 488).

On the other hand, SR, in spite of its predictive equivalence with Lorentz's theory, was not in conflict with the quantum hypothesis. Einstein's theory, as we saw above, is a theory of principle, and as such, it is silent with respect to issues concerning the ultimate structure of matter, fields or energy. The conflict of Lorentz's theory with quantum mechanics was grounded on the model of the electron included in the former. In special relativity there simply was no model of the electron at all. This feature became totally explicit in the eyes of the scientific community of the times through the work of Minkowski. This essential difference between the theories was not initially noticed by the scientific community. Actually, the expression *Lorentz-Einstein theory* was common before Minkowski's work. He showed that special relativity is 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, by 1909 the expression *Lorentz-Einstein theory* began to disappear from physicists' vocabulary. Minkowski's work had shown that they were two theories of a different nature (although Minkowski himself did not clearly notice such a difference), and by 1911 the rise of quantum mechanics turned the balance decidedly in Einstein's favor. As Helge Kragh explains in his book on the history of 20th physics:

> 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

[103] Lorentz's ether theory is not a theory about thermal phenomena, of course. However, the main tenets of the electron model it includes can be applied to thermodynamics—as Lorentz did between 1900 and 1903. Thus, the observed spectrum of emission of black-body radiation does not *directly* falsify the ether theory, but Planck's law cannot be derived using Lorentz's model of the electron. The theoretical assumptions required for the derivation of Planck's correct law of black-body radiation are thus incompatible with the ether theory.

> 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).

Furthermore, notice that this view allows us to make sense of the case of predictive equivalence at issue both from the conceptual and the historical point of view. We have just seen that the decision made by the scientific community of the time was crucially determined by the conflict between Lorentz's theory and quantum physics. The unease between the latter and electrodynamics was historically relevant also in the *formulation* of SR. Einstein was fully aware of the deep crisis that the quantum hypothesis was about to provoke. With particular foresight, just after the publication of Planck's work he noticed that the quantum hypothesis was to lead to a deep reformulation of the basic framework of physics. Besides, a few months before his paper on SR appeared, Einstein published his work on the light-quantum hypothesis—hypothesis which relied on Planck's concept—so the fact that electrodynamics was going to dramatically crash against the new physics of the quantum was particularly clear in front of Einstein's eyes. Einstein narrated this episode with particular clarity in his *Autobiographical Notes*:

> Planck got his radiation-formula if he chose his energy elements ε of the magnitude $\varepsilon = h\nu$. The decisive element in doing this lies in the fact that the result depends on taking for ε a definite finite value, i.e., that one does not go to the limit $\varepsilon = 0$. This form of reasoning does not make obvious the fact that it contradicts the mechanical and electrodynamical basis, upon which the derivation depends. Actually, however, the derivation presupposes implicitly that energy can be absorbed and emitted by the individual resonator only in "quanta" of magnitude $h\nu$, i.e., that the energy of a mechanical structure capable of oscillations as well as the energy of radiation can be transferred in such quanta—in contradictions to the laws of mechanics and electrodynamics. […] All of this was quite clear to me shortly after the appearance of Planck's fundamental work; so that, without having a substitute for classical mechanics, I could nevertheless see to what kind of consequences this law of temperature-radiation leads, as well as for other related phenomena of the transformation of radiation-energy, as well as for the specific heat of (especially) solid bodies. All my attempts, however, to adapt the theoretical foundation of physics to this [new type of] knowledge failed completely. It was as if the ground had been pulled out from under one, with no firm foundation to be seen anywhere, upon which one could have built. (Einstein 1959, 49)

This awareness, in turn, was a crucial factor in the formulation of SR. The theory to be created had to elude the quantum conflict, and the way to achieve that was to avoid commitments to electrodynamic models, that is, by an approach *of principle*:

> Reflections of this type [the quantum-issues] made it clear to me as long ago as shortly after 1900, i.e., shortly after Planck's trailblazing work, that neither mechanics nor thermodynamics could (except in limiting cases) claim exact validity. By and by I despaired of the possibility of discovering the true laws by means of constructive efforts based on known facts. The longer and the more despairingly I tried, the more I came to the conviction that only the discovery of a universal formal principle could lead us to assured results. The example I saw before me was thermodynamics. The general principle was there given in the theorem: the laws of nature are such that it is impossible to construct a *perpetuum mobile* (of the first and second kind). (Einstein 1959, 51-3).[104]

[104] Janssen does not attribute the rise of quantum physics a crucial important in the resolution of the case we are considering: 'According to this electromagnetic program, the common origin underlying universal Lorentz invariance is that all matter is made of electromagnetic fields and is thus governed by Maxwell's equations. Since Maxwell's equations are Lorentz invariant, all systems in motion must contract. The program initially showed promise, but it did not pan out. Special relativity thus carried the day, but not thanks to this or any other *COI*. Variants of the *COI* laid out above, in conjunction with arguments from nascent quantum theory, removed Lorentz's theory from serious consideration, but did not decide between special relativity and the electromagnetic theory' (Janssen 2002b, p. 499). I disagree with the last sentence. As I just argued, the empirical confirmations of early quantum theory provided empirical evidence to reject Lorentz theory, whereas that evidence was neutral with respect to special relativity.

Before we consider a second way in which empirical considerations can settle the case of Einstein vs. Lorentz, it is interesting to mention that the previous remarks show that, after all, explanatory issues are conceptually and historically relevant. However, the explanatory features of SR are significant to make a definitive choice if we consider them not *only* as non-empirical features. The (kinematic) explanations of principle of special relativity are superior to the (dynamic) constructive explanations of Lorentz's theory *because the former are not in conflict with the quantum hypothesis and its empirical confirmations*[105]. That is, the explanatory superiority of Einstein's theory is grounded on the fact that the kind of explanations it offers do not lead to empirical problems.

2.5.6 Special and general relativity

In order to explain the third reason that can be used to decide the case between the theories at issue it is necessary to take a glimpse on Einstein's road to the formulation of the general theory of relativity (GR)[106]. Soon after he introduced SR, he thought that the principle of relativity should be extended and generalized to encompass also non-inertial motion. He found that the reference to inertial forces included in his 1905 theory –which in turn refer to absolute space-time– was a problematic feature, for example. By applying the relativity postulate to accelerated motion, he believed, the reference to inertial forces would be no longer needed. On the other hand, Einstein also wanted to develop a theory of gravitation grounded on his results of 1905.

The crucial insight that allowed him to tackle both issues came from what he called *the happiest idea of his life*. This idea, which is the seed of his crucial *principle of equivalence*, was simply that a freely falling observer in a gravitational field does not feel his own weight, so that if the observer were to perform experiments in physics their outcome would be the same as if the he was at rest or in inertial motion. The general conclusion of this thought experiment is that a frame which freely falls in a homogenous gravitational field is equivalent to an inertial frame –and this statement constitutes the 'first half' of the principle of equivalence. On the other hand, it is also the case that an observer in an elevator at rest in a homogeneous gravitational field –which is therefore resisting the pull of the massive object which generates the field– is indistinguishable from a case in which an observer is inside an elevator in outer space, free from any gravitational attraction, but in accelerated motion—with the acceleration having the same value of the gravitational pull but opposite direction. More generally, a frame at rest in a homogenous gravitational field is physically equivalent to a frame accelerating with the same value but in the opposite direction of the attraction of the field –and this statement constitutes the second part of the principle of equivalence[107].

The connection just stated between acceleration and gravitation was thus the milestone for Einstein's path to GR. It allowed a solution for a long-lasting puzzle in the context of Newton's theory: the coincidental equality of the values of inertial and gravitational mass for any object. Though in this theory *inertial mass*, the 'stubbornness' of bodies to remain in their state of motion, and *gravitational mass*, the measure of a body's tendency to gravitate, represent different physical properties, have mutually identical values for and each and every body—and the theory does not provide any explanation for this coincidence[108]. The link that the principle of equivalence establishes between acceleration and gravitation allowed a way out of the puzzle: inertial and gravitational mass are the same property. And if acceleration

[105] If we want to avoid to get troubled by the dispute about the correct interpretation of special relativity, we can simply omit the adjectives 'kinematic' and 'of principle'.

[106] On this subject I closely follow (Torretti 2003, § 1.4).

[107] The restriction to homogeneous gravitational fields is required to neglect the effects of the different values of the gravitational potential in different points of the field in the real cases, and the effects produced by the fact that the vectors of the field are radials, not parallel. Such effects would, of course, break the equivalence.

[108] More precisely, consider the Newtonian equations $\mathbf{F} = m_i\mathbf{a}$, where m_i is the inertial mass; and $\mathbf{W} = m_g g$, where $\mathbf{W}$ is the weight, m_g the gravitational mass, and g the acceleration of gravity. Applying these equations for the acceleration in

and gravitation are two sides of the same coin, Einstein's goal of extending his theory to accelerated motion and to gravitational phenomena could be achieved by the same theoretical stroke.

The full significance of the principle of equivalence was put in front of Einstein's eyes by means of a group of problematic features he found. Maybe the most important one was given by the analysis of what happens in a rigid rotating disk. Imagine a flatlander equipped with a measuring rod living on the surface of the disk. If he were to make a measurement of the circumference of the disk he would find it to be greater than $2\pi r$, for his rod would get contracted when put along circumference but would remain the same when put along the radius –the contraction occurs only in the direction of motion, and the radius is always transversal to it. The surprising result is then that the geometry of the rotating disk is not Euclidean, and given the principle of equivalence, the geometry in a frame at rest in a gravitational field is not Euclidean either.

The theoretical framework that Einstein created to provide an explanation of this issue was the result of a sort of analogical reasoning. During the 19th century Gauss had created mathematical methods to deal with the intrinsic geometric properties of curved surfaces. If x^1 and x^2 are Cartesian coordinates on a surface, the length of a line C in that surface is given by the expression $\int_C ds = \int_C \sqrt{(dx^1)^2 + (dx^2)^2}$. The *line element* ds can also be written by means of the Kronecker symbol, such that $\int_C ds = \int_C \sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2} \delta_{ij} dx^i dx^j}$. If the Cartesian coordinates are replaced by *curved* coordinates u^1 and u^2, then the lines defined by $u^1 = constant$ and $u^2 = constant$ are curves which form variable angles in their intersection points—whereas the corresponding lines defined by $x^1 = constant$ and $x^2 = constant$ in the Cartesian case *always* cut each other orthogonally. The length of a line in terms of the curved coordinates can be expressed in a similar way, but the constant factors δ_{ij} must be replaced by factors that vary with position. Designating those factors as g_{ij}, then the length of a curve can be stated as $\int_C ds = \int_C \sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2} g_{ij} du^i du^j}$; and a geodesic line in the curved surface is given by $\delta \int_C \sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2} g_{ij} du^i du^j} = 0$. The line elements $\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2} \delta_{ij} dx^i dx^j}$ and $\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2} g_{ij} du^i du^j}$ are, of course, different[109]. The former is defined for Cartesian coordinates, whereas the latter is defined for curved ones. However, they coincide when infinitesimally or locally considered. Locally or infinitesimally speaking, the expression for length in curved coordinates reduces to the expression for Cartesian ones. Finally, the approach introduced by Gauss also allowed to quantitatively evaluate the degree of curvature of a surface. For example, flat surfaces have a 0 curvature, the surface of a sphere is constant and positive, and the surface of an egg is positive but varies with position.

The adoption of the Minkowskian point of view allowed Einstein to notice the analogy between the geometrical work of Gauss and the results he obtained from his principle of equivalence. In an article of 1912, entitled *On the Theory of the Static Gravitational Field*, Einstein wrote the equation for the world-line of a material point freely falling in a static gravitational field as $\delta\left\{\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}\right\} = 0$. As I showed above, this is the equation for a time-like geodesic in a Minkowskian space-time, the equation for an *inertially moving* particle. However, one of the results that his principle of equivalence entailed was that light gravitates, and therefore in inhomogeneous gravitational fields the speed of light depends on the specific position-dependent value of the gravitational potential. Consequently, a more accurate and general way to write the equation was $\delta\left\{\sqrt{[c(x,y,z)]^2dt^2 - dx^2 - dy^2 - dz^2}\right\} = 0$. If we replace x, y, z, t with x^0, x^1, x^2, x^3, and if we define $g_{00} = c(x^1, x^2, x^3)$, $g_{kk} = -1$ if k is greater than 0, and $g_{kh} = 0$ if $k \neq$

free falling—substituting **W** for **F**—one obtains $\mathbf{a} = (m_g/m_i)g$. As experience shows, the gravitational acceleration exerted on *any* freely falling body is the same. This can only be the case if the ratio between the two different masses is the same for any object. Within the context of Newton's gravitation theory, this is the unexplained coincidence.

[109] More precisely, only in a Euclidean plane it is possible to define Cartesian coordinates x^1 and x^2 such that the g_{ij} quantities—expressed as a function of the coordinates—satisfy the relation $g_{ij} = \delta_{ij}$.

h; then the equation can be written as $\delta\int_C\sqrt{\sum_{i=0}^{3}\sum_{i=0}^{3}g_{ij}dx^i dx^j}=0$. A decisive point in Einstein's road to his gravitational theory was to notice that this equation is different from a Gaussian geodesic only by the amount of dimensions at stake, so that the decisive analogy that Einstein saw was that it represents the equation for a geodesic in space-time. It had to be a *curved* space-time because he had already discovered that the spatial metric of an accelerated frame—and of its equivalent frame at rest in a gravitational field—is not Euclidean, and thus the spatio-temporal metric in those frames could not be Minkowskian; i.e., the g_{ij} factors are position-dependent. The curvature expressed by the specific g_{ij} was associated to acceleration and gravitation determining the space-time considered: the non-Euclidean geometry of the rotating disk was the outcome of its acceleration, and by the principle of equivalence one has that gravitation results in the same feature.

This line of reasoning naturally suggested Einstein that gravitation is not a force acting at a distance between distant massive bodies, but simply the curvature of space-time that the bodies produce. Since $\delta\int_C\sqrt{\sum_{i=0}^{3}\sum_{i=0}^{3}g_{ij}dx^i dx^j}=0$ defines a time-like geodesic in a curved space-time, it defines the *inertial* motion of a particle in a region of space-time determined by a gravitational field: free-fall is simply inertial motion in a curved region of space-time, and the curvature is produced by the presence of mass-energy. This is gravitation, not a force nor an instantaneous action at a distance. Yet another important feature of the analogy with the Gaussian approach was that the 'Minkowskian line element' relates to the line-element defined for the curved space-time case just as the Cartesian line element relates to the one defined for curved coordinates; i.e., as a local-infinitesimal region of the global surface.

This was, in a nutshell, the line of reasoning that led Einstein from SR to GR. The *Gaussian analogy* he discovered got reinforced by Einstein's collaboration with Marcel Grossman, who introduced him to the work of Riemann on the generalization of the Gaussian approach for n-dimensional manifolds. The road to the specific quantitative relation between the mass-energy distribution and the specific metric of space-time was a difficult and intricate one, but the goal was finally achieved in 1916 with the introduction of the field equations that form the core of the new theory.

Coming back to our subject, we have that the relevance that GR has with respect to the Einstein *vs.* Lorentz case is that it reduces the special theory to a limiting case. The Minkowskian space-time that SR defines is, from the point of view of the general theory, a specific solution of the field equations in which there is no mass-energy to produce any curvature-gravitation-field, or simply a local-infinitesimal piece of a global space-time defined by the field equations—just as the line element defined for Cartesian coordinates can be considered as determining an infinitesimal piece of a curved surface, according to the Gaussian approach. More simply, the Minkowskian world describes either an empty universe, or a tiny piece of a universe which does contain mass-energy.

The essential mathematical and *physical* connection between the special and general theories does not exist between the LPT and the GR. If we remind ourselves that in the context of the Lorentz-Poincaré view the invariant interval does *not* refer to the metric of a four-dimensional space-time—it is nothing but a *mathematical nicety*—we can easily see that we cannot say that the world the LPT describes is a specific solution of the field equations or a local-infinitesimal piece of a global curved space. It is true that the LPT can be presented as a *mathematical* consequence of Einstein's field-equations under certain constraints—its mathematical structure is identical to SR, after all. However, from a physical-semantic point of view the connection does not hold. GR states that an empty universe would be determined by a Minkowskian *metric*, and that locally speaking the *metric* of a curved space-time can be described by the 'Minkowskian line element'. But the invariant interval, within the context of the LPT, has nothing to do with the metric of space-time, it is simply a mathematical nicety produced by the specific value of compensating dynamical effects such as the Lorentz-Fitzgerald contraction and local time. The real metric of the space-time that the theory defines is Newtonian. More generally, it is the *physical meaning* of GR what

precludes that the LPT could be understood as the expression of a limiting case. In order to do that, the meaning of GR should have to be severely altered, in a way such that the theories could become coherent. Therefore, the acceptance of GR does *not* entail the acceptance of the LPT. Moreover, since the former claims that an empty universe would have a *Minkowskian metric* and that a *Minkowskian metric* describes an infinitesimal portion of a curved space-time, its acceptance entails a *rejection* of the LPT, for this theory claims that the metric of space-time would be *Newtonian* in both cases. More generally, even though the mathematical structure of the LPT can be derived from GR, the different meanings of the two theories make them incompatible.

The crucial point is that GR entails empirical consequences that SR does not. As a result, SR 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 provide empirical evidence for general relativity, therefore, those phenomena are also evidence for the view that the Minkowski geometry is the geometry of the tangent space-time associated to each point of a global 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 theory. Thus, the evidence given by the successful predictions of GR that flow to SR but that cannot flow to the LPT work as a tool to break the UD between the theories, even though the EE between them remains[110].

It is important to remark, that, as mentioned above, the acceptance of Einstein's theory and the abandonment of Lorentz's by the scientific community of the time occurred around 1911. That is, by 1919—the year when the bombastic acceptance of general relativity produced by Eddington's observation of the 'light-bending' effect—the case of Einstein vs. Lorentz had been already decided. Therefore, this second argument based on non-entailed empirical evidence was historically irrelevant, it only works from a conceptual point of view.

Now that we have gone through a detailed evaluation of the case of Einstein vs. Lorentz, we can move on to the second case-study to be considered.

---

[110] This solution in terms of non-entailed empirical evidence could be challenged if, besides GR, there were an alternative gravitational theory able to encompass the LPT, or at least a suitable interpretation of GR that allows its compatibility with the LPT—thus restoring the UD (recall Bangu's objection explained in section 1.4.2). Lorentz actually prposed some arguments stating that GR implied a restoration of the ether. Even Einstein himself expressed that since GR endows 'space itself' with physical properties (it is 'metrically sensitive' to the presence of matter), then metric field could be interpreted as a kind of 'ether'—though such an 'ether' cannot be associated to a privileged reference frame. However, the ether that Lorentz was talking about was still the medium on which electromagnetic waves propagate, that is, the ether that determines a preferred frame. Lorentz's views on this matter are incorrect, so that his attempt to reintroduce the substantial ether did not work (see Illy 1989, and Kox 1988).
Another strategy could be not to interpret GR as a chrono-geometric theory. That is, to state that the geodesics of a curved space-time are just a mathematical devices that allow to refer to gravitation, whereas the *real* picture is that gravitation is universal force that acts in a Newtonian space-time. However, this is mere speculation. As Kyle Stanford puts it: 'While Eddington, Reichenbach, Schlick and others have famously agreed that General Relativity is empirically equivalent to a Newtonian gravitational theory with compensating "universal forces", the Newtonian variant has never been given a precise mathematical formulation (the talk of universal forces is invariably left as a promissory note), and it is not at all clear that it can be given one. (David Malament has made this point to me in conversation). The "forces" in question would have to act in ways no ordinary forces act (including gravitation) or any forces could act insofar as they bear even a family resemblance to ordinary ones; in the end, such "forces" are no better than "phantom effects" and we are left with just another skeptical fantasy. At a minimum, defenders of this example have not done the work needed to show that we are faced with a credible case of non-skeptical empirical equivalence' (2001, S6). That is, the choice favoring SR in terms of non-entailed empirical evidence provided by GR remains unchallenged.

CHAPTER 3

STANDARD QUANTUM MECHANICS VS. BOHM'S THEORY

The second example of EE and UD I will address is given by the case of standard quantum mechanics (SQM) vs. David Bohm's quantum theory (BQT). I will proceed in an analogous way as in chapter two. In the first section of this chapter I provide a historical outlook of the development of quantum theory. In the second section I offer a detailed exposition of the essentials of quantum mechanics as axiomatized by von Neumann, paying special attention to the features of the theory that are relevant for a comparison with Bohm's theory. In the third section present an overview of the main interpretations of quantum mechanics in Hilbert space, namely, Bohr's complementarity, decoherent histories, Everettian approaches, and the modal interpretation. In the fourth section I describe the motivations of the 'hidden variables' view of quantum physics, and I analyze the constraints that theoretical considerations and empirical adequacy impose on any viable hidden variable theory. I the fifth section I offer a detailed presentation of Bohm's quantum theory. In the sixth and final section I undertake an evaluative comparison between the theories at issue.

3.1 HISTORICAL OUTLOOK[1]

In order to setup the conceptual background required for an adequate grasp of SQM, I will first present a historical outlook of the main stages in the development of the theory. This outlook will also allow us to understand all the strange and revolutionary physical conceptions that quantum physics brought about. This is especially important, since most these strange features are relevant when it comes to an evaluative comparison between SQM and BQT.

3.1.1 Planck and the quantum of energy

The first stage in the historical development of QM began with the work of Max Planck on the problem of blackbody radiation. This problem has its roots on the work of Gustav Kirchhoff on the absorption of heat and light. In a famous paper of 1859, Kirchhoff determined that the ratio of emissive power to absorptivity is the same for all bodies. He showed this by considering two parallel plates facing each other and covered by ideal reflectors. Assuming that one of the plates emits and absorbs radiations of all wavelengths and that the other emits and absorbs only in the frequency λ, it turns out that all the radiation of wavelength different from λ is absorbed by the first plate after repeated reflections, so that the calculations can be restricted only to the λ-wavelength radiation. The result that Kirchhoff obtained was $e/a = E/A$, where e and E stand for the emissive power of the plates, and a and A stand for their absorptivity. The relevance of this law is clear, the ratio between the emissive power and the absorptivity of bodies is a function only of the temperature and wavelength (or frequency) of the radiation involved, and independent of the specific materials they are made of.

In order to stress out the fundamentality of his law, Kirchhoff introduced the concept of a *blackbody*, a body which absorbs all incident radiation, and he also established that cavity radiation inside an enclosure with walls at an equal temperature T is equivalent to the emissive power E of a blackbody. The

[1] This historical overview is based on (Segre 1980), Hermann (1971), Hund (1974), and, especially, on (Jammer 1966). Full references to all the relevant original papers can be found in Jammer's book.

absorptivity index of a blackbody is 1, and therefore its emissive power is given by a function of the temperature and frequency. More formally, Kirchhoff showed that u_ν, the *density of radiation in a frequency interval* ν within a blackbody, is given by a universal function $u(\nu, T)$.

Given the fundamental and universal character of Kirchhoff's law, to determine the specific form of the mentioned function became one of the most important tasks in thermodynamics during the last decades of the 19th century. The first step in the achievement of this goal was given by the work of J. Stefan. In an 1879 paper, he proposed that u is proportional to T^4. Stefan's law, $u(T) = \sigma T^4$, was provided a theoretical proof by Boltzmann in 1884 and received empirical confirmation in 1897 through experiments carried out by Paschen, by Lummer and Pringsheim, and by Menenhall and Saunders. A second step forward came with the work of Wien in 1894. He proposed a formula $u(\nu, T) = \nu^3 f(\nu/T)$, where f was an as yet undetermined function. Taking the derivative with respect to ν in Wien's formula and determining where it takes a 0 value, we find the frequency at which, at a given temperature, $u(\nu, T)$ is maximum. Calculations and graphic representation show that as T increases the corresponding ν-value at which u is maximum slightly increases—that is, $\nu_{max} \propto T$. For this reason Wien's formula became commonly known as Wien's *displacement* law, and it got experimentally confirmed by Paschen in 1899.

Wien attempted a definitive and complete specification of the function $u(\nu, T)$ in 1896. Considering a model in which molecules emit radiation whose frequency depends on the molecules' velocity, and using statistical mechanics, he proposed a formula $u(\nu, T) = \alpha\nu^3 e^{-(\beta\nu/T)}$, where α and β are constants. According to the experimental data then available, Wien's formula seemed correct. This was the situation when Max Planck, an authority in classic thermodynamics, got first involved in the subject. In a series of papers he presented between 1897 and 1899, Planck undertook the task of providing a firm theoretical basis for Wien's proposal[2]. His key assumption was—exploiting the fact that the specific nature of the body was irrelevant—to consider the walls of the blackbody as composed of harmonic oscillators in order to establish a state of radiation equilibrium. By means of electrodynamic reasoning he obtained the formula $u(\nu, T) = \frac{8\pi}{c^3}\nu^2 U(\nu, T)$, where U stands for the average energy of the resonators. Had he used the equipartition theorem[3] to determine the form of the function U, he would have inexorably obtained the Rayleigh-Jeans law—which only worked in the low frequency range of the spectrum and had absurd consequences in the short wavelengths case. However, for some reason[4], he took a different approach. He defined the entropy of a resonator as $S = -\frac{U}{b\nu}\log\frac{U}{ea\nu}$, where a and b are constants, and from this expression he obtained a formula for the energy of resonators $U = a\nu^3 e^{-(b\nu/T)}$ [5], and plugging this formula in $u(\nu, T) = \frac{8\pi}{c^3}\nu^2 U(\nu, T)$, Wien's formula can be obtained in the form $u(\nu, T) = \frac{8\pi a}{c^3}\nu^3 e^{-(b\nu/T)}$ [6]. Planck interpreted this result as a proof that Wien's formula was a necessary consequence of the core principles of thermodynamics.

---

[2] As Max Jammer puts it: 'since cavity radiation is a process related to electromagnetism rather than to the kinetic theory of gases, Planck decided that he could do the same for Maxwell's theory of the electromagnetic field as Boltzmann had done for mechanics with his famous, though still disputed, H theorem. Planck thus attempted to prove that Maxwell-Hertz equations, if applied to resonators with arbitrary small conditions, would lead to irreversible processes converging toward a stationary state whose energy distribution is that of cavity radiation and which therefore determines the energy spectrum of black-body radiation. In short, what Planck seems to have had in mind was a translation of the reasoning that led to the Maxwell-Boltzmann velocity distribution on the basis of the kinetic theory into the conceptual structure of electromagnetic theory' (Jammer 1966, 11).

[3] This theorem states that the kinetic energy of the molecules of a gas is equally distributed among their degrees of freedom.

[4] Jammer (1966, 12-14) states that it is not clear why Planck did not use the equipartition theorem, possible motives might have been his unfamiliarity with the statistical methods in thermodynamics, the experimental difficulties that the theorem was facing at the turn of the century, his aversion to the molecular approach, and his conviction regarding the fruitfulness of thermodynamical reasoning grounded on the concept of entropy.

[5] Since this entropy must satisfy the equation $dS = \frac{dU}{T}$, Planck could derive $\frac{1}{T} = \frac{dS}{dU} = -\left(\frac{1}{b\nu}\log\frac{U}{ea\nu} - \frac{1}{b\nu}\right) = -\frac{1}{b\nu}\log\frac{U}{a\nu}$, and then solving for U, $U = a\nu e^{-(b\nu/T)}$.

[6] In this equation Planck's constant appears for the first time. Planck determined the value of constant $a = 6{,}885 \times 10^{-27}$erg sec. In current terminology, $\beta = h/k$, where k is Boltzmann's constant.

The empirical success of the Wien-Planck law did not last long. In 1899 experiments carried out by Lummer and Pringsheim showed important deviations between the results obtained and the predictions of the equation in the range of small frequencies of the spectrum. On the other hand, in June 1900, Lord Rayleigh introduced yet another formula for $u(\nu,T)$. He derived the expression $u(\nu,T)\sim\nu^2U(\nu,T)$, and, unlike Planck, determined the function U from the equipartition theorem to obtain $U(\nu,T)=kT$, where k is Boltzmann's constant. Therefore, Rayleigh's law took the form $u(\nu,T)\sim\nu^2kT$. James Jeans introduced some correction in Rayleigh's reasoning that in 1905 allowed to formulate the Rayleigh-Jeans law in a more specific way, namely, $u(\nu,T)=\frac{8\pi}{c^3}\nu^2kT$. This formula agreed with experimental results where the Wien-Planck law failed, but it led to absurd consequences in the range of high frequencies. As a quick inspection of the formula show, for high frequencies the integral for the total energy density u diverges[7]. However, Rubens and Kurlbaum reported to Planck that the new experimental results made clear that for low ν and high T, the energy density u had to be proportional to T, in accordance with the Rayleigh-Jeans law and in disagreement with the Wien-Planck law.

Rubens and Kurlbaum's observation convinced Planck that a new law was needed. His point of departure was again his definition of the entropy of a resonator $S=-\frac{U}{b\nu}\log\frac{U}{ea\nu}$, though this time he knew that he needed a formula which in the case of high ν and low T agrees with Wien's formula and in the case of low ν and high T reduces to a proportionality of u to T. He used the second derivative of S with respect to U as a reference. In the case of his 1899 definition of entropy it held that $\frac{d^2S}{dU^2}\propto\frac{1}{U}$, but this result failed for the low ν and high T conditions. To obtain a proportionality of u to T under these conditions he needed that $\frac{d^2S}{dU^2}\propto-\frac{1}{U^2}$. Planck noticed that he simply needed an interpolation of these two second derivatives, so that he proposed $\frac{d^2S}{dU^2}=\frac{\alpha}{U(\beta+U)}$, where α and β are constants. From this expression he obtained a corrected equation for U, namely, $U=\frac{\beta}{e^{(\beta/\alpha T)}-1}$ [8]. Finally, by inserting this equation in $u(\nu,T)=\frac{8\pi}{c^3}\nu^2U(\nu,T)$, Planck obtained his famous law:

$$u(\nu,T)=\frac{8\pi\nu^2}{c^3}\,\frac{\beta}{e^{(\beta/\alpha T)}-1}$$

Planck communicated this result on October 19th, 1900, at the University of Berlin. That same night Rubens and Kurlbaum checked the formula against the experimental data and found close agreement. It is clear that Planck obtained his law from a mere mathematical interpolation between $\frac{d^2S}{dU^2}\propto-\frac{1}{U}$ and $\frac{d^2S}{dU^2}\propto-\frac{1}{U^2}$, so he naturally felt the urge to provide a firm theoretical foundation for it[9]. To do so, Planck turned

[7] Helge Kragh (2000) compellingly argues that the relevance of the Rayleigh-Jeans law in the historical and conceptual assessment of the formulation of Planck's definitive law of blackbody radiation has been overestimated. The absurd consequences of the formula became a central point of discussion only after 1905. As we saw in chapter 2, Lorentz showed that this law was the unavoidable consequence of the application of the principles of classical mechanics (equipartition theorem) in 1908 (see Kox 2013), and Ehrenfest coined the term 'ultraviolet catastrophe' only in 1911. That is, when Planck formulated his law there was not yet a sense of a deep crisis of classical mechanics arising from blackbody thermodynamics. He wanted a correct law for blackbody radiation, he was not responding to a breakdown of classical physics.

[8] Integrating $\frac{d^2S}{dU^2}=-\frac{\alpha}{U(\beta+U)}$ with respect to dU he got $\frac{dS}{dU}=\frac{1}{T}=\frac{\alpha}{\beta}\log\frac{\beta+U}{U}$, and solving this last equation for U, one obtains $U=\frac{\beta}{e^{(\beta/\alpha T)}-1}$.

[9] As Jammer puts it: 'This interpolation, though mathematically a mere trifle, was one of the most significant and momentous contributions ever made in the history of physics. Not only it lead Planck, in his search for his logical corroboration, to the proposal of his elementary quantum of action and thus initiate the early development of quantum theory […]; it also contained certain implications which, once recognized by Einstein, affected decisively the very foundations of physics as well as their epistemological presuppositions. Never in the history of physics was there such an inconspicuous mathematical interpolation with such far-reaching physical and philosophical consequences', (Jammer, 1966, 18).

to Boltzmann's probabilistic conception of entropy, which he had as yet avoided and disliked. He defined the entropy of a system of N resonators as $S_N = k \log W$, where k is Boltzmann's constant and W stands for the number of distributions compatible with the energy of the system. To determine W, Planck assumed that the total energy $U_N = NU$ consists of an integer number P of energy elements ϵ, so that $U_N = P_\epsilon$—for the conception of U_N as a continuous magnitude did not admit a combinatorial method to establish W. Then he assumed that the energy of each energy element is given by $\epsilon = h\nu$, where h is the famous constant that takes his name. This means that the total energy U_N is composed by discrete packets of energy[10]. Under this heuristic framework and applying combinatorial calculations Planck obtained an expression for the energy of the oscillators $= \frac{\epsilon}{e^{(\epsilon/kT)}-1}$[11], so that the definitive form of Planck's law is given by

$$u(\nu, T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{(h\nu/kT)} - 1}$$

This formula was consistent both with Stefan's law and Wien's displacement law, it was in close agreement with experimental data, and allowed to determine values for the constants $h = 6{,}55 \times 10^{-27}$erg sec and $k = 1{,}346 \times 10^{-16}$ erg/°C, for Avogadro's number $N = 6{,}175 \times 10^{23}\text{mole}^{-1}$, and for the elementary unit charge $e = 4{,}69 \times 10^{-10}$esu.

Planck presented these results in December 14th, 1900. In standard historical accounts this date is considered to be the birth of QM, for the discrete quantization of energy is taken to be the first essential feature of quantum physics ever discovered. However, this standard view has been challenged. The main reason is that neither Planck himself nor other members of the scientific community of the time interpreted $\epsilon = h\nu$ as an expression of such discrete quantization. The only place where Planck refers to the meaning of the constant h in connection with the energy elements ϵ is the passage I quoted in footnote 9, and that text is not enough to unmistakably understand that he did consider the energy elements as really quantized—he might have conceived the division in discrete elements as a mathematical tool, pretty much in the spirit of Boltzmann's statistical and combinatorial methods. Planck showed an openly and explicit commitment to the physical quantization view only in 1908[12]. Anyhow, in 1905 Einstein made another groundbreaking contribution in the development of QM, and he was fully and clearly aware that a radical change in the course of physics was taking place. I now turn to his work.

[10] As Planck himself expressed the point: 'Now we have to consider the distribution of the energy U_N among the N resonators of frequency ν. If U_N were regarded as an infinitively divisible quantity, the distribution could be performed in an infinite number of ways. We consider, however—and this is the cardinal point of the whole computation—U_N as composed of a finite number of discrete equal parts and employ for this purpose the natural constant $h = 6{,}55 \times 10^{-27}$ erg sec. This constant multiplied by the common frequency ν of the resonators gives the energy element ϵ in ergs, and by dividing U_N by ϵ we obtain the number P of energy elements which are distributed among the N resonators', (from Planck 1900, quoted in Jammer 1966, 20).

[11] Planck interpreted W in $S_N = k \log W$ as the number of possible ways to distribute P energy elements ϵ among N oscillators, so that $W = \frac{(N+P-1)!}{(N-1)!} = \frac{(N+P)^{N+P}}{N^N P^P}$. With this expression he obtained $S_N = k[(N+P)\log(N+P) - N\log N - P\log P] = kN\left[\left(1+\frac{U}{\epsilon}\right)\log\left(1+\frac{U}{\epsilon}\right) - \frac{U}{\epsilon}\log\frac{U}{\epsilon}\right]$. From equation $u(\nu, T) = \frac{8\pi}{c^3}\nu^2 U(\nu, T)$ and Wien's displacement law $u(\nu, T) = \alpha\nu^3 f(\nu/T)$ it follows that the entropy of the resonators must be a function of the form $S(U/\nu)$, so that $S_N = kN\left[\left(1+\frac{U}{h\nu}\right)\log\left(1+\frac{U}{h\nu}\right) - \frac{U}{h\nu}\log\frac{U}{h\nu}\right]$. Finally, from $\frac{dS}{dU} = \frac{1}{T}$, Planck obtained $U = \frac{h\nu}{e^{(h\nu/kT)}-1}$.

[12] Kuhn (1978) argues that Planck actually unquestionably understood his use of $\epsilon = h\nu$ as a mere mathematical device. For a critical assessment of Kuhn's view, see (Gearhart 2002). There the author compellingly argues that Planck's attitude with respect to the mentioned formula was essentially ambiguous.

3.1.2 Einstein and the quantum of light

Aside from the specific way in which the scientific community understood Planck's quantum, it was nevertheless clear that the quantization was confined to the interaction between resonators and emitted energy. That is, it was the material oscillators that could emit or absorb energy in multiples of $h\nu$. In Einstein's work, however, the idea of quantization was applied to the radiation itself. In his famous paper of 1905 entitled *On a Heuristic Viewpoint Concerning the Production and Transformation of Light*, Einstein undertook a similar task as Planck, he applied thermodynamic and statistical reasoning to a problem in which electrodynamics was involved.

His starting point was to consider low density radiation inside a cavity of a volume V for which Wien's radiation law $u(\nu,T)=\alpha\nu^3 e^{(-\beta\nu/T)}$ is valid, that is, in the range of short wavelengths. He introduced a function $\varphi(\nu,T)$ for the entropy of such radiation and set himself to determine φ, but as a function of (u,ν), where u stands for the energy density. Exploiting the fact that $dS=\frac{dU}{T}$, Einstein wrote $\frac{d\varphi(\nu,T)}{du(\nu,T)}=\frac{1}{T}$, and solving Wien's radiation formula for $1/T$ he obtained $\frac{1}{T}=-\frac{1}{\beta\nu}\log\frac{u}{\alpha\nu^3}=\frac{d\varphi}{du}$, and after integration he got $\varphi(u,\nu)=-\frac{u}{\beta\nu}\left(\log\frac{u}{\alpha\nu^3}-1\right)$. Accordingly, the entropy of the radiation in a frequency interval $d\nu$ in a volume V is $S=-\frac{U}{\beta\nu}\left(\log\frac{U}{\alpha V\nu^3 d\nu}-1\right)$, and thus the expression for entropy change of the radiation originally within a volume V_0 and later within a volume V is given by $S-S_0=\frac{U}{\beta\nu}\log\frac{V}{V_0}$. Since $\beta=h/k$, this last equation can be written as $S-S_0=\frac{kU}{h\nu}\log\frac{V}{V_0}=k\log\left(\frac{V}{V_0}\right)^{\frac{U}{h\nu}}$. On the other hand, according to Boltzmann's entropy formula, Einstein noticed, the entropy change for a gas of n particles that is first contained in a volume V_0 and after in a volume V is given by $S-S_0=k\log\left(\frac{V}{V_0}\right)^n$. By the identical structures of the formulas Einstein set

$$U=n(h\nu),$$

which means that the radiation within volume V can be considered as built up of n energy quanta of magnitude $h\nu$—or *photons*, as we nowadays call them[13].

As it is clear, by means of Einstein's maneuver radiation itself gets quantized. This result was blatantly incompatible with the wave theory of light that had been developed since the times of Huygens and Young, and also with the standard understanding of Maxwell's equations of the electromagnetic fields. Einstein proposed a justification for his disruptive argument by stating that the observations which strongly confirm the classical theory of light considered only time averages, and suggested that when it comes to instantaneous values of the wave functions the theory might break down.

Einstein's formula $\epsilon=h\nu$ allowed a simple explanation for a curious empirical phenomenon, the photoelectric effect. In a series of experiments in the late 1890s and early 1900s, P. Lenard observed that incident light on a metallic plates provoked the emission of electrons from the latter. The strange features of this effect were that, contrary to Maxwell's theory of light, electrons were emitted only if the incident light was of a frequency higher than a minimum threshold, and that the energy of the emitted electrons independent of the light's intensity—though, over the threshold, the amount of electrons emitted was indeed proportional to the intensity. However, according to Einstein's proposal it could be considered that each quantum of light transferred its energy $h\nu$ to a single electron in the plate. The minimum threshold of frequency got naturally explained because the removal of an electron from the plate required an amount of work P to be performed (P is a constant for each metal), but if the frequency of an incident

[13] In Einstein's terminology, $n=NU/R\beta\nu$, where R is the constant and N is Avogadro's number. Since $R/N=k$ and $\beta=h/k$, then $U=n(h\nu)$.

photon is such that $h\nu < P$ no electron is emitted. This also explains the independence of the energy of the electrons with respect to the intensity of light. If a beam of light is constituted by photons such that their energy is lower than P, it will not liberate any electrons, no matter how intense the beam may be—but if the energy of the photons in the beam is higher than P, higher intensity beams (more photons), liberate more electrons. Finally, provided that $h\nu \geq P$, the higher the frequency of the incident photon, the higher the energy of the emitted electron. This was a novel prediction contained in Einstein's work. The maximum kinetic energy of the emitted electron is given by $h\nu - P$, and the stopping potential—the potential needed to prevent an electron of reaching a second metallic plate, for example—for an electron must be a linear function of the frequency of the incident light, whose slope is independent of the nature of the metal from which the electron is emitted. This prediction was confirmed in the laboratory by R. Millikan in in a series of experiments performed between 1908 and 1916. The constant of proportionality thereby obtained was in close agreement with the value of h determined by Planck.

Another revolutionary feature of Einstein's proposal was recognized by Planck in 1908. He strengthened the view that $\epsilon = h\nu$ as applied to radiation itself means that light is made out of particles. Elaborating on Einstein's relativity and electrodynamics he showed that the mentioned formula implies that light quanta carry a momentum given by $h\nu/c$, and if particles are considered as carriers of energy and momentum, it follows that photons are indeed particles. However, and in spite of the successful account of the photoelectric effect, Einstein's photon was considered as too revolutionary by the community of the time. Lorentz, for example, rejected the hypothesis on the basis of the experimental results of Lemmer and Gehrcke of 1902—if their experiments concerning interference were considered it followed that the spatial locations of photons had to be spread along one meter. He also mentioned other interference phenomena that seemed to be in contradiction with a particulate view of light. As it is clear, Lorentz remarks were an early manifestation of the perplexities aroused by the wave-particle duality of quantum physics that de Broglie and Schrödinger were to introduce as a central feature of the new theory some years later.

Wider and more definitive acceptance of Einstein's quantum of light came many years later through a series of experiments performed by A. Compton. In 1921 he had shown that if an X-ray is 'scattered' by a particle, such as an electron, the scattered ray got its wavelength increased and its frequency diminished. He tried to explain this effect according to classic electrodynamics using the Doppler effect, but he got convinced that an explanation along this line was just impossible. In 1923, though, he decided to refer to Einstein's hypothesis of the corpuscle of light: he simply assumed that each quantum in the X-ray was concentrated in a single particle and that it so interacted with a single electron. Under this assumption a natural explanation became available if the principles of conservation of momentum and energy were considered. That is, the scattering could be considered simply as a collision between particles. Compton found that the increase of wavelength was given by $\Delta\lambda = \left(\frac{h}{mc}\right)(1 - \cos\theta)$, where θ stands for the scattering angle and, mc represents the momentum of the quantum in the X-ray. This formula was in full agreement with observational data. That is, the way to explain the Compton effect was to assume Einstein's hypothesis of particulate radiation.

Three other contributions on the quantum made by Einstein are worth to be mentioned. First, in 1906 paper entitled *On the Theory of Light Emission and Absorption*, he showed that Planck's derivation of the blackbody law was based on inconsistent assumptions. He remarked that Planck obtained $u(\nu, T) = \frac{8\pi}{c^3}\nu^2 U$ from classic electrodynamics, so that he implicitly endorsed the view that the energy of the oscillators must be capable to range over a continuum. However, he then determined $U = \frac{h\nu}{e^{(h\nu/kT)}-1}$ which implied energy quantization in discrete packages. Einstein, with his particular insight, did not take this issue as a reason to reject Planck's reasoning, but as an indication that the foundations of classical electrodynamics had to be profoundly revised.

Second, in a 1907 paper, *Planck's Theory of Radiation and the Theory of Specific Heats*, Einstein showed that Planck's formula was the key to a problem that came up in the last two decades of the 19th century.

The Dulong-Petit formula for the specific heat of metals—introduced in 1819—that had been many times corroborated got threatened by measurements performed at low temperatures. According to statistical mechanics, the energy of a particle with three degrees of freedom is given by $3kT$[14], so that in a mole the energy U is this expression times Avogadro's number N, $3NkT$, and since $k = R/N$, then $U = 3RT$. The specific heat per mole $C = dU/dT$ is thus $3R$—Dulong-Petit formula—an expression which is clearly independent of the temperature. The problem with this law was that at low temperatures the specific heat of some measured metals decreased. Einstein showed that the formula for specific heat per mole obtained from Planck's law is $= 3R\frac{(h\nu/kT)^2 e^{(h\nu/kT)}}{(e^{(h\nu/kT)}-1)^2}$ [15], which explains the observed dependency of C on T—when $kT \gg h\nu$ the formula reduces to $3R$, but quantization becomes relevant at low T. This formula was not quite correct, but it showed the right path to solve the problem, and it was definitively amended by P. Debye in 1912. Einstein's solution of the specific heat problem contributed to clarify the relevance and fundamentality of Planck's quantum.

Finally, in a 1909 paper entitled *On the Present State of the Problem of Radiation*, from Planck's blackbody radiation law Einstein derived the expression $\overline{\varepsilon^2} = \bar{U}h\nu + \frac{c^3\bar{U}^2}{8\pi\nu^2 V d\nu}$ for the mean square energy fluctuation of energy within a cavity of volume V at a certain temperature. The second term in the sum gives the mean square energy fluctuation due to interference between partial waves—which is therefore the expected value for $\overline{\varepsilon^2}$ on the basis of classic electrodynamics. The first term in the sum, following Einstein's photon hypothesis, is the expression for $\overline{\varepsilon^2}$ if the radiation were considered as built up of light quanta of magnitude $h\nu$. Einstein interpreted the presence of this term as support for his light quantum hypothesis. Max Jammer states that this is the first manifestation of the wave-particle duality typical of quantum phenomena[16]. He interestingly argues that, rather than energy discrete quantization, it is this duality that paradigmatically defines QM[17]. He also observes that in the case of low frequencies—where the Rayleigh-Jeans law is valid—the first term in the expression for $\overline{\varepsilon^2}$ can be neglected in comparison with the second, and in this case the expression becomes consistent with $\frac{d^2S}{dU} \propto -\frac{1}{U^2}$; whereas in the case of high frequencies—where Wien's law is valid—$\overline{\varepsilon^2}$ reduces to the first term in the sum and the formula becomes consistent with $\frac{d^2S}{dU} = -\frac{1}{U}$. Therefore, the full formula $\overline{\varepsilon^2} = \bar{U}h\nu + \frac{c^3\bar{U}^2}{8\pi\nu^2 V d\nu}$ is consistent with Planck's interpolation $\frac{d^2S}{dU} = \frac{\alpha}{U(\beta+U)}$, and Jammer concludes that since the wave-particle duality expressed by the equation is already contained in such interpolation, the birth date of QM is October 19th, 1900, rather than December 14th (see Jammer 1966, 44-46).

The quantization of energy implied by Planck's law was not noticed or not seriously considered until ca. 1908. Einstein in 1905 elaborated on the hypothesis and established ostensibly revolutionary consequences. However, he was a rather lonely voice. It took several years for the light quantum hypothesis to be accepted in its full meaning by the physicists' community. It was the work of Niels Bohr in 1913 on

---

[14] More precisely, to each degree of freedom an energy $\frac{1}{2}kT$ corresponds, but since there two energy forms per degree of freedom (potential and kinetic), we have that $3kT$ is the total energy corresponding to the particle.

[15] Recall that in December 1900 Planck obtained $U(\nu, T) = \frac{h\nu}{e^{(h\nu/kT)}-1} = kT\frac{h\nu/kT}{e^{(h\nu/kT)}-1}$, which if applied to a mole and to three degrees of freedom becomes $U = 3RT\frac{h\nu/kT}{e^{(h\nu/kT)}-1}$. Taking the derivative with respect to T Einstein got $C = \frac{dU}{dT} = 3R\frac{(h\nu/kT)^2 e^{(h\nu/kT)}}{(e^{(h\nu/kT)}-1)^2}$.

[16] In Einstein's own prophetic words from his 1909 paper on energy fluctuations: 'It is undeniable that there is an extensive group of data concerning radiation which show that light has certain fundamental properties that can be understood much more readily from the standpoint of the Newtonian emission theory than from the standpoint of the wave theory. It is my opinion therefore that the next phase of the development of theoretical physics will bring us a theory of light that can be interpreted as a kind of fusion of the wave and emission theory' (quoted in Segre 1980, 89).

[17] Jammer (1966, 44) supports this view by underscoring that Planck's constant h was already present in Planck's derivation of Wien's law, that is, in a formula that belongs to pre-quantum physics.

the structure of the atom that confirmed the sense of crisis of classical physics and paved the way for the definitive rise of quantum physics.

3.1.3 Bohr's quantum atom

Niels Bohr got involved in the development of quantum physics when he tackled the problem of the instability of the electron orbit in Rutherford's model of the atom. In 1911 Ernest Rutherford had introduced the 'planetary' or nuclear model in order to account for the observations of large scattering of α-particles—experimental results that J. J. Thomson's plum-cake model could not explain. In Rutherford's model, a tiny electron orbits a massive and positively charged atomic nucleus. The main problem with this proposal was that, according to classic electrodynamics, the electron must lose energy as an effect of its orbital motion, leading to a very fast collapse of the electron into the nucleus—Rutherford's atom could not be stable if classic electrodynamics was right. Besides, there was nothing in Rutherford's model to provide a determined and fixed radius for the orbits in the atom; although it was clear from chemical considerations that atoms of a same element had exactly the same properties.

In 1912-13 Bohr published a paper divided in three parts entitled *On the Constitution of Atoms and Molecules* in which he proposed a solution for this difficulty. He realized that Planck's constant could be used to postulate a constraint in the energy levels that electrons could take in their orbits around the nucleus[18]. He argued that from the constant parameters in Rutherford's model, namely, the charge of the electron e and its mass m, it was impossible to construct a constant with the dimensions of length to determine the orbit. However, with the introduction of h it was possible to construct a quantity of the desired nature, h^2/me^2. He considered an electron of charge e orbiting a nucleus of charge e'. If no energy is radiated, the electron will describe an elliptic orbit with a major axis $2a$ with an orbital frequency ω. If U is the energy required to remove the electron from its orbit to infinity, Bohr showed that, on the basic of classical physics, $\omega = \frac{\sqrt{2}}{\sqrt{m}}\frac{U^{3/2}}{\pi e^2}$ and $2a = \frac{e^2}{U}$ [19]. As just mentioned, combinations of the constant parameters e and m in these formulas cannot yield a constant with the dimension of length for the orbits. Moreover, according to Maxwell's theory, the value of the radiated energy U should increase according to the square of the electron's acceleration, and thus the dimensions of the orbit would continuously decrease until the electron collapses in the nucleus.

As a conceptual way out of this problem Bohr introduced two crucial assumptions: that there is a discrete set of allowed stationary orbits, and that in such orbits electrons do not radiate energy. Armed with these assumptions, plus Planck's hypothesis of energy quantization, he found a successful path[20].

---

[18] In a famous private memorandum directed to Rutherford, Bohr wrote: 'In the investigation of the configuration of the electrons in the atom we immediately meet with the difficulty ... that a ring, if only the strength of the central charge and the number of electrons in the ring are given, can rotate with an infinitely great number of different times of rotation, according to the assumed different radius of the ring; and there seems to be nothing ... to allow, from mechanical considerations to discriminate between the different radii and times of vibration. In the further investigation we shall therefore introduce and make use of a hypothesis, from which we can determine the quantities in question. This hypothesis is: that for any stable ring (any ring occurring in the natural atoms) there will be a definite ratio between the kinetic energy of an electron in the ring and the time of rotation. This hypothesis, for which no attempt at a mechanical foundation will be given (as it seems hopeless), is chosen as the only one which seems to offer a possibility of an explanation of the whole group of experimental results, which gather around and seems to confirm conceptions of the mechanismus of radiation as the one proposed by Planck and Einstein' (quoted in Jammer 1966, 74).

[19] In the special case of a circular orbit of radius a, the equilibrium condition between centripetal force and Coulomb force states that $\frac{mv^2}{a} = \frac{ee\prime}{a^2}$, from which follows that the total energy is given by $U = \frac{mv^2}{2} - \frac{ee'}{a} = \frac{e^2}{2a}$, so that $2a = \frac{e^2}{U}$; and since $v = 2\pi\omega a$—and after some algebra—it follows that $\omega = \frac{\sqrt{2}}{\sqrt{m}}\frac{U^{3/2}}{\pi e^2}$.

[20] In the first paper of his 1913 trilogy on the quantum atom, Bohr wrote: 'Now the essential point of Planck's theory is that the energy radiation from an atomic system does not take place in the continuous way assumed in the ordinary electrodynamics, but that it, on the contrary, takes place in distinctly separated emissions, the amount of energy radiated out from

He considered an electron at great distance form a nucleus with no sensible velocity relative to the latter that is brought to one of the stationary orbits around the nucleus with orbital frequency ω. In the process, Bohr assumed, the electron has emitted radiation U of a frequency ν which is equal to the average frequency between the initial orbital frequency 0 and the final frequency ω—and considering Planck's hypothesis, this means that $U = \frac{nh\omega}{2}$. Inserting this expression in the classical formulas for ω and $2a$ he had already obtained, Bohr got

$$U = \frac{2\pi me^4}{n^2h^2} \qquad \omega = \frac{4\pi me^4}{n^3h^3} \qquad 2a = \frac{n^2h^2}{2\pi^2me^4}.$$

With ($\tau = 1, 2, 3, 4, \ldots$) these equations yield the energy, orbital frequency and radius of each stationary orbit. For $n = 1$, and inserting the values for the constants e, m, and h, Bohr obtained $U = 13$ eV, $\omega = 6{,}2 \times 10^{15}$ sec^{-1}, and $a = 1{,}1 \times 10^{-8}$ cm; all values consistent with experimental data.

Bohr then turned to spectroscopy as a heuristic tool in the formulation of his atomic theory. During the 19th century it was discovered that the radiation spectrum of many elements showed apparent regularities. The electromagnetic spectrum of emission of these elements did not constitute a continuum, but a discrete pattern of 'emission lines' of a certain frequency. In 1885, Joseph Balmer introduced a formula for such regularities in the case of hydrogen, namely, $\nu = R\left(\frac{1}{(n_2)^2} - \frac{1}{(n_1)^2}\right)$, where R is a constant (commonly known as Rydberg's constant), ν is the frequency of the emission line, and n_1 and n_2 are integer numbers such that $n_2 < n_1$. With $n_2 = 2$, the emission lines in the visible part of the spectrum were accounted for—such lines formed Balmer's series. Successful as it was in terms of observational predictions, this formula had not been provided any kind of plausible theoretical foundation up to 1913, but Bohr's theory did offer an explanation. His second assumption—that atoms emit energy only when electrons 'jump' between stationary states—along with Planck's quantum hypothesis entailed that this emission, in the case of an electron jump from energy level 2 to level 1 is given by $U_{n_2} - U_{n_1} = h\nu$. Bohr plugged in the left hand side of this formula the expression for the energy levels he had already obtained to get $U_{n_2} - U_{n_1} = \frac{2\pi^2me^4}{h^2}\left(\frac{1}{(n_2)^2} - \frac{1}{(n_1)^2}\right)$, and therefore

$$\nu = \frac{2\pi^2me^4}{h^3}\left(\frac{1}{(n_2)^2} - \frac{1}{(n_1)^2}\right),$$

which is Balmer's formula. That is, Bohr determined the value of Rydberg's constant from theoretical reasoning and showed that Balmer's formula is simply the expression for the frequency of emitted radiation in electron's quantum jumps—and he so explained the discrete nature of the patterns of emission observed in spectroscopy.

In the third section of the first paper of the trilogy on the atomic model, Bohr provided a more general derivation of Balmer's law. Instead of assuming that a distant electron that is brought to a stationary state emits radiation of energy $U = \frac{nh\omega}{2}$ in the process, he expressed this energy as $U = f_{(n)}h\omega$, where $f_{(n)}$ is an indeterminate function of n. Then he obtained $U = \frac{\pi^2mZ^2e^4}{2h^2f^2}$ and $\nu = \frac{\pi^2mZ^2e^4}{2h^3}\left(\frac{1}{f^2_{(n_2)}} - \frac{1}{f^2_{(n_1)}}\right)$. According to the variable factor in Balmer's formula $\left(\frac{1}{(n_2)^2} - \frac{1}{(n_1)^2}\right)$, the function f must have the form $f_{(n)} = cn$. To determine the factor c, Bohr considered the transition between two successive stationary states N and $(N-1)$

an atomic vibrator of frequency ν in a single emission being equal to $\tau h\nu$, where τ is an entire number, and h is a universal constant' (quoted in Heilbron & Kuhn 1969, 268).

and obtained $\nu = \frac{\pi^2 me^4}{2c^2h^3}\frac{2N-1}{N^2(N-1)^2}$, and for the orbital frequencies before and after the transition he got $\omega_N = \frac{\pi^2 me^4}{2c^3h^3N^3}$ and $\omega_{N-1} = \frac{\pi^2 me^4}{2c^3h^3(N-1)^3}$. Bohr argued that if N is large, the ratio between the orbital frequency before and after the emission should be nearly equal to 1, and the same holds for the ratio between the radiative frequency and the orbital frequency, and the mentioned equations show that this condition is met only if $c = 1/2$—and if $c = 1/2$, Balmer's formula can be obtained. This derivation is historically important because it is the first use of the famous *correspondence principle*, which states that when the dimensions of action involved are large enough for quantum effects to be negligible—in this case the condition that N is large—the quantum formulas must reduce to the corresponding classical physics equations[21].

The link between Bohr's model of the atom and Balmer's spectroscopy formula was also the source for the former's empirical confirmation. Assigning different values to n_2 new series of emission lines were predicted in different regions of the electromagnetic spectrum. All those predicted series were soon observed. For $n_2 = 1$ an emission series in the ultraviolet spectrum was observed by Lyman in 1914, and for $n_2 = 4, 5$ series were observed in the infrared by Brackett in 1922 and by Pfund in 1924, respectively (for $n_2 = 3$ an infrared series had already been predicted by Ritz and observed by Paschen in 1908). The most important spectroscopic empirical confirmation of Bohr's theory came from an empirical result that at first sight looked like an anomaly. In 1896 Pickering had observed an emission series in the spectrum of the star ξ Puppis that he identified as six hydrogen lines that could not be accounted by Balmer's formula, but by the expressions $\nu = R\left(\frac{1}{2^2} - \frac{1}{(m+1/2)^2}\right)$ and $\nu = R\left(\frac{1}{(1/2)^2} - \frac{1}{m^2}\right)$, with $m = 2, 3, \ldots$. However, Bohr noticed that these two formulas could be written as $\nu = 4R\left(\frac{1}{4^2} - \frac{1}{k^2}\right)$—with $k = 5, 7, \ldots$—and $\nu = 4R\left(\frac{1}{3^2} - \frac{1}{k^2}\right)$—with $k = 4, 6, \ldots$—respectively. Bohr then showed that applying his atomic theory to ionized helium with nuclear charge $2e$ these formulas could be obtained. That is, according to Bohr, the Pickering lines corresponded to ionized helium emission, not to hydrogen. If he was right, an equivalent series should be obtained in the laboratory in the emission spectrum of a tube filled with helium and chlorine. Experiments carried by E. J. Evans and A. Fowler in 1914-5 confirmed this prediction.

Two other sources of empirical confirmation came from experiments developed by J. Franck and G. Hertz and by H. G. J. Moseley. In 1919 Franck and Hertz bombarded molecules of a low pressure gas with electrons. If the electrons carried low energy, only elastic collision happened and no emission was observed. However, if the electrons' energy was equal or higher than a specific threshold (4,9 eV for mercury vapor), then collisions and radiation were observed. The detected loss of energy in the bombarding electrons neatly corresponded to the difference between the ground state ($n = 1$) and the excited state ($n = 2$) of the bombarded atom, and the measured frequency of the emitted radiation clearly corresponded to the frequency of radiation associated to a quantum jump from the excited to the ground state—that is, the emission happened because of the electron in the atom excited by the collision returns to the ground energy level. All these results, of course, strongly confirmed Bohr's theory.

Moseley's experiments dealt with X-rays. According to Bohr, this kind of radiation occurred when an electron in the innermost atomic orbit is knocked out and an electron from a higher energy level goes down to fill the empty place. Applying his theory it was possible to determine a fixed relation between the frequency of the X-ray emitted and the atomic number[22]. Moseley measured the radiation spectrum of X-rays emitted by all then-known elements between calcium and zinc. According to Bohr's theory, the

[21] For a detailed account of the relevance of the correspondence principle in the development of the 'old quantum theory' (quantum theory as developed between 1913 and 1924), see (Jammer 1966, section 3.2).

[22] The formula $U = \frac{2\pi^2 me^4}{n^2h^2}$ is a special case for atomic number $Z = 1$ (hydrogen). The form of the general formula is $U = \frac{2\pi^2 mZ^2e^4}{n^2h^2}$, so that $\nu = \frac{2\pi^2 mZ^2e^4}{h^3}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right)$. The mentioned relation between X-ray frequency and atomic number is thus $\sqrt{\nu} \propto Z$.

relation between X-ray frequency and atomic number is $\sqrt{\nu} \propto Z$, so its graphic representation must be a straight line. Since some parts of the line were missing, Moseley's predicted the existence of elements with atomic numbers 42, 43, 72 and 75, corresponding to the absent frequencies. All four elements were eventually discovered, confirming Bohr's theory.

Bohr's theory was, in general, very positively received in the scientific community. For example, Einstein, who by that time was already a leading figure in theoretical physics, described Bohr's work as a great discovery and an enormous achievement. Older important characters such as J. Jeans were also very positive about it. This positive reception, however, happened together with a clear recognition that the quantum atom was at deep and fundamental odds with classical physics. Bohr's two postulates and his open and substantial assumption of Planck's energy quantization hypothesis were absolutely inconsistent with classical mechanics and electrodynamics. A very clear example of this was the notion of a quantum 'jump'. To say that an electron jumps from an excited state to the ground state (and vice-versa) was simply a manner of speaking, for in such jump the electron does not traverse the space between the stationary states involved—the jump we are considering is an intrinsically discontinuous physical event. Features like these were the basis for some dissenting voices regarding Bohr's theory. For example, O. Stern and Max von Laue stated that if the theory turned out to be correct they would quit physics. Anyhow, the empirical and theoretical success of Bohr's theory contributed to clarifying that a scientific revolution was occurring. His 1913 model of the atom strongly reinforced the sense of crisis already brought about by Planck's work.

3.1.4 Sommerfeld's quantum conditions

Though Bohr's atomic model was indeed a great achievement, both from the theoretical and empirical point of view, there were still some important loose ends. Two empirical problems in spectroscopy showed that some refinements and modifications in Bohr's theory were needed: the observation of 'fine structure' in emission lines, and the 'Zeeman effect'. As early as 1892 it was known that some of the hydrogen emission lines are not single lines, but 'doublets' of lines, and the observation of X-rays emission lines drew similar results around 1915—their emission lines were usually doublets. That is, there was a fine structure underlying the emission lines. The Zeeman effect, discovered by Peter Zeeman in 1897, simply consists in that when a magnetic field is applied to an emission source, a single spectral line gets split into a number of lines. Bohr's theory of 1913 did not contain an explanation for any of these features.

The first step towards an explanation was given in a 1915 paper by Arnold Sommerfeld entitled *On the Theory of the Balmer Series*. He noticed that Bohr's work was applicable only to systems of one degree of freedom, and he undertook the task of introducing a generalization of quantum conditions for every degree of freedom required in order to describe the motion of a quantum system—such as an orbiting electron. Using two polar coordinates to describe the orbital motion of an electron orbiting in a plane, Sommerfeld introduced two quantum conditions in the form of a restriction of two phase-integrals connected to the radius and angle components:

$$\oint p_r dr = n'h \qquad \oint p_\psi d\psi = kh,$$

where the integer numbers k and n' establish the quantum conditions for the radius r and the azimuthal angle ψ. From these conditions Sommerfeld deduced that $k/(k+n') = b/a$, where a and b are the major and minor semi-axes of the elliptical orbit, respectively. He also showed that the energy for a stationary state is given by $U = \frac{-RhZ^2}{(k+n')^2}$, and comparing this expression to the one that Bohr obtained it follows that

Bohr's quantum number n is equal to $k + n'$. The 0 value for number k had to be excluded—for it would represent a straight line going through the nucleus, not an orbit—so that the possible values for k and n' were given by $1 \leq k \leq n$ and $0 \leq n' \leq (n-1)$. Taking all these features together it follows that electrons orbit in a circle around the nucleus when $k = n$, and when $k < n$ the orbit is an ellipse—whose degree of eccentricity increases as k decreases.

Sommerfeld quantum conditions gave a more detailed picture of Bohr's model, but given that $n = k + n'$, no further energy levels were introduced, and such extra levels were required for an explanation of the fine structure of emission lines. The solution came when Sommerfeld introduced Einstein's relativity. At the quantum level, the variations in the speed of an electron along an elliptical orbit are enough in order to produce a variation in the electron's mass, which in turn brings along a variation in the electron's energy. Without this mass-energy variation, the electron would orbit in a closed ellipse (if $k < n$), but when the alteration is considered the orbit becomes an open ellipse whose perihelion precesses with an angular velocity dependent on the eccentricity of the ellipse eccentricity which depends on the value of k. Therefore, for a quantum level n, there are n possible quantum orbits, each with a slightly different energy. The formula that Sommerfeld obtained to express these facts is the following:

$$U_{(n,k)} = \frac{-RhZ^2}{n^2}\left[1 + \left(\frac{\alpha^2 Z^2}{n}\right)\left(\frac{1}{k} - \frac{3}{4n}\right) + 0\alpha^4\right]$$

The correction factor in square brackets—which contains the term $\alpha = 2\pi e^2/hc$, commonly known as the fine structure constant—is a function of n and k, and its presence in the formula shows that the precession decreases the energy by a fraction which depends on these quantum numbers. This precession, thus, is associated to a quantum jump between two possible quantum orbits given for a quantum level n. Therefore, Sommerfeld's result naturally explains the fine structure of emission lines.

Paschen's observations on the spectrum of ionized helium mentioned above—which were important in the empirical confirmation of Bohr's theory—were also relevant to provide empirical support to Sommerfeld's modifications. Given the presence of the factor Z^4 in the formula above, fine structure was easier to observe in spectra of elements heavier than hydrogen, such as helium. Paschen's measurements on helium spectral emission corresponded to the results expected from Sommerfeld's formula[23]. X-rays spectroscopy was also a source for confirmation of Sommerfeld's work. As mentioned above, X-rays emission lines turned out to be normally doublets, and the above formula could be applied to explain their fine structure and to refine Moseley's theoretical conclusions.

The first step in the long line of reasoning that led to a solution of the Zeeman effect problem was given by Sommerfeld's generalization of his phase-integrals conditions to the case of three dimensions in polar coordinates. This generalization is given by

$$\oint p_r dr = n'h \qquad \oint p_\varphi d\varphi = n_1 h \qquad \oint p_\theta d\theta = n_2 h\,,$$

where φ and θ give latitude and the azimuthal angle with respect to a polar axis, respectively, and n_1 and n_2 are the equatorial and latitudinal numbers, respectively. According to the formula for the kinetic

[23] A Jammer points out (1966, 95), Paschen's results were also relevant as empirical confirmation for Einstein's formula of the velocity-dependence of mass. Actually, K. Glitscher showed in 1917 that the velocity-dependence predicted by Abraham's model of the electron would have implied a very different fine structure for the helium emission lines. By 1916 it was rather clear that Lorentz's theory—which, as we saw, predicts exactly the same v-dependence of mass as special relativity—was incompatible with the quantum hypothesis. The quantum atom strengthened the view that this hypothesis had arrived to stay, so it was natural that Paschen's 1916 results were not considered as empirical support for Lorentz's theory.

energy he now obtained, and in comparison with $U = \frac{-RhZ^2}{(k+n')^2}$, Sommerfeld noticed that $k = n_1 + n_2$. Now, since the total angular momentum $p_\psi = kh/2\pi$ is normal to the orbital plane and since p_φ is its projection on the polar axis, he concluded that $n_1 = k\cos\alpha$ (where α is the angle between the direction of p_ψ)—or, equivalently, that $\cos\alpha = n_1/(n_1 + n_2)$. This result means that the quantum number n_1 represents a quantization for the inclination of the orbital plane with respect to the polar axis. This feature becomes relevant in connection with the Zeeman effect when the polar axis becomes uniquely determined by means of the direction of a magnetic field. That is, the presence of a magnetic field produces a 'space quantization' affecting the possible inclination angles that an electron orbit can take.

In 1916 P. Debye obtained a similar result that offers a precise account of the Zeeman effect in terms of energy values. From the equations of motion of an electron in a magnetic field, Debye derived a Hamiltonian written in polar coordinates. From this Hamiltonian he derived expressions for n', n_1 and n_2 in terms of energy, and from those expressions he obtained $U = \frac{2\pi^2\mu e^4}{h^3} + \frac{mh\omega}{2\pi}$, where μ stands for the electron's mass and $\omega = eH/2\mu c$ (with H representing the magnetic field and c a constant). The second term in the sum expresses a correction with respect to Bohr's formula that depends on the 'magnetic quantum number' $m = n_1$. Finally, applying $U_{n_2} - U_{n_1} = h\nu$, Debye obtained

$$\nu = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{(\tau_2)^2} - \frac{1}{(\tau_1)^2}\right) + \frac{\omega}{2\pi}(m_2 - m_1).$$

In this case the numbers principal quantum numbers n_2 and n_1 of Bohr's formula have been replaced by τ_2 and τ_1 in order to avoid confusion with the numbers $n_1 + n_2 = k$ that Sommerfeld introduced, and, as usual, in Debye's formula the numbers with sub-indexes 1 and 2 represent the numbers corresponding to the initial and final states of the jumping electron, respectively.

That the quantum number m corresponds to the magnetic moment of the orbiting electron can be explained in the following way. Since $\omega = eH/2\mu c$ and $n_1 = k\cos\alpha$, the second term in Debye's energy formula can be written as $mh\omega/2\pi = (kh/2\pi)(e/2\mu c)H\cos\alpha$. From electrodynamics it can be shown that $(kh/2\pi)(e/2\mu c)$ represents the magnetic moment generated by a circling electron with angular momentum $p_\psi = kh/2\pi$. Then, and in vector notation, it follows that $mh\omega/2\pi = \mathfrak{m}\cdot\mathbf{H}$, that is, that the correction term in Debye's formula corresponds to the quantized energy of the interaction between a magnetic moment $\mathfrak{m}$ and a magnetic field $\mathbf{H}$. And this result offers an explanation of the emission lines splitting of the Zeeman effect: the presence of an external magnetic field splits the normal emission lines into different lines that are described by the extra energy levels that the correction term $mh\omega/2\pi$ in Debye's formula determines.

3.1.5 Pauli's exclusion principle and electron spin

However, the solution for the empirical problem was not complete. Debye's formula worked for the 'normal' Zeeman effect associated with magnetic fields applied to singlet lines, that is, to emission lines without a fine structure. The magnetic splitting of doublets, triplets and multiplets, the 'anomalous' Zeeman effect, could not be accounted for. In the early 1920s, Sommerfeld and Landé formulated a sketch for a theory that was known as the 'magnetic core hypothesis'. This hypothesis postulated that the atomic core, the nucleus and the non-optical inner electrons, possesses a characteristic angular momentum associated to a magnetic moment, which in turn produces a magnetic field whose symmetric axis coincides with the core's angular momentum. That is, Landé and Sommerfeld postulated a sort of internal Zeeman effect within the atom to try to explain the anomalous splitting. In 1919 Landé had introduced vectors **R**

and **K** to represent the angular momentum of the core and of the optical electron, respectively, that combined formed the vector **J** representing the total angular momentum of the atom. When an external magnetic field is acting, **J** oscillates as a gyroscope about the direction of the field. Numbers j, l and s were introduced for the magnitudes of **J**, **K**, and **R,** respectively, with j the 'inner quantum number' for the total angular momentum and with $l = (k - 1)$ [24]—so that for a principal quantum number n, l can assume the values $0, 1, \dots, n - 1$. The basic idea was thus that the anomalous splitting depended on the values of j and l, i.e., the value of the angular momentum of the atomic core was thought to be responsible for the anomalous Zeeman effect.

Landé obtained a formula for the energy of the split lines $U = m_1 ghv$, where g is a ('splitting factor') constant dependent on J, K and R; and m_1 is a new quantum number. Just as the quantum number m introduced by Sommerfeld and Debye represented the projection of the angular momentum of the optical electron on the direction of the external field, Landé's number m_1 represented the projection of the total angular momentum J on the direction of the field. A strange point was that four even multiplicities, the value of m_1 was given by half-integers, whereas for odd multiplicities it corresponded to integers. A possible explanation was grounded on an argument introduced by Heisenberg. If an electron joins an atom, its interaction with the electrons that were already there, the core electrons, was such that the incoming electron gives up a fraction of its angular momentum given by half a unit of $h/2\pi$ that is absorbed by the core, and retaining the rest.

The magnetic core theory gave some clues: the need for yet another quantum number that could take half-integer values, and the notion of an angular momentum other than the one associated to the orbital motion of the optical electron as responsible for the anomalous splitting, for example. However, in 1925, W. Pauli took the magnetic core theory to a relativistic context and demonstrated that its resulting predictions could not be put in agreement with experience, and he concluded that the postulated angular momentum of the core, given by **R** and s, was not relevant for the anomalous Zeeman effect. Accordingly, he attempted a solution based on a different outlook. Moseley's and Kossel's work on X-rays had clarified that they were emitted when an electron in the innermost orbits is knocked out and an electron jumps down to cover the empty place from an outer orbit. The argument implied that there was a maximum number of electrons occupying each of the inner orbits. Bohr elaborated on this idea based on chemical evidence, in order to determine the structure of the periodic system of elements The energy needed to ionize noble gases, with atomic numbers $2, 10, 18, 36, 54, 86$, was comparatively high and that atoms of such elements were rather reluctant to chemically bond with other atoms—hence their chemical stability. In the case of hydrogen and halogens, with atomic numbers $1, 9, 17, 35, 53, 85$, their atoms easily formed compounds and became negative ions. Alkali atoms, finally, were prone to lose an electron and become positive ions. Since the chemical behavior of atoms depends on their outer 'valence' electrons, this information suggested that noble gases had their outer orbital shells full, that in hydrogen and halogens only one empty place was left in the outermost shell, and that in alkalis the outer shell is inhabited by one single electron.

When he knew of Bohr's work on the structure of the periodic system of the elements, Pauli got intrigued by one question that, as Bohr admitted, the theory was not capable to answer, namely, why not all electrons of an atom occupy its innermost shell? The clue that Pauli followed to provide an answer was given by the 1924 work of E. C. Stoner. Stoner realized that the amount of electrons in a completed energy level n equals twice the total possible assignments of the three quantum numbers n, k and j (in Stoner's original notation, n, k_1, k_2), so that the total number of electrons for a completed level is given

[24] In modern terminology, the magnetic quantum number m_l has $2l + 1$ values ranging between $-l$ and $+l$. The positive values defining an angular momentum parallel to the direction of the magnetic field, the negative values defining antiparallel orbits, and the 0 value the perpendicular case.

by $2n^2$ [25]. Stoner's scheme was consistent with chemical and X-ray spectroscopic empirical data, and it allowed an accurate description of the structure of atomic shells—comparison with the atomic numbers of noble gases shows that successive shells have a maximum capacity of 2, 8, 8, 18, 18, 32 electrons. A crucial point implied in Stoner's work was that for alkali atoms, when an external magnetic field is applied, the splitting of the emission lines shows that the lonely atom in the outermost shell can adopt an amount of different energy levels corresponding to the number of possible electrons that can populate the outermost completed shell of the next noble gas in the periodic system.

This feature was the key for Pauli's formulation of the exclusion principle in 1925. That the optical electron could take energy values given by the amount of 'electron-seats' available in the outermost shell suggested Pauli that a sort of prohibition rule governed atomic structure. Describing the configuration of each electron by means of the quantum numbers n, l, j and m_j (in Pauli's original notation, which followed Stoner's, n, k_1, k_2, m_1), the exclusion principle states that only one electron in an atom can occupy each state defined by those numbers. An empirical test was readily available. The magnetic splitting of emission lines of alkaline-earth atoms, with two external electrons, was such that some lines were missing (the triplets corresponding to $k = 1$), but Pauli's work showed that they were banned, because for the lines to be present, the two outer electrons had to be described by the same set of quantum numbers. Another important feature in Pauli's work was that, as the theoretical sketch of Landé had already suggested, a factor of $1/2$ was necessary to account for the total angular momentum and magnetic moment of the atom. More precisely, Pauli was led to the introduction of a magnetic moment and an angular momentum that could not be accounted for by means of the properties of the orbital motion of the optical electron—and, as he had already shown, that could not be explained by the atomic core either—with an intrinsic two-valuedness impossible to explain in terms of classical physics. In other words, Pauli showed that the total angular momentum was given by $j = l + s$, where $s = \pm 1/2$, in units of $\hbar = h/2\pi$, represents the 'extra' angular momentum[26].

The half-integer two-valuedness of the 'extra' angular momentum provided the missing energy levels required to explain the anomalous Zeeman effect. The exclusion principle explained why some expected emission lines were not actually observed, and provided a principle for atomic structure that explained why not all the electrons of an atom orbit in the ground level shell. However, Pauli did not propose a definite physical underpinning for the two-valued extra angular momentum. The first attempt to fill this theoretical gap was made by R. Kronig also in 1925. He proposed that the difference between the total angular momentum j and the orbital angular momentum l was due to an intrinsic angular momentum of the electron, momentum that he interpreted as based on a spinning motion of the electron about its own axis. He quickly found some problem with the hypothesis, though. Sommerfeld's spin-free formula provided a very accurate account of the spectrum of hydrogen, but the introduction of spin would destroy the agreement with experience unless it was possible to interpret Sommerfeld's formula as expressing a compensation effect between the spin-orbit coupling and the relativistic energy variation due to the electron's orbit precession; but Kronig was not able to introduce spin in Sommerfeld's formula in a way such that the suitable compensation was attained. On the other hand, it became readily clear that the hypothesis required that for a spinning electron with intrinsic angular momentum $\hbar/2$ the velocity of a point in its surface would be many times greater than the speed of light, in blatant conflict with special

[25] In modern notation, Stoner's view is more clearly explained with the quantum numbers n (principal), l (azimuthal) and m_l (magnetic). Recalling that $0 < k \leq n$, that $l = k - 1$ and that $|m_l| \leq l$, then for a level given by $n = 1$ there is one single possible assignment $(1, 0, 0)$ and it gets completed with 2 electrons, for the level $n = 2$ there are four assignments $(2, 1, -1)$, $(2, 1, 0)$, $(2, 1, 1)$, $(2, 0, 0)$ and it gets completed with 8 electrons. The same reasoning shows that for levels $n = 3, 4$ the total of electrons when completed is 18 and 32, respectively.

[26] In modern terms, Pauli's exclusion principle means that only one electron can take each state defined by the numbers (n, l, m_l, m_s), where m_s is the spin projection number. The factor 2 in Stoner's expression $2n^2$ for the total amount of electrons corresponding to a level n, gets readily explained by the intrinsic two-valuedness $\pm 1/2$ of m_s.

relativity. Discouraged by these problems and the severe criticism of Pauli, Kronig decided not to publish his ideas.

Almost simultaneously, the Dutch physicists G. E. Uhlenbeck and S. Goudsmit proposed an equivalent hypothesis. The experimental data related to the Zeeman effect suggested that the three degrees of freedom for an orbiting electron given by the quantum numbers, n, k and m_l were not enough to account for all possible energy levels, and that the introduction of something like a fourth degree of freedom was necessary. Uhlenbeck and Goudsmith found the key for the extra degree of freedom in Pauli's work. The idea of a point-electron orbiting an atom admitted only three degrees of freedom, but after reading Pauli's paper, where four quantum numbers were assigned, they concluded that picturing the electron as small rotating sphere the idea of a fourth degree of freedom associated to spin was rather natural. That is, both in Kronig's and Uhlenbeck and Goudsmith's proposal, the vector **R** in Landé's model was associated to the intrinsic angular momentum of the electron rather than to the core. Unlike Kronig, the Dutch physicists received encouraging support from Eherenfest and Lorentz and, in spite of the faster than light rotation problem, they published their hypothesis. Bohr's approval, and the solution of the problem that bothered Kronig in connection to Sommerfeld's formula provided by L. H. Thomas and J. Frenkel in 1926, paved the way for a general acceptance of the spin hypothesis.

The introduction of spin closes a period that is commonly known as 'the old quantum theory'. As it can be seen from what has been said so far, this old quantum theory is a sort of patchwork of independent hypothesis concerning several forms of discrete quantization involving h, not a unified theoretical corpus. Furthermore, the extensive use of Bohr's correspondence principle was somewhat problematic as well, for the introduction of suitable quantum conditions that in the classical limit delivered the empirically correct results was a rather *ad-hoc* maneuver. This state of things gradually led to the awareness that a quantum theory with tight and strong foundations was necessary. The formulation of such theory, QM, was the result of the work of Heisenberg, Jordan and Born, on the one hand—matrix mechanics—, and of the work of de Broglie and Schrödinger, on the other hand—wave mechanics.

3.1.6 Matrix mechanics

Bohr's method was based on a consideration of motion of atomic systems in classical terms, to which the quantum conditions were applied, and guided by the correspondence principle. Heisenberg abandoned the classical conception of motion and replaced it by a description given by what he considered as *observable* magnitudes. He adopted an epistemological stance rather close to logical empiricism—a school of thought that was flourishing by those years. Heisenberg rejected the classical kinematic notions of position, velocity, momentum and trajectory of a quantum particle simply because they were not open to direct observation[27]. Instead, he focused on the optical observable quantities of frequency and intensity. On the other hand, Heisenberg did use the correspondence principle, but in a much more fundamental way than the creators of the old theory. Instead of using it as a guessing principle to solve a diversity of problems—the principle had to be adapted to every single case—Heisenberg used it as a guide to formulate the very mathematical apparatus of the new mechanics in his 1925 epoch-making paper *On a Quantum Theoretical Interpretation of Kinematical and Mechanical Relations*.

Heisenberg started from the fact that in classical physics a time-dependent quantity $\xi_n(t)$ can be represented by a Fourier series $\xi_n = \sum_\tau \xi(n,\tau) = \sum_\tau x(n,\tau)e^{2\pi i\nu(n,\tau)t}$. The quantum conditions implied that the classical amplitude $x(n,\tau)$ and classical frequency $\nu(n,\tau)$ correspond to the quantum amplitude $x(n,n-\tau)$ and quantum frequency $\nu(n,n-\tau)$, respectively. Though the expression $\sum_\tau x(n,n-\tau)e^{2\pi i\nu(n,n-\tau)t}$ does not have a quantum counterpart, Heisenberg assumed a correspondence between the

[27] For an interesting challenge to the generally accepted view concerning the positivist spirit of Heisenberg's work, see (Wolff 2014).

set of all the terms in the classical Fourier series and the set of all the terms in the quantum Fourier series, so that $\xi_n \leftrightarrow \{x(n, n-\tau)e^{2\pi i\nu(n,n-\tau)t}\}$. His next step was to look for the form of $(\xi_n)^2$ in the quantum case. His sets-correspondence assumption meant that $(\xi_n)^2 \leftrightarrow \{[x(n, n-\tau)]^2 e^{2\pi i\nu(n,n-\tau)t}\}$, so the question of the quantum form of $(\xi_n)^2$ reduced to the quantum form of $[x(n, n-\tau)]^2$. Heisenberg found that the expression for this squared quantum amplitude was given by the multiplication rule

$$[x(n, n-\tau)]^2 = \sum_{\tau'} x(n, n-\tau')\, x(n-\tau', n-\tau)$$

Generalizing this rule for two different quantities ξ_n and η_n, with $\eta_n \leftrightarrow \{y(n, n-\tau)e^{2\pi i\nu(n,n-\tau)t}\}$, Heisenberg noticed that, in general, $\xi_n\eta_n \neq \eta_n\xi_n$—for, in general, $\sum_{\tau'} x(n, n-\tau')y(n-\tau', n-\tau) \neq \sum_{\tau'} y(n, n-\tau')x(n-\tau', n-\tau)$. That is, the quantum multiplication turned out to be non-commutative. The multiplication rule provided Heisenberg with a unified kinematics quantum formalism. He completed his work with the dynamics given by the equation for energy of a harmonic oscillator

$$U = \hbar\omega(n + 1/2)$$

where ω is the angular velocity $2\pi\nu$. He obtained this equation—which contrasted with the result of the older quantum theory $U = h\nu n = \hbar\omega n$—guided by his fundamental interpretation of the correspondence principle.

The non-commutative multiplication rule in Heisenberg's work was rather puzzling, until Max Born realized that it was simply the 'row times column' multiplication rule of matrices. That is, Heisenberg's sets of time-dependent observable quantities were matrices whose multiplication was governed by the row time columns rule, rule that determines a non-commutative operation. What we nowadays call 'linear algebra' was not a widely known branch of mathematics for physicists of the time. P. Jordan was an exception. He was highly proficient in matrix mathematics, so Born recruited him as a collaborator. Together they elaborated on Heisenberg's work and obtained the important matrix equation

$$\mathbf{pq} - \mathbf{qp} = \left(\frac{h}{2\pi i}\right)\mathbf{1}$$

where $\mathbf{p}$ and $\mathbf{q}$ stand for the momentum and position matrices, respectively, $\mathbf{pq} - \mathbf{qp}$ is a diagonal matrix, and $\mathbf{1}$ stands for the unit matrix. This simply means that the matrices defining momentum and position are non-commuting. Heisenberg, Born and Jordan systematized and generalized the matrix approach to QM in a famous subsequent 1925 paper commonly known as the 'three men paper'. This paper was the first comprehensive exposition of the foundations of QM.

At this point a flash-forward digression is conceptually useful. In 1927 Heisenberg tried to provide some intuitive content to this strange property of non-commutability. Elaborating on the relation $\mathbf{pq} - \mathbf{qp} = (\hbar/i)\mathbf{1}$ he found out that the product between the uncertainty in our knowledge of two non-commuting physical quantities, such as momentum and position, is always greater than a value proportional to Planck's constant h—the precise expression being

$$\Delta q \Delta p \geq \frac{\hbar}{2}$$

This means that we cannot simultaneously know with exact precision the values of non-commuting physical quantities corresponding to a physical system, in that case the product of the uncertainties should be

zero. For this reason, the mentioned formula is commonly known as the *uncertainty principle* (Heisenberg also found out that an equivalent relation holds for the case of energy and time).

Heisenberg proposed that the physical underpinning of this relation is given by the unavoidable disturbance in the system that observation entails. He gave the example of the observation of an electron with a γ-ray microscope. He reasoned that in order to find out the position of the electron, radiation of a certain wavelength should be pointed at it. In order to increase the accuracy in the position measurement, the frequency of the radiation should also increase. However, Heisenberg thought, the effect of a high frequency radiation ray on an electron would produce a strong Compton effect, such that the momentum of the electron would be altered in an unpredictable way, increasing the uncertainty of our knowledge of it. This early interpretation was physically and epistemologically criticized by Bohr. He pointed out that Heisenberg did not notice that the uncertainty in the momentum of the electron was related to the width of the microscope lens. The momentum of the recoiling electron could be measured after illuminating the electron. However, for this to be possible it was necessary that the angular aperture of the lens was significant, and in that case the sensitivity and precision of the microscope to position observations would decrease. On the epistemological side, Bohr did not think that a sort of unsurmountable technical limit in measurement processes was the root of the uncertainty relations, he was convinced that the real underpinning was rooted in the wave-particle duality.

Coming back to 1925, we have that given the instrumentalist epistemological considerations underlying Heisenberg's endeavor, matrix mechanics was a theory for which no clear physical meaning could be assigned. The matrices describing observable quantities and the rule to operate with them did not depict any intuitive model to provide an understanding of the physical processes involved. Besides, the fact that the mathematics of matrices were not widely known at the time did not contribute to an open-arms reception of the three men work by the scientific community. It was clearly a very important contribution and a clear advance regarding the foundations of the theory, but there were still some gaps concerning the understanding of the theory. However, the matrix approach to QM was not the only one available at the time.

3.1.7 Wave Mechanics

The formulation of the wave approach to QM originates in the work of Louis de Broglie in 1923, *Ondes et Quanta*. De Broglie pondered about the strange situation concerning the physical description of light. According to classical physics light is a wave, exhibiting effects such as interference and diffraction. Einstein's work on quantum physics, on the other hand, showed that, at least in some contexts, light gets better described as a particle. De Broglie thus attempted the formulation of a theory in which both the particle and the wave nature of light could be accounted for in a unified way. In his early works in physics he showed that if blackbody radiation is considered as a gas consisting of light-quanta, the application of classical physics to this model inevitably leads to Wien's law; and he also made a first attempt to reconcile Einstein photon hypothesis with interference and diffraction phenomena. However, his breakthrough came only when he included some features of special relativity in the picture.

As we saw in the previous chapter, in Einstein's theory the inertial mass of a particle is given by $m = m_0\gamma$, where m_0 is the rest mass and $\gamma = 1/\sqrt{1 - v^2/c^2}$. Another feature of the theory is the mass-energy relation, that can be written as $U = m_0c^2\gamma$. With these formulas as starting point, de Broglie's main result can be explained in the following way. The energy-mass relation can be squared to give $U^2\left(1 - \frac{v^2}{c^2}\right) = m_0^2c^4$. The classical expression for momentum is $p = mv$. If we replace m with U/c^2 in the momentum expression and then solve for v we get $v = pc^2/U$. Plugging this expression in the formula for the squared energy it readily follows that $U^2 = p^2c^2 + m_0^2c^4$. Since photons move with speed c and have thus a zero rest mass, the second terms in the sum vanishes. Therefore, the energy formula obtained reduces to $U =$

pc. We then equate de Broglie's energy formula with Planck's, that is, $U = h\nu = pc$. Now, since the frequency of light (a photon) is given by $\nu = c/\lambda$, where λ is the wavelength, de Broglie's famous formula that connects momentum and wavelength follows, namely

$$\lambda = \frac{h}{p}$$

Though the fact that this formula has λ, a wave property, on one side and p, a particle property, on the other was received as rather weird, it accomplishes the goal of providing a formally unified foundation for the dual nature of light. The particle nature is given by the momentum p in the formula, whereas the wavelike behavior is accounted for by the wavelength λ. On the other hand, a similar dual nature can be postulated for physical objects that were normally considered as particles, such as electrons. To a particle with a momentum p there corresponds a wavelength λ determined by de Broglie's formula. He applied this reasoning to the case of an electron in orbital motion, and argued that in order to establish a system of stationary orbits it is required that the orbits must contain an integral number of wavelengths. That is, the orbit of an electron is described by a standing wave around the nucleus. De Broglie also demonstrated that this view neatly corresponds to Sommerfeld's quantum conditions.

Empirical confirmation of de Broglie's hypothesis came from experiments that measured diffraction effects—normally associated with waves—in the case of electrons. Actually, in 1921 C. J. Davisson and C. Ramsauer had already performed experiments with electrons—designed to measure scattering-collision effects—whose results were interpreted in 1923 by J. Franck and W. Elsasser as particle-diffraction in accordance with de Broglie's formula. Between 1925 and 1927 Davisson performed further experiments involving electrons going through a nickel crystal that neatly showed the diffraction for particles predicted from $\lambda = h/p$. Several similar experiments were performed in the upcoming years. One of them was carried by G. P. Thomson, son of J. J. Thomson. The latter received the Nobel prize for discovering the electron through its particle properties, his son received the Nobel prize for unraveling its wave-like properties.

The precise way in which de Broglie understood his own work was not that quantum entities were at the same time particles and waves, nor that in some contexts they behaved as particles and in some others as waves. De Broglie thought that quantum particles were 'surrounded' by a guiding or pilot wave determining their path and which was responsible for their wavelike properties[28]—as we will see below, Bohm's quantum theory is similar in this respect. De Broglie elaborated on this view in 1927, and presented his work as 'the theory of double solution', according to which the equations of wave mechanics (Schrödinger's equations) admit a continuous solution given by a wave function with statistical significance (the pilot wave), and a singularity solution that constitute the particles in the pilot waves. Several conceptual objections were leveled against this approach, and it was not well received by the scientific community of the time.

De Broglie's work readily suggested a deep question. If his momentum-wavelength formula assigned wave properties to particles, then there should be a wave equation as well. The formulation of this for-

[28] For example, this is how he conceived the Young double-slit experiment: 'some atoms of light pass through the holes and diffract along the ray of neighboring part of their phase [guiding] waves. In the space behind the wall, their capacity of photoelectric action will vary from point to point according to the interference state of the two phase waves which have crossed the two holes. We shall then see interference fringes, however small may be the number of diffracted quanta, however feeble may be the incident light intensity. The light quanta do cross all the dark and bright fringes; only their ability to act on matter is constantly changing. This kind of explanation, which seems to remove at the same time the objections against light quanta and against the energy propagation through dark fringes, may be generalized for all interference and diffraction phenomena' (from de Broglie's 1924 *A Tentative Theory of Light Quanta*, quoted in Jammer 1966, 246).

mula, that would become the essential equation of QM, was the outcome of the work of Erwin Schrödinger in 1926, *Quantization as an Eigenvalue Problem*. Startled by the work of de Broglie, he undertook the task of finding the wave equation that governs quantum particles—strange as this task might have sounded from the point of view of classical mechanics.

A simple explanation of Schrödinger's reasoning can be given in the following way. Consider the classical wave equation $\frac{\partial^2\Psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\Psi}{\partial t^2}$ (we assume that the motion occurs along the x-axis) and its solutions given by wave functions of the form $\Psi = Ae^{i(kx-\omega t)}$, where A stands for the wave amplitude, $k = 2\pi/\lambda$ is the wave vector, and $\omega = 2\pi\nu$ is the angular frequency. Considering only the time independent component of the equation and taking the second partial derivative with respect to x, we have that $\frac{\partial^2\Psi}{\partial x^2} = -k^2\Psi = -\frac{4\pi^2}{\lambda^2}\Psi$. Now we can introduce the de Broglie's quantum formula $\lambda = h/p$ to obtain $\frac{\partial^2\Psi}{\partial x^2} = -\frac{4\pi^2p^2}{h^2}\Psi = -\frac{p^2}{\hbar^2}\Psi$. The total energy of the system is given by $U = \frac{1}{2}mv^2 + V = \frac{p^2}{2m} + V$, where the terms in the sum represent the kinetic and potential energy, respectively. From this we can get $p^2 = 2m(U - V)$. Plugging the right hand side of this last formula in $\frac{\partial^2\Psi}{\partial x^2} = -\frac{p^2}{\hbar^2}\Psi$ we finally obtain the time independent version of Schrödinger's equation, namely,

$$\frac{\partial^2\Psi}{\partial x^2} = -\frac{2m}{\hbar^2}(U - V)\Psi$$

The solution of this equation has the form of an eigenvalue problem: $-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi = U\Psi$. Applying the formula to the case of the harmonic oscillator Schrödinger obtained the expression for energy $U = \hbar\omega(n + \frac{1}{2})$, that is, the same result Heisenberg arrived to in the matrix approach. Just as in matrix mechanics, and unlike the old quantum theory, Schrödinger's approach allowed the derivation of energy levels in the atom from a unified mathematical framework.

In the case of the time dependent component of the classical mechanics wave equation, if we take the partial derivative of the wave function with respect to t we have that $\frac{\partial\Psi}{\partial t} = -i\omega\Psi = -2\pi\nu i\Psi$. The Planck-Einstein energy formula implies that $\nu = U/h$, so that $\frac{\partial\Psi}{\partial t} = -\frac{2\pi Ui}{h}\Psi = -\frac{Ui}{\hbar}\Psi$. Multiplying both sides of this last equation by $i\hbar$ we get $i\hbar\frac{\partial\Psi}{\partial t} = U\Psi$. Finally, comparing this result with the time independent equation in its energy eigenvalue form, we obtain the time dependent Schrödinger equation:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi = i\hbar\frac{\partial\Psi}{\partial t}$$

If the expression $-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V$ is defined as an operator to be applied on a certain wave function Ψ, the *Hamiltonian operator* H, the equation can be succinctly written as $i\hbar\dot{\Psi} = H\Psi$, where $\dot{\Psi} = \partial\Psi/\partial t$.

Besides the derivation of energy levels, the quantum numbers that had been introduced in a patchwork way in the old theory, neatly followed from Schrödinger's equation (except spin numbers, which were later introduced in the formalism through Pauli's spin matrices). The only requirement was to consider some restrictions for the form of the wave function: it had to be single-valued (only one value for a given set of coordinates), finite, and continuous (for the wave equation is a second order differential equation).

Unlike Heisenberg's matrix mechanics, Schrödinger wave mechanics, at least at first sight, allowed an intuitive visualization of the physical processes it describes. In that sense, it was received as a comforting return to quasi-classical concepts. Heisenberg's approach was a sort of algorithm or black box in which observable initial conditions were plugged and observable quantities came out as the output, without

any reference to a physical underpinning. Schrödinger's description in terms of a wave equation, it was thought ca. 1927, opened a window into the black box. For these reasons, the reception of wave mechanics was warmer than the reception of Heisenberg's theory. This situation turned into a harsh competition. Schrödinger was quite outspoken about his disregard of matrix mechanics, and Heisenberg was also very loud about his distaste for wave mechanics[29]. Anyhow, the fact that the empirical predictions of both approaches were the same suggested that they were simply two alternative formulations of the same theory. Later in 1926 Schrödinger himself demonstrated a formal, mathematical identity between the theories, allowing a more peaceful coexistence between them[30].

The visualization that wave mechanics seemed to allow became quickly problematic. First, the wave functions that are possible solutions of the wave equations could be complex-valued, and the meaning of such a complex wave function was far from clear. Second, and this is a point that Schrödinger himself pointed out in his seminal work, many-particles systems were described by wave functions that yielded a number of dimensions greater than 3, the space in which the 'waves' inhabit is configuration space, not physical space. Finally, Schrödinger flirted with the idea that waves were the ultimate constituent of matter and that particles were nothing but wave packets—a wave with a large amplitude situated in a specific place. However, as Lorentz quickly pointed out, dispersion effects precluded that particles (wave packets) could show a stable behavior—after a short time the wave packet would spread in a large region and there would be no particle anymore. For all these reasons, the intuitive picture that wave mechanics provided had to be understood only as a useful analogy. The precise meaning of the theory remained obscure.

3.1.8 Quantum probabilities

Max Born was one of the leading physicists who did not accept the wave interpretation of Schrödinger's work. Experiments involving electron scattering performed by J. Franck in his home-town Göttingen convinced him that such entities were indeed corpuscles in a literal way—this did not mean a negation of their wave-like properties, of course. Using Schrödinger's formalism in order to explain the result of scattering experiments, Born arrived at a new interpretation of wave mechanics that would become a central feature of QM. In a nutshell, his reasoning was that the wave function describing an electron that collides with an atom gives a measure of the probability of finding it in a certain region of the detection screen. More generally, Born argued that the expression $|\Psi|^2$ (or, more precisely, $\Psi\Psi^*$, given the possibility that Ψ may be complex-valued) represents the measurable probability for particles in specific states.

---

[29] Jammer vividly describes this episode: 'it is instructive to compare Schrödinger's wave mechanics with Heisenberg's matrix mechanics. It is hard to find in the history of physics two theories designed to cover the same range of experience, which differ more radically than these two. Heisenberg's was a mathematical calculus, involving noncommutative quantities and computation rules, rarely encountered before, which defied any pictorial interpretation; it was an *algebraic* approach which, proceeding from observed discreteness of spectral lines, emphasized the element of *discontinuity*; in spite of its renunciation of classical description in space and time it was ultimately a theory whose basic conception was the *corpuscle*. Schrödinger's, in contrast was based on the familiar apparatus of differential equations, akin to the classical mechanics of fluids and suggestive of an easily visualizable representation; it was an *analytical* approach which, proceeding from a generalization of the classical laws of motion, stressed the element of *continuity*; and, as its name indicates, it was a theory whose basic conception was the *wave*. Arguing that the use of multidimensional (> 3) configuration spaces and the computation of the wave velocity from the mutual potential energy of particles is "a loan from the conceptions of corpuscular theory", Heisenberg criticized Schrödinger's approach as "not leading to a consistent wave theory in de Broglie's sense". In a letter to Pauli he even wrote: "The more I ponder about the physical part of Schrödinger's theory, the more disgusting it appears to me". Schrödinger was not less outspoken about Heisenberg's theory when he said: "… I was discouraged, if not repelled, by what appeared to me a rather difficult method of transcendental algebra, defying any visualization" (Jammer 1966, 271-2).

[30] F. A. Muller (1997) has shown that matrix and wave mechanics were not strictly equivalent before von Neumann's introduction of the projection postulate in 1932.

He also realized that a wave function Ψ can be expressed in terms of a complete orthonormal set of eigenfunctions Ψ_n determined by the Schrödinger equation; that is, Ψ can be written as $\sum_n c_n \Psi_n$, where the terms c_n are the coefficients corresponding to the Ψ_n components of the orthonormal basis. Guided by the relation $\int |\Psi(q)|^2 dq = \sum_n |c_n|^2$, and assuming that $\Psi(q)$ is a normalized function describing a single particle so that $\sum_n |c_n|^2 = 1$, he noticed that the integral $\int |\Psi(q)|^2 dq$ can be regarded as the number of particles and that $|c_n|^2$ can be considered as the statistical frequency of the occurrence of the state characterized by the eigenfunction in the orthonormal basis determined by the index n. Born also defined the expectation value in his statistical approach—the weighted sum of the probabilities corresponding to each eigenfunction Ψ_n. In case of the energy U, the expected value is given by $U = \sum_n |c_n|^2 U_n$, where U_n is the energy eigenvalue corresponding to the eigenfunction in the orthonormal basis determined by the index n. Born's statistical formulas were soon strongly confirmed in the field where its application was natural, atomic scattering. Tests carried by Faxén and Holstmark and by Bethe and Mott yielded results in neat correspondence with Born's predictions. Moreover, Wentzel was able to derive Rutherford's classical scattering formula, a formula that twenty years earlier had been so important for unraveling the structure of the atom.

The conceptual cogency and empirical success of Born's statistical approach undermined the hope of a return to a semi-classical conceptual framework through wave mechanics. Wave mechanics did not really answer the question 'what is the state of a particle after a collision?', rather, it answered the question 'what is the possibility of a definite state after the collision?' As Born himself put it in his seminal paper, 'the motion of particles conforms to the laws of probability, but the probability itself is propagated in accordance with the law of causality' (quoted in Jammer 1966, 285). Moreover, it became quickly clear that the probability involved in Born's approach was not like the probability of classical statistical mechanics, in which it simply represents a measure of our ignorance of the specific state of microsystems. For instance, if a system represented by a wave function Ψ_1 with a probability density $P_1 = |\Psi_1|^2$ is superposed with another system given by a function Ψ_2 with a probability density $P_2 = |\Psi_2|^2$, then the probability density corresponding to the superposed system $\Psi_1 + \Psi_2$ is not $P_1 + P_2$, as classical statistics would have it, but $P_1 + P_2 + \Psi_1 \Psi_2^* + \Psi_2 \Psi_1^*$, where the last two terms corresponding to 'interference features' can be different from zero—we will see below how this gets clearly manifested in the double-slit experiment. The wave-particle duality was at the basis of the non-classical nature of Born's probability. The first conclusion that theorists drew was that these probabilities were somehow objective, that is, they do not represent a measure of our ignorance, but a feature of reality and the physical systems themselves[31].

Born's statistical approach is the last conceptual step in the process of creation of QM. At that point it was blatantly clear that the theory was of a very awkward nature. The highlights of this strangeness were given by the wave-particle duality, non-commuting physical quantities and the uncertainty relations, and the objective nature of the probabilities defined by the wave functions. The perplexity that the new theory caused among the physicists of the time brought along a big deal of philosophical reflection regarding what is the most appropriate interpretation of the theory. The question was: what does this theory mean? Before undertaking an outlook of the different interpretations proposed, it is necessary to take a look at the main features of the standard formal framework in which QM is presented. Although the creation of the theory was crowned with Born's statistical approach, the formulation of the theory was further refined by Paul Dirac and John von Neumann. The work of the latter, the presentation of QM in

[31] 'Laws of nature, as Born and Heisenberg contended from now on, determined not the occurrence of an event, but the probability of the occurrence. For Heisenberg, as he later explained, such probability waves are "a quantitative formulation of the concept of *dýnamis*, possibility, or in the later Latin version, *potentia*, in Aristotle's philosophy. The concepts that events are not determined in a peremptory manner, but that the possibility or 'tendency' for an event to take place has a kind of reality, halfway between the massive reality of matter and the intellectual reality of the idea or image—this concept plays a decisive role in Aristotle's philosophy. In modern quantum theory this concept takes on a new form; it is formulated quantitatively as probability and subjected to mathematically expressible laws of nature"' (Jammer 1966, 286-7).

vector spaces, is the most common formalism used to deal with QM and it is specially clear in order to illustrate its most important conceptual features and problems. This formalism is what we can form now on dub as 'standard quantum mechanics' (SQM).

3.2 STANDARD QUANTUM MECHANICS

Pondering on the mathematical structure and relations between matrix mechanics and wave mechanics, the German mathematician John von Neumann wrote a paper in 1926 in which he established the basIs of his monumental work of 1932 entitled *Mathematical Foundations of Quantum Theory*. In this book von Neumann presented the new physical theory within the framework of Hilbert spaces[32].

In order to provide a preliminary and intuitive picture of how QM fits in this framework we can consider the following simple case. Consider a die that is to be thrown into a black box, such that once the dice is in the box we cannot see what face is up[33]. We can use a vector space of six dimensions to represent the probabilities of the 'face-states' within the box once it has been released. To each dimension of the vector space a face of the die $(1, 2, 3, 4, 5, 6)$ corresponds, and a basis vector corresponds to each of the six possible face-states—the basis vectors are orthogonal (perpendicular to each other). Obviously, these basis vectors represent the degrees of freedom for the 'face-state' of the die, so we can label them $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$ and $\mathbf{x}_6$ (the sub-indexes establish which face-state each basis vector corresponds to, of course). Considering the physical properties of the die and the initial conditions in which it is thrown, we can define a vector $\mathbf{v}$ in our vector space in order to determine the probability of each face-state once the die is in the box. For this strategy to make sense, we need the vector $\mathbf{v}$ to be a unit vector (its modulus must be equal to one). We can describe this vector, in terms of its coordinates, as $\mathbf{v} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 + c_4\mathbf{x}_4 + c_5\mathbf{x}_5 + c_6\mathbf{x}_6 = \sum_{i=1}^{6} c_i\mathbf{x}_i$. That our vector is a unit vector means that $\sum_{i=1}^{6} c_i^2 = 1$ [34]. The vector $\mathbf{v}$ gives us the information about the probability of each face-state, in the sense that c_1^2 represents the probability for the die to fall with the face 1 up, c_2^2 represents the probability for the die to fall with the face 2 up, and so on. The probability of each case is equal to or less than one, and the sum of all probabilities is one—this is the reason for the requirement that the vector $\mathbf{v}$ must be a unit vector.

Now we can introduce some of the technical vocabulary of vector spaces. Assume that we are interested in the probabilities of different dice that are thrown inside the box under exactly the same initial conditions (by a robot that always throws them with the same force and from the same position and orientation, for example). The coordinate-coefficients of the probability vector that describes each die depends of course on the initial conditions under which it is thrown. So, if we want the dice to be interesting for our research, first we have to 'adapt' the vector to the initial conditions we have stipulated. We do so by means of a suitable *operator*, that is, a suitable mathematical receipt that converts the vector of any die in a certain face-state in the vector that describes the face-states probabilities corresponding to the initial conditions that interest us. That is, the operator 0 allows us to represent the probabilities at stake when the die is in a ready-to-be-thrown state. Within the vector space in which this operator operates, there exist some vectors that the operator only stretches or shrinks, leaving their orientation unaltered. More precisely, if we consider the operator 0, there are some vectors $\mathbf{r}$ such that $0\mathbf{r}_n = \lambda_n\mathbf{r}_n$, where

[32] *Hilbert spaces* are a special type of vector spaces in which an infinite number of dimensions can be handled. In what is coming we can harmlessly simplify the language and simply refer to *vector spaces*.

[33] I thank Alfred van Herk for this analogical explanation of the way in which vectors in Hilbert space represent quantum states.

[34] For a perfect or 'fair' dice, all the c_i have the same value, namely, $1/\sqrt{6}$, so that the probability of every side to stand up is also the same: $1/6$, but for a biased dice the coefficients and probabilities will be different. Anyhow, even for a fair dice, the specific initial conditions of throwing might imply that the coefficients and probabilities are different anyway.

λ is a real number that describes the shrinking or stretching: if $\lambda = 2$, the vector has been stretched to twice its original modulus, if $\lambda = 1/2$ the vector has been shrunk to half its original size, and if λ is a negative number the direction of the original vector has been inverted. The vectors $\mathbf{r}_n$ which are unaltered in their orientation by the operator 0 are the *eigenvectors* of 0, and the values of λ_n are the *eigenvalues* of the eigenvectors. Any vector in the space in which 0 operates can be described in terms of its coordinate-coefficients corresponding to each eigenvector[35]. In our die example, the operator 0 that sets up the dice vectors according to our stipulated initial conditions has six mutually orthogonal eigenvectors, namely $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$ and $\mathbf{x}_6$. To these eigenvectors the eigenvalues $1, 2, 3, 4, 5, 6$ correspond, respectively. With the set of eigenvectors as basis, any vector in our space can be expressed in terms of the corresponding coordinate-coefficients c_i. We can tie up all this and express the probability information that the vectors give us in the following way: the possible defined values that the face-state of a dice can take—its *eigen-states*—are given by the eigenvalues of the operator 0, and the probability that each of those eigenvalues obtains is given by the square of the coefficient of the vector $\mathbf{v}$ on each eigenvector of 0.

Von Neumann's formulation of SQM provides us information about quantum systems in a way analogous to this dice example. But this is just an analogy. In the case of the dice the probability expressed is just a measure of our ignorance: we talk about probabilities because we cannot see inside the box they fall into, but we know that each dice is indeed in a specific face-state once it falls in the box. In SQM probabilities have a more objective meaning. Actually, unless we do look inside the box, we cannot really say that the face-state of a quantum die in the box is an eigenstate. Furthermore, this analogy does not consider the essential relevance of superpositions—states represented by a vector which does not lie on an eigenvector. However, I think that this analogy is useful in order to grasp an intuitive grasp of how the mathematical framework of SQM depicts the physics. So, we can now move on and take a look at the postulates of the theory.

3.2.1 The postulates of SQM

In SQM the wave functions of wave mechanics are replaced by *state vectors*. These vectors inhabit Hilbert space. State vectors are represented by *kets*, a term introduced by Paul Dirac. In classical vector space the inner product between two vectors is usually denoted as $(\mathbf{m} \cdot \mathbf{n})$, but Dirac introduced a notation in which the inner product between two quantum state vectors is denoted as $\langle m|n\rangle$. He split this symbol in two parts, a *bra* $\langle m|$ and a *ket* $|n\rangle$, where the ket $|n\rangle$ represents our state vector and contains all the physical properties and information that a wave function defines[36]. With this in mind we can formulate the first postulate of the theory:

> *Postulate 1. State postulate.* Every physical system has a Hilbert space $\mathcal{H}$ as its state space, and the state of the system is represented by a unit vector in $\mathcal{H}$. A composite physical system corresponds to the direct product of the Hilbert spaces of the subsystems.

As it is clear from the dice example, that the vector $|n\rangle$ has to be a unit vector ($\langle n|n\rangle = 1$) is required for the consistency of the probabilities the vector contains as information. Postulate 1 also mentions that the Hilbert space of composite systems—systems of interacting particles—is given by the direct or tensor product of the spaces of the subsystems. This feature will be relevant when we consider some conceptual issues in which entangled particles are involved, so I will postpone this discussion.

---

[35] For an operator to have eigenvectors that are mutually orthogonal and that span the whole space—one eigenvector with a different eigenvalue for each dimension—, such that they can be used as an orthogonal basis, the operator must satisfy certain formal conditions. We take for granted that our operator satisfies them.

[36] If the ket $|n\rangle$ can be expressed as a column matrix, the bra $\langle n|$ is the adjoint of that matrix. Given a matrix A, the adjoint matrix $A^\dagger$ is obtained by interchanging rows and columns and taking the complex conjugates of the elements of A.

We saw in our dice example that in order to set up a vector that provides us with probabilistic information for certain initial conditions we needed to apply a certain operator on the vector dices. In the case of SQM, if, given a certain system, we want to obtain information regarding a specific observable physical property, we need to apply a suitable operator to the corresponding state vector $|n\rangle$. For this reason, in SQM observable physical properties are identified with operators in Hilbert space. This is what postulate 2 states:

> *Postulate 2. Observables postulate*. Every physical quantity $\mathcal{A}$ of the system corresponds to a self-adjoint or Hermitian operator[37] A in $\mathcal{H}$.

There is no general agreement on whether or not to each possible Hermitan operator there corresponds an observable physical quantity, and *vice versa*. There are certain possible operators for which it is very difficult to conceive an experimental setup to measure the corresponding observable quantity. On the other hand, the measurement of time in physical processes, for example, is not represented by a Hermitian operator.

In the dice example above I mentioned that any vector $\mathbf{v}$ can be expressed as a linear combination of basis vectors given by the eigenvectors of a certain operator 0. We can now express this in the technical jargon of SQM. The *spectral theorem* tells us that a normal operator[38] A in a space $\mathcal{H}$ has a set of mutually orthogonal eigenvectors $|\alpha_i\rangle, \dots, |\alpha_n\rangle$ associated to a set of respective eigenvalues $a_1, \dots, a_n$—the latter set is called the *spectrum* of A. Thus, the operator A can be written as $A = \sum_i a_i |\alpha_i\rangle\langle\alpha_i|$. This last expression can be clarified by explaining the meaning of $|\alpha_i\rangle\langle\alpha_i|$. A projector $P_\phi = |\phi\rangle\langle\phi|$ is an operator that projects any vector $|\psi\rangle$ onto the vector $|\phi\rangle$, that is, $P_\phi|\psi\rangle \longmapsto \langle\phi|\psi\rangle\,|\phi\rangle = |\phi\rangle\langle\phi|\psi\rangle$ [39]. Thus, the operator $|\alpha_i\rangle\langle\alpha_i|$—let us dub it P_{α_i}—projects any vector $|\psi\rangle$ onto the eigenvector α_i of the operator A. Therefore, A can be written as $A = \sum_i a_i P_{\alpha_i}$. This expression is called the *spectral decomposition* of A. With all this in mind we can now address the next two postulates of SQM:

> *Postulate 3. Spectrum postulate*. The only possible outcomes that can be found upon a measurement of a physical quantity $\mathcal{A}$, corresponding to an operator A, are values from the spectrum of A.

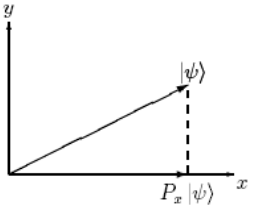
In simple words, this means that if we decide to measure the value of a physical quantity on a system in a state described by a certain vector $|\psi\rangle$, the only values that we can obtain are given by the eigenvalues of the corresponding operator. It is important to remark that the eigenvalue restriction of the possible

---

[37] A self-adjoint or Hermitian operator is an operator which is identical to its adjoint, that is, $A = A^\dagger$. The adjoint of an operator is obtained from the original operator by interchanging the indexes and taking the complex conjugate of each element. If the operator is a matrix, this simply means that rows become columns and columns become rows, and complex conjugates are taken. A Hermitian operator has only real eigenvalues (the values of observable quantities must be real numbers, of course) and its eigenvectors are mutually orthogonal (this allows that the probabilities defined are consistent).

[38] An operator A is normal if it commutes with its adjoint, that is, if $AA^\dagger = A^\dagger A$. All Hermitian operators are normal, so the theorem holds for all observables.

[39] The following figure provide a clear example of what a projector does:



In this case the operator P_x projects the vector $|\psi\rangle$ onto the x-axis. An operator P is a projector if it is Hermitian and idempotent, that is, if $P = P^\dagger$ and $P = P^2$.

values to be obtained holds only for measurements: Postulate 3 does not intend a description of systems that we are not looking at.

If the eigenvalues in the spectrum of the operator A are the only possible values that we can obtain in an experiment designed to measure the quantity $\mathcal{A}$, we now need to know what is the probability for each of those eigenvalues to obtain in the mentioned experiment. This is what the fourth postulate states:

> *Postulate 4. Born postulate.* If the system is in a state $|\psi\rangle \in \mathcal{H}$, and a physical quantity $\mathcal{A}$, corresponding to an operator A with a discrete spectrum[40] Spec A, is being measured, then the probability of finding the outcome $a_i \in$ Spec A, is equal to $\text{Prob}^{|\psi\rangle}(a_i) = \langle\psi|P_{\alpha_i}|\psi\rangle$, where P_{α_i} is the projector from the spectral decomposition of A, namely, $A = \sum_i a_i P_{\alpha_i}$.

We saw above that Born noticed that for a wave function expressed as $\sum_i c_i \Psi_i$, the probability of that the outcome of an experiment is represented bu the eigenvalue corresponding to the eigenfunction Ψ_i is given by $|c_i|^2$. The connection of Born's formulation and the fourth postulate is simple. We know that the operator P_{α_i} projects the state vector $|\psi\rangle = \sum_i c_i|\alpha_i\rangle$ (where the c_i are the coordinate-coefficients corresponding to each eigenvector of A) onto the eigenvector $|\alpha_i\rangle$. This means that the coordinate-coefficients of the vector $|\psi\rangle$ corresponding to the other eigenvectors $|\alpha_j\rangle$, with $j \neq i$, become zero once the projector has operated on $|\psi\rangle$, i.e., $P_{\alpha_i}|\psi\rangle = c_i|\alpha_i\rangle$. Now, since $\langle\psi|P_{\alpha_i}|\psi\rangle$ is the inner product between the vectors $|\psi\rangle$ and $c_i|\alpha_i\rangle$, it is clear that the result of this operation is $|c_i|^2$, so that $\text{Prob}^{|\psi\rangle}(a_i) = \langle\psi|P_{\alpha_i}|\psi\rangle = |c_i|^2$ [41].

An important connection between postulates 3 and 4 is that according to the former experiments to measure an observable quantity $\mathcal{A}$ on a system represented by a vector $|\psi\rangle$ can only produce outcome values belonging to Spec A. According to postulate 4, it follows that the probability of obtaining a value a_i in an experimental result which does not belong to Spec A is zero.

Before we move on to the next postulate it is interesting to take a further look at the probabilistic framework of SQM allowed by postulates 3 and 4. The expectation value of a measurement of a physical quantity $\mathcal{A}$ corresponding to an operator A for a certain state vector $|\psi\rangle$ is the weighted sum of all the possible results, that is, we multiply each possible value by its probability and sum up all these products. More formally $\langle A_\psi\rangle = \sum_i |c_i|^2 a_i$. In the formalism of SQM the expectation value or probability density $\langle A\rangle$ can also be expressed as $\langle A_\psi\rangle = \langle\psi|A|\psi\rangle$. Let us assume that the vector $|\psi\rangle$ corresponds to one of the eigenvectors of A, that is $A|\psi\rangle = a_i|\psi\rangle$, so that $\langle A_\psi\rangle = \langle\psi|a_m|\psi\rangle = a_m\langle\psi|\psi\rangle$. Since $|\psi\rangle$ is assumed to be normalized (a unit vector), then $\langle\psi|\psi\rangle = 1$, so that $\langle A_\psi\rangle = a_m$ (or, equivalently, $|c_m|^2 = 1$). In this case we say that the vector $|\psi\rangle$ is an *eigenstate* of A. If a state vector $|\psi\rangle$ is an eigenstate of an operator A we know with certainty that if measure the quantity $\mathcal{A}$ we will obtain the corresponding eigenvalue.

Let us now assume that $|\psi\rangle$ is not an eigenstate of A and define it as $|\psi\rangle = c_m|\alpha_m\rangle + c_n|\alpha_n\rangle$, where $|\alpha_m\rangle$ and $|\alpha_n\rangle$ determine the set of eigenvectors of A (our space $\mathcal{H}$ has two dimensions)—we also assume that they form an orthonormal basis, that is, that $|\alpha_m\rangle$ and $|\alpha_n\rangle$ are unit vectors. In this case[42],

---

[40] There are operators, such as the position operator, that have a continuous spectrum, that is, its eigenvalues range over a continuum. Accordingly, the set of eigenvectors of that operator is infinite, so that the corresponding Hilbert space $\mathcal{H}$ has an infinite number of dimensions. The mathematical techniques to handle these cases are more complicated and the postulates need to be somewhat adapted. Fortunately, reference to these complicated cases are not necessary to grasp the main conceptual framework of SQM.

[41] A simple proof is presented in (Hughes 1989, 70). We have that $\text{Prob}^{|\psi\rangle}(a_i) = \langle\psi|P_{\alpha_i}|\psi\rangle$. By idempotence of P_{α_i}, $\langle\psi|P_{\alpha_i}|\psi\rangle = \langle\psi|P_{\alpha_i}P_{\alpha_i}|\psi\rangle$, and by the Hermiticity of P_{α_i}, $\langle\psi|P_{\alpha_i}P_{\alpha_i}|\psi\rangle = \langle P_{\alpha_i}\langle\psi||P_{\alpha_i}|\psi\rangle\rangle$. Now, since $P_{\alpha_i}|\psi\rangle = c_i|\alpha_i\rangle$, we have that $\langle P_{\alpha_i}\langle\psi||P_{\alpha_i}|\psi\rangle\rangle = \langle c_i\langle\alpha_i||c_i|\alpha_i\rangle\rangle = c_i^* c_i\langle\alpha_i|\alpha_i\rangle = |c_i|^2$.

[42] Notice that $\langle\psi| = c_m^*\langle\alpha_m| + c_n^*\langle\alpha_n|$. If we think of $|\psi\rangle$ as given by a column matrix, then $\langle\psi|$ is the adjoint or transpose conjugate of that matrix.

$$\langle A_\psi \rangle = \langle \psi | A | \psi \rangle = \langle c_m^* \langle \alpha_m | + c_n^* \langle \alpha_n | | A | | c_m | \alpha_m \rangle + c_n | \alpha_n \rangle \rangle$$
$$= \langle c_m^* \langle \alpha_m | + c_n^* \langle \alpha_n | | c_m A | \alpha_m \rangle + c_n A | \alpha_n \rangle \rangle$$
$$= |c_m|^2 \langle \alpha_m | A | \alpha_m \rangle + c_m^* c_n \langle \alpha_m | A | \alpha_n \rangle + c_n^* c_m \langle \alpha_n | A | \alpha_m \rangle + |c_n|^2 \langle \alpha_n | A | \alpha_n \rangle$$

Now, since $|\alpha_m\rangle$ and $|\alpha_n\rangle$ are unit eigenvectors of A, it follows that $\langle \alpha_m | A | \alpha_m \rangle = a_m$ and that $\langle \alpha_n | A | \alpha_n \rangle = a_n$, where a_m and a_n are the eigenvalues of the corresponding eigenvectors. On the other hand, since $|\alpha_m\rangle$ and $|\alpha_n\rangle$ are orthogonal eigenvectors of A, it follows that $\langle \alpha_m | A | \alpha_n \rangle = a_n \langle \alpha_m | \alpha_n \rangle = 0$ and that $\langle \alpha_n | A | \alpha_m \rangle = a_m \langle \alpha_n | \alpha_m \rangle = 0$. Therefore $\langle A_\psi \rangle = |c_m|^2 a_m + |c_n|^2 a_n$. We can generalize for state vectors in n-dimensional spaces to obtain $\sum_i |c_i|^2 a_i$, as we would expect.

To determine the *expansion coefficients* c_m and c_n, we can simply take the inner product between the corresponding eigenvector and the state vector $|\psi\rangle$. That is, $\langle \alpha_m | \psi \rangle = \langle \alpha_m | c_m | \alpha_m \rangle + \langle \alpha_m | c_n | \alpha_n \rangle = c_m \langle \alpha_m | \alpha_m \rangle + c_n \langle \alpha_m | \alpha_n \rangle = c_m$, and by the same token, $\langle \alpha_n | \psi \rangle = c_n$. Thus, we can write the state vector as $|\psi\rangle = |\alpha_m\rangle\langle \alpha_m | \psi \rangle + |\alpha_n\rangle\langle \alpha_n | \psi \rangle$. The inner products $\langle \alpha_m | \psi \rangle$ and $\langle \alpha_n | \psi \rangle$ are sometimes called *projection amplitudes*, and we can also recognize the projector operators $P_{\alpha_m} = |\alpha_m\rangle\langle \alpha_m|$ and $P_{\alpha_n} = |\alpha_n\rangle\langle \alpha_n|$. In other words, we can express the state vector $|\psi\rangle$ in two equivalent ways: in terms of the expansions coefficients (or the projection amplitudes and their corresponding eigenvectors), or in terms of the projectors operators P_{α_m} and P_{α_n} acting on $|\psi\rangle$.

Now we can move on to consider the dynamics of SQM. The basic idea is that given a system that at a certain time t_0 and under certain initial conditions is described by a state vector $|\psi_0\rangle$, we can calculate the state vector $|\psi_i\rangle$ describing the system at a later time t_i. Moreover, this time evolution proceeds deterministically, that is, if at time t_0 and under suitable initial conditions the system is in a state $|\psi_0\rangle$, at time t_i the system will necessarily be in the state $|\psi_{t_i}\rangle$. The receipt to perform the calculation connecting $|\psi_0\rangle$ and $|\psi_i\rangle$ is given by the time dependent Schrödinger equation. Let us recall that this equation can be written in the form $i\hbar \frac{\partial |\psi\rangle}{\partial t} = H|\psi\rangle$. Solving this equation we get $|\psi_t\rangle = e^{(-iH/\hbar)t}|\psi_0\rangle$. The term t in the exponential in the right hand side denotes the interval $(t_i - t_0)$, so that given a system described by a state vector $|\psi\rangle$ at a certain time, we can calculate the state vector for that system at a later time by applying the operator $U = e^{(-iH/\hbar)t}$. Bearing this in mind we can formulate the next postulate

> *Postulate 5. Schrödinger postulate*. As long as no measurements are performed on the system, its time evolution is described by the unitary transformation $|\psi_t\rangle = U|\psi_{t_0}\rangle$.

As the postulate states, U is a unitary operator[43], which means that its application to any two vectors leaves the inner product between the vectors unaltered, and also their modulus. That is, if we apply U to a unit vector $|\psi_{t_0}\rangle$, the resulting vector $|\psi_t\rangle$ is also a unit vector. Another property of the operator U that has very important conceptual consequences is the following. Consider a vector $|\psi_{t_0}\rangle = c_m|\alpha_m\rangle + c_n|\alpha_n\rangle$ that is not an eigenstate of the operator A ($|\alpha_m\rangle$ and $|\alpha_n\rangle$ are the eigenvectors of A). Since U is a unitary operator, it only changes the direction of the vector $|\psi_{t_o}\rangle$, not its length. Therefore, the resulting vector $|\psi_t\rangle$ will, in general, be given by $|\psi_t\rangle = c'_m|\alpha_m\rangle + c'_n|\alpha_n\rangle$. That is, if the operator U is applied to a state vector that is not an eigenstate of a certain observable, the resulting state vector will, in general, not be an eigenstate of that observable either. This conceptual feature is one of the kernels of the measurement problem, which I will address later on.

[43] An operator U is unitary if its self-adjoint is equal to its inverse, that is, if $U^\dagger = U^{-1}$. That the operator U^{-1} is the inverse of an operator U means that $UU^{-1} = U^{-1}U = \mathbf{1}$, where $\mathbf{1}$ is the unity or identity operator. In turn, the identity operator is such that $A\mathbf{1} = \mathbf{1}A = A$.

3.2.2 Superposition

Let us recall the state vector of our die in the example above. Let us assume that the die is already inside the box and that its state vector is not an eigenstate of the 'face-up' property. This simply means that the probabilities defined by the vector are a measure of our ignorance of the determined 'face-up' state of the dice inside the box—we talk about probabilities simply because we cannot look inside, but we know that the dice is showing one of its faces up. That is, in a classical framework, the only possible objective and determined states that a system can take, regardless of whether we are observing it or not, are eigenstates. In SQM the situation is dramatically different. That a system is described by a state vector which is not an eigenstate of a certain operator A does not mean that such a vector expresses a measure of our ignorance with respect to its determined value of the quantity $\mathcal{A}$. In QM a state vector $|\psi\rangle = |\alpha_i\rangle$ is on an equal footing as a vector $|\varphi\rangle = c_i|\alpha_i\rangle + c_j|\alpha_j\rangle$, where $|\alpha_i\rangle$ and $|\alpha_j\rangle$ are orthonormal eigenvectors of operator A. This feature constitutes the principle of superposition in QM:

> *Principle of superposition.* If $|\psi_1\rangle$ and $|\psi_2\rangle$ represent possible states of a system, then any linear combination of these vectors, that is, any normalized vector of the form $c_1|\psi_1\rangle + c_2|\psi_2\rangle$, also represents a possible state of that system.

In other, words, that the vector representing the state of a system is not an eigenstate means that the state of that system is a superposition of eigenstates—that the system is in a state of superposition.

That superposed states are not simply an expression of our ignorance can be clearly seen in the empirical effects that this kind of states have in certain experiments. A paradigmatic example is given by the Stern-Gerlach experiment. The essentials of this experiment are the following. A beam of silver atoms is made to pass between the poles of a specially shaped magnet to be finally detected on a screen. The result was that the beam was split in two, with half the atoms deflected upwards and the other half deflected downwards. From a classical point of view, the most plausible hypothesis would be to consider the atoms as tiny magnets that due to their interaction with the magnetic field get deflected. However, we would expect that the orientation of the magnetic axes of these tiny magnets were randomly distributed, and under this assumption the result of the experiment would be a smeared line of silver atoms on the screen, not two spots, one upwards and one downwards.

The quantum explanation of the experiment is given by the property of electron spin. As we saw above, electrons possess an intrinsic angular momentum, known as spin, associated to an intrinsic magnetic moment. This spin can take, in any direction, only two possible values $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$. Silver atoms have 47 electrons, 46 of which are arranged in pairs that mutually cancel their spins (recall Pauli's exclusion principle). Therefore, the total spin of a silver atom is given either by $\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$ (the magnetic moment associated to the spin of the protons in the nucleus is negligible given their comparatively large mass). Now, we can assume that the orientation of the device that produces the magnetic field in the S-G is vertical (V). Since the spin of the silver atoms can take only the two mentioned values, the result is that the beam is split in the way explained above: atoms with spin $\frac{1}{2}\hbar$ are deflected upwards and atoms with spin $-\frac{1}{2}\hbar$ are deflected downwards. If we then block the beam deflected downwards and repeat the experiment on the upward beam—with the magnetic field in orientation V—we find that all the atoms get deflected upwards in the second application of the experiment. We might thus conclude that in the original beams half the atoms had a spin up and the other half had a spin down, and that the magnetic field simply separates them accordingly. But this conclusion does not generally hold.

Instead or repeating the experiment on the upward beam with the field in orientation V, we now take this beam and make it pass between the poles of the magnet horizontally oriented (H). Again, the spin of the atoms in the H direction can take only two values, so the upward beam gets split in two, one towards

the right and one towards the left. Let us now block the beam deflected leftward and apply once again the experiment with orientation V to the remaining beam. We might suppose that all the atoms should be deflected upwards—we could assume that the results of the first two experiments left us with a beam of atoms with spin-up and spin-right. However, the observed result is that the beam is again split in two, one deflected upwards and one deflected downwards, just as in the first experiment. The physical underpinning of this strange result is given by the connection between incompatible (non-commuting) observables and superposition.

Let us suppose that a state vector $|\psi\rangle$ is an eigenstate of an operator A. Therefore, $A|\psi\rangle = a_\psi|\psi\rangle$ and we know that the expectation value of A is a_ψ. That is, if we measure the quantity $\mathcal{A}$ on the system given by $|\psi\rangle$ we will get the value a_ψ. Suppose now that, simultaneously, we want to measure the property $\mathcal{B}$ associated to the operator B on the system defined by $|\psi\rangle$. If we are to obtain an eigenvalue b_ψ on this measurement, such that we could say that the system described by $|\psi\rangle$ has a definite value for the properties $\mathcal{A}$ and $\mathcal{B}$ simultaneously, then $|\psi\rangle$ must simultaneously be an eigenstate of both A and B. However, this is, in general, not possible. If a certain state vector is an eigenstate of an operator A, it cannot, in general, be also an eigenstate of an operator B that does not commute with A[44].

This is exactly what happens in the S-G experiment. The operators corresponding to perpendicular orientations of spin do not commute[45]. Therefore, if a silver atom is in an eigenstate of spin in the vertical direction, it is in a superposition of states of spin in the horizontal direction. Let us assume that the spin eigenvectors for the V direction are $|v_+\rangle$ and $|v_-\rangle$ and that the spin eigenvectors for the H direction are $|h_+\rangle$ and $|h_-\rangle$. After the first application of the experiment in the V direction we kept only the beam containing the atoms in the eigenstate $|v_+\rangle$. To this beam we then applied the experiment in the H direction and kept only the atoms in the eigenstate $|h_+\rangle$. Since the spin operators of perpendicular directions do not commute, the vector describing the state of the silver atoms cannot be an eigenstate $|v_+\rangle$ and $|h_+\rangle$ simultaneously. Actually, if an atom is in an eigenstate $|h_+\rangle$, its state vector expressed in terms of the eigenvectors of V is $\frac{1}{\sqrt{2}}|v_+\rangle + \frac{1}{\sqrt{2}}|v_-\rangle$[46]—which explains the result of the second application of the experiment in the V direction described above. As it can be seen, the result of the S-G experiment shows that if a system is described by a superposition-of-eigenstates vector, it represents an objective feature of reality, not a measure of our ignorance. It is clear that $|h_+\rangle$ describes the state of a system, therefore, and since $|h_+\rangle = \frac{1}{\sqrt{2}}|v_+\rangle + \frac{1}{\sqrt{2}}|v_-\rangle$, a superposition of eigenstates of vertical spin also represents the state of a system.

Another general conclusion that can be drawn from these considerations concerns the active role of experiments form in SQM. A measurement is not a passive, unraveling discovery of the value of a certain property in a given system. As it can be noticed in the second application of the S-G experiment in the V direction, the measurement actively determines the observed value of the property we are looking for—according to the statistics delivered by the Born rule. In SQM, the process of observation becomes a part of the physical process that is being investigated. The classical separation between system observed and observation method is not possible here.

---

[44] An exception for this rule can occur if the noncommuting operators A and B share an eigenvector, for example, so that a state vector that corresponds to this eigenstate (the shared eigenvector) is actually an eigenstate of both operators.

[45] The operators for spin in the x, y and z directions are given by the *Pauli spin matrices* $S_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $S_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $S_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. It is easy to show that these matrices do not commute. For example, $[S_x, S_y] = 2iS_z$. In this formulation the operators have eigenvalues ± 1, so we assume that the corresponding property values are given in units of $\frac{1}{2}\hbar$.

[46] This holds in general for all orthogonal directions of spin: if $|x_+\rangle$ is an eigenstate of spin in the x-direction, then $|x_+\rangle = \frac{1}{\sqrt{2}}|y_+\rangle + \frac{1}{\sqrt{2}}|y_-\rangle = \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle$.

3.2.3 Indistinguishable particles

In classical physics—and also in everyday experience and ordinary language—individual objects are distinguished from each other in terms of properties. No matter how alike two entities may be, if they are two different ones there must be a property to distinguish them. With the advent of modern chemistry and electrodynamics the physical properties that distinguish elementary particles became restricted. For example, it became clear that all electrons have exactly the same mass and the same electric charge, and all atoms of a certain element were exactly alike. That is, the 'intrinsic' non-relational properties of objects like these could not count as the distinguishing features. However, one of these intrinsic properties—though common to all elementary particles—, impenetrability, could be the ground to distinguish between otherwise identical objects in terms of their spatio-temporal features. If atoms, electrons and the like are impenetrable bodies, they cannot occupy the same position at the same time, and, accordingly, their spatio-temporal histories must be necessarily different. To say it in more precise, relativistic words, if the motion of a particle is represented by its four-dimensional world-line in space-time, then a family of nonintersecting world-lines corresponding to atoms or electrons represents a family of different atoms or electrons. Conversely, if two electrons or atoms are two different objects, their world-lines cannot intersect. Therefore, the world-line of a particle can be used as a label to differentiate it from all other particles of the same kind.

However, in SQM the spatio-temporal history of elementary particles cannot be invoked in order to distinguish them. The reason is given by the wave-particle duality and the uncertainty relations. According to the latter, there is no such thing as a spatio-temporal trajectory for an elementary particle. A spatio-temporal trajectory is a continuous process—to every instant there corresponds a specific position—, but Heisenberg noticed that even though we are able to observe something like a trajectory in an electron cloud chamber, what we really see is a consecutive record of discrete interactions of an electron with ionized atoms—pretty much like when we watch a sequence of discrete photograms in a movie theater. In each interaction—the electron ionizes a molecule in the chamber and a droplet is formed—the position of the electron gets determined and recorded, but its momentum gets undetermined according to the uncertainty principle and the wave function of the electron spreads once again, until it gets localized in the next interaction. Given the fast pace of the interactions we observe something like a trajectory, but we cannot really say that the electron is at a specific place between the interactions.

The wave-particle duality entails that in SQM we cannot really talk about impenetrability. Think of two electrons. In the absence of position measurements, their respective positions are given by their wave functions, that is, the electrons are spatially spread out. It might be the case that their wave functions overlap and therefore interfere. If this is so, the 'positions' of the two particles spread out along the same spatio-temporal region, so the principle of impenetrability does not strictly hold in SQM. More importantly, if in a region of constructive interference due to the wave function overlap we do find an electron, there is no way to determine to which of the two original electrons it corresponds, not because we cannot find that out, but because there is not fact of the matter as to the question 'which of the two original electrons is this?'

The fact that spatio-temporal features fail to work as criteria to distinguish between quantum particles is usually taken to imply that in SQM two identical particles can be essentially indistinguishable. The example that is normally used to explain this feature is given by the Bose-Einstein statistics that rule the behavior of bosons. Let us review first the classical case. Consider two classical particles a_1 and a_2 inside a box with two compartments (states) α and β that they can occupy:

	α	β
(1)	a_1, a_2	
(2)		a_1, a_2
(3)	a_1	a_2
(4)	a_2	a_1

As it is clear, we have four different arrangements for the distributions of particles in compartments in the box, each with a probability of $\frac{1}{4}$. Let us assume that all the physical and chemical properties of a_1 and a_2 are the same, so that we can only distinguish and label them in terms of their spatio-temporal histories. This distinction is enough to assure that the distributions (3) and (4) are indeed different distributions—particle interchanges result in different configurations. This way of counting corresponds to Maxwell-Boltzmann statistics, and since it is relevant for the derivation of formulas that describe the entropy of a system (a gas, for example), this mode of counting can (and has been), empirically tested.

Let us now suppose that the particles are photons. Instead of compartments in a box we have to think in terms of possible quantum states that the two photons can take. Unlike classical physics, the table for the possible different arrangements is given by

	α	β
(1)	a, a	
(2)		a, a
(3)	a	a

Now we have only three different possible arrangements, each with a probability of $\frac{1}{3}$. Actually, this is the correct mode of counting for black-body radiation. Einstein and Bose showed that if the Maxwell-Boltzmann method is used, then Wien's law follows, whereas the correct formula (Planck's) could be derived by using this alternative way of counting, which is therefore known as Bose-Einstein statistics.

The indistinguishability between the photons becomes manifest when we notice that particle interchange does not result in a different arrangement, as row (3) shows. That is, there are no empirical differences associated to a substitution of particles when there is one in each possible state. If we recall that spatio-temporal histories are not available as a criterion to label the particles, it then follows that in SQM there are cases in which two particles are completely indistinguishable—this the reason why sub-indexes labeling the particles in the classical case have now been omitted. More formally, we have that in the case of *bosons*, the quantum particles that are subject to Bose-Einstein statistics, the possible states are given by $|\alpha\rangle_1|\alpha\rangle_2$, $|\beta\rangle_1|\beta\rangle_2$, and $\frac{1}{\sqrt{2}}(|\alpha\rangle_1|\beta\rangle_2 + |\beta\rangle_1|\alpha\rangle_2)$[47], and the fact that particle interchange has no observable consequences whatsoever is expressed in the

[47] If we were considering *fermions*, the particles which are subject to the Fermi-Dirac statistics, the only possible state would be $\frac{1}{\sqrt{2}}(|\alpha\rangle_1|\beta\rangle_2 - |\beta\rangle_1|\alpha\rangle_2)$, so that the counting of possible states would be given by

	α	β
(1)	a	a

Electrons, for example, are fermions. We saw above that they must respect Pauli's exclusion principle: no two electrons (fermions) can be in the same quantum state. This restriction can be seen in that for fermions $\frac{1}{\sqrt{2}}(|\alpha\rangle_1|\alpha\rangle_2 - |\alpha\rangle_1|\alpha\rangle_2) = 0$. Under index permutation the state of a boson remains exactly the same. For example, in the case of a boson given by the ket $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\alpha\rangle_1|\beta\rangle_2 + |\beta\rangle_1|\alpha\rangle_2)$, index interchange gives $\frac{1}{\sqrt{2}}(|\alpha\rangle_2|\beta\rangle_1 + |\beta\rangle_2|\alpha\rangle_1) = |\Psi\rangle$, that is, indistinguishable bosons stand in *symmetric* states. On the other hand, a fermion given by the ket $|\Phi\rangle = \frac{1}{\sqrt{2}}(|\alpha\rangle_1|\beta\rangle_2 - |\beta\rangle_1|\alpha\rangle_2)$ becomes $\frac{1}{\sqrt{2}}(|\alpha\rangle_2|\beta\rangle_1 - |\beta\rangle_2|\alpha\rangle_1) = -|\Phi\rangle$ under index permutation, that is, indistinguishable fermions stand in *anti-symmetric* states. Since the probability for experimental results is given by the modulus square of the expansion coefficients, the + and –

Indistinguishability postulate. If a particle interchange is performed to any ket for an assembly of particles, then there is no way to distinguish the resulting permuted ket from the original one by means of observation. Let $\hat{P}$ be the permutation operator. It thus holds that $\hat{P}$ commutes with any observable O, such that $\langle O_\psi \rangle = \langle \psi | O | \psi \rangle = \langle \hat{P}\psi | O | \hat{P}\psi \rangle$.

Together with the requirement that the states describing indistinguishable particles must be either symmetrical or anti-symmetrical, this postulate entails that Bose-Einstein and Fermi-Dirac statistics are the only allowed ones in SQM.

If in SQM there are indistinguishable particles, then how is the individuality or identity of each particle in an indistinguishable pair instantiated? In modern physics the individuality of an object is normally given by its properties (intrinsic, relational or spatio-temporal), but we have just seen that this is not possible in SQM. Now, if being an object presupposes being an individual, then in what sense are quantum particles objects? Leibniz's famous principle of the identity of indiscernibles (PII) merges the criteria to *distinguish* an object from all other objects and to *individuate* it as an object: two individuals cannot share all their properties. SQM seems to violate Leibniz's PII, therefore, a metaphysical problem is involved: where does the individuality of quantum particles come from? The philosophical and foundations literature on this subject is enormous, and there are several, very different, positions about it. A discussion of this question would take as far afield, so I will only mention three characteristic views. One way out of the problem would be given by stating that the individuality of objects somehow transcends their properties. That is, the individuality of a (quantum) object may be given by a *haeccity* or 'primitive thisness'. The strong metaphysical, even scholastic, commitments of a position like this are obvious. Another possible way out consists in relaxing the PII, such that so called irreflexive relations that indistinguishable particles possess might weakly differentiate them in terms of properties after all. Finally, we could simply deny that the concept of particle is appropriate for states like $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\alpha\rangle_1|\beta\rangle_2 \pm |\beta\rangle_1|\alpha\rangle_2)$. The basic idea is to save the notion of individuality by speaking of *field quanta* instead of particles—the individuality of $|\Psi\rangle$ resides then in the formalism of quantum field theory—and reserve the term 'particle' for entities that emerge from quantum states in the classical limit[48].

3.2.4 Pure states, mixtures and improper mixtures

State vectors are not the only way to describe the state of a system in SQM. Another important element of the formalism that can accomplish this function is known as the *density operator*. In general, density operators are used to describe the state of a *mixed* system, a system constituted by several subsystems, i.e., an *ensemble*, whereas state vectors describe *pure* individual systems which are not constituted in this way. A density operator ρ is defined in the following way, $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. The vectors $|\psi_i\rangle$—which are not necessarily orthogonal—describe the pure states that constitute the ensemble, and the p_i represent the fraction of the ensemble in each pure state $|\psi_i\rangle$: if n is the total number of subsystems, $p_i = \frac{n_i}{n}$—so that $\sum_i p_i = 1$. In other words, if $n = 3$, our ensemble may be found in a state $|\psi_1\rangle$ with a probability p_1, in a state $|\psi_2\rangle$ with a probability p_2, or in a state $|\psi_3\rangle$ with a probability p_3. The most important properties of

signs do not yield different observable predictions (states like $|\alpha\rangle|\beta\rangle$ are called composite systems, and they are given by the tensor product of the Hilbert spaces of the component states, that is, the Hilbert space corresponding to $|\alpha\rangle|\beta\rangle$ is $\mathcal{H}^\alpha \otimes \mathcal{H}^\beta$, where $|\alpha\rangle \in \mathcal{H}^\alpha$ and $|\beta\rangle \in \mathcal{H}^\beta$—see postulate 1 above).

[48] For indistinguishable particles, see (Teller 1998). For *weak discernibility*, see (Saunders 2003 and 2006), (Muller & Saunders 2008) and (Muller & Seevinck 2009); for criticism of this position, see (Dieks & Veerstegh 2008) and (Dieks 2010). For individuality and *field quanta* and particles as emergent entities, see (Dieks 1990) and (Dieks & Lubberdink 2011). For a general discussion of quantum indistinguishable particles and individuality, see (French & Krause 2006) and the references therein.

a density operator are *i)* it is Hermitian, so that its eigenvalues are real numbers; *ii)* it is positive definite, i.e., for any vector $|u\rangle$, $\langle u|\rho|u\rangle \geq 0$, so that its eigenvalues are positive; and *iii)* its trace $\mathrm{Tr}\rho = 1$, that is, the sum of its main diagonal elements, is equal to 1[49]. We have extended then our Postulate 1, for quantum systems can also be represented by density operators

Given a mixture described by a density matrix ρ, and given an observable operator A, the expected value of A is defined in the following way: $\langle A_\rho\rangle = \sum_i p_i\langle\psi_i|A|\psi_i\rangle = \sum_i\sum_j p_i a_j|\langle\alpha_j|\psi_i\rangle|^2 = \mathrm{Tr}(\rho A)$, where $|\alpha_j\rangle$ and a_j are the eigenvalues and eigenvectors of A, respectively. In words, the expected value of an observable A for a mixture ρ is given by the sum of the expectation values for each pure state $|\psi_i\rangle$ in ρ, weighted by the corresponding probabilities p_i. The probability that a specific value a_i is obtained in a measurement of A on ρ is given by $\mathrm{Prob}^\rho(a_i) = \mathrm{Tr}(\rho P_{\alpha_i})$, where P_{α_i} is the corresponding projection operator in the spectral decomposition of $A = \sum_i a_i P_{\alpha_i}$. In the case of the dynamical evolution of a system represented by a density matrix, we have that $\rho_t = U\rho_{t_0}U^{-1}$. These remarks give us the generalizations of Postulates 4 and 5 for the case of states represented by density operators.

Density matrices can indeed be used to describe pure states. However, for a density matrix ρ to describe a pure state it must hold that $\rho = \rho^2$, that is, the density matrix must be a projector operator, whereas in the case of mixtures, it holds that $\rho \neq \rho^2$. Beyond this technical difference, we can clearly appreciate the difference between a pure state and a mixture through the following example. Consider an electron described by the state vector $|x_+\rangle$ which is an eigenstate of the operator S_x corresponding to the property of spin in the x-direction—that is, our electron spin in the x-direction has the value $\frac{1}{2}\hbar$. The density matrix for this state is $\rho = |x_+\rangle\langle x_+|$, and it can be shown that, also in the case of pure states represented by vectors, the expectation value $\langle S_x\rangle = \langle x_+|S_x|x_+\rangle = \mathrm{Tr}(\rho S_x) = \frac{1}{2}\hbar$[50]. Since $|x_+\rangle = \frac{1}{\sqrt{2}}(|y_+\rangle + |y_-\rangle)$, we know that the probability for a measurement of S_y on ρ yields a probability of $\frac{1}{2}$ for the possible results $\pm\frac{1}{2}\hbar$ of spin in the y-direction. Let us now compare this pure state with the mixture given by $\rho' = \sum_i p_i|y_i\rangle\langle y_i|$, where the $|y_i\rangle$ are given by the pure states $|y_+\rangle$ and $|y_-\rangle$—which are in turn the eigenstates of the operator S_y for spin in the y-direction—and both the $p_i = \frac{1}{2}$. That is, the probability to find the system ρ' in each of the possible eigenstates of S_y is $\frac{1}{2}$, just as in the case of the pure state ρ. But what is the probability of the system ρ' to be found in the state $|x_+\rangle$ after a measurement of S_x? From what was said in the discussion of superposition, we know that if the system represented by ρ' is in the state $|y_+\rangle$ the probability of obtaining a result $|x_+\rangle$ is $\frac{1}{2}$, but since the probability of system ρ' to be in the state $|y_+\rangle$ is also $\frac{1}{2}$, we can 'save' a probability of $\frac{1}{4}$ for $|x_+\rangle$. The same result obtains when we consider $|y_-\rangle$ in ρ': we also get a probability of $\frac{1}{4}$ for $|x_+\rangle$. Therefore, the total probability of obtaining a result $|x_+\rangle$ in a measurement of S_x in ρ' is $\frac{1}{2}$, whereas this same probability in the case of the pure state ρ is 1, of course.

The mixture ρ' is very illustrative regarding some important conceptual issues about the interpretation of mixed states in SQM. As I said in the beginning of this section, the natural interpretation of a state $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ is that the ensemble is formed by pure states $|\psi_i\rangle$, each of them in a proportion p_i. Another

---

[49] More formally, the trace TrA of an operator A is defined as $\mathrm{Tr}A = \sum_i\langle\gamma_i|A|\gamma_i\rangle$, where the $|\gamma_i\rangle$ constitute an arbitrary orthonormal basis in the n-dimensional space $\mathcal{H}$ in which A operates. The fact that this basis is arbitrary shows that the value of TrA is an intrinsic property of A. Besides, if we choose the set of eigenvectors of A as basis, it is clear that TrA is equal to the sum of the eigenvalues of A.

[50] Consider the pure state represented by the state vector $|\psi\rangle = c_m|\alpha_m\rangle + c_n|\alpha_n\rangle$, where the c_i are the expansion coefficients and the $|\alpha_i\rangle$ are the eigenvectors of A. As we saw above, $\langle A_\psi\rangle = \langle\psi|A|\psi\rangle = |c_m|^2\langle\alpha_m|A|\alpha_m\rangle + c_m^*c_n\langle\alpha_m|A|\alpha_n\rangle + c_n^*c_m\langle\alpha_n|A|\alpha_m\rangle + |c_n|^2\langle\alpha_n|A|\alpha_n\rangle = |c_m|^2a_m + |c_n|^2a_n = \sum_i|c_i|^2a_i$, where the a_i are the eigenvalues of A. Since A is considered in terms of its eigenvectors, we know that $A = \begin{pmatrix} a_m & 0 \\ 0 & a_n \end{pmatrix}$, and since for a two-state expansion vector like $|\psi\rangle$, the corresponding density matrix is given by $\rho = \begin{pmatrix} |c_m|^2 & c_m^*c_n \\ c_mc_n^* & |c_n|^2 \end{pmatrix}$. Therefore, $\langle A_\psi\rangle = \mathrm{Tr}(\rho A)$.

natural example of the meaning of a mixture is the following. Think of a ball that enters a box through a narrow funnel. Within the box there are several pins, so that we can think of the device as a pinball machine. At the bottom of the box there are six compartments, so that the ball will end up in one of them when it reaches the bottom. A density matrix can represent the state of the ball at the end of the process. There are six possible pure states, and that probability that each of them may be the final result is given by initial conditions, the properties of the box, etc. since the system is very complicated, we do not know all the variables needed, so the density matrix describes a measure of our ignorance. There are cases in QM where an interpretation like this is possible. For instance, a particle accelerator that does not work completely accurately produces electrons in different pure states—we know that the accelerator produces a certain range of pure states—so the ensemble of electrons can be described by a density matrix that in turn can be given the ignorance interpretation just described.

However, there are mixtures for which the ignorance interpretation is very problematic—if possible at all. Consider the case of $\rho' = \frac{1}{2}(|y_+\rangle\langle y_+| + |y_-\rangle\langle y_-|)$ that we just reviewed. It is not difficult to notice that the probability of finding the system in any of the possible eigenstates for the *three* spin-directions x, y and z is $\frac{1}{2}$. We can thus say that this mixed state is completely *unpolarized*. Moreover, we can write density matrices for the spin state of the mixture which are formally different than ρ', but that yield exactly the same probabilities for all spin measurements, namely $\rho = \frac{1}{2}(|x_+\rangle\langle x_+| + |x_-\rangle\langle x_-|)$ and $\frac{1}{2}(|z_+\rangle\langle z_+| + |z_-\rangle\langle z_-|)$[51]. Recall that the natural ignorance interpretation tells us that the system is in one of the pure states, but that we do not know which; or that the mixture is in a weighted combination of pure states. However, none of these options seems tenable in this case, for we cannot even determine which are the possible pure states of the mixture—the degeneracy implies that the families of pure states that can work as projectors in the density matrix are, in principle, infinite. It might be replied that our ignorance is even bigger than what we thought, but then it seems odd to say that we do not know what are the possible pure states, but at the same time we can assign to a particular pair of them determined probabilities that add to unity.

In the case of the totally unpolarized system we have that the degeneracy of the eigenvalues p_i is the reason why the corresponding density matrix can be expressed in terms of different pure states. However, it holds for any density matrix whatsoever that its decomposition is not unique. As we saw above, given a density matrix $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$, the pure states $|\psi_i\rangle$ need not be mutually orthogonal. But since density matrices are Hermitian, we know by the spectral theorem that there exists a decomposition of ρ in terms of its orthogonal eigenvectors, namely, $\rho = \sum_i a_i P_{\alpha_i}$, so that, in general, $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i| = \sum_i a_i P_{\alpha_i}$, but $|\psi_i\rangle\langle\psi_i| \neq P_{\alpha_i}$. Actually, for any density operator $\rho = \sum_i a_i P_{\alpha_i}$ which is not a projector—that is, such that $\rho \neq \rho^2$ and the represented system is indeed a mixture, not a pure state—there is an infinite number of ways of formulating ρ in terms of a weighted sum of projectors onto non-orthogonal states $|\psi_i\rangle$. Therefore, the objection against the ignorance interpretation in the case of the unpolarized spin mixture can be run, in principle, for any mixture expressed by a density matrix[52].

---

[51] The reason why we can write the same density matrix in different ways is, in this case, the degeneracy of the eigenvalues p_i of ρ'—they are all equal to 1/2. As Hughes succinctly explains it: 'the possibility of degeneracy, however, is one reason we cannot guarantee a unique decomposition for **D** [ρ in our notation] […] Assume, for instance, that we have $a_j = a_k$ [$p_j = p_k$]. Then the rays onto which $\mathbf{P}_j$ and $\mathbf{P}_k$ [$|\psi_j\rangle\langle\psi_j|$ and $|\psi_k\rangle\langle\psi_k|$] project span a plane L_{jk} in $\mathcal{H}$, and if $\mathbf{P}'_j$ and $\mathbf{P}'_k$ [$|\psi'_j\rangle\langle\psi'_j|$ and $|\psi'_k\rangle\langle\psi'_k|$] are projectors onto any two orthogonal rays of L_{jk}, we can replace $\mathbf{P}_j$ and $\mathbf{P}_k$ in $\{\mathbf{P}_i\}$ [$\{|\psi_i\rangle\langle\psi_i|\}$] by $\mathbf{P}'_j$ and $\mathbf{P}'_k$ to form a new family $\{\mathbf{P}'_i\}$ of projectors (such that for $j \neq i \neq k$, $\mathbf{P}'_i = \mathbf{P}_i$). We then obtain $\mathbf{D} = \sum_i a_i P_i = \sum_i a_i P'_i$ [$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i| = \sum_i p_i|\psi'_i\rangle\langle\psi'_i|$]' (Hughes 1989, 139).

[52] As Harvey Brown clearly explains, 'the ambiguity in the resolution of a given statistical [density] operator W can be said to have two sources. First, being self-adjoint, W has a spectral resolution; if W is degenerate, the spectral resolution is not unique, as is well known. Second, one can in general construct from the spectral resolution of W another resolution containing projections over one-dimensional subspaces which are not pairwise orthogonal' (Brown 1986, 859, fn. 10).

To conclude this subsection, one final conceptual distinction must be drawn. Consider the system S in a pure state described by a normalized ket $|\psi\rangle = \sum_{i,j} c_i|v_i\rangle\, c_j|u_j\rangle = \sum_{i,j} c_{i,j}|v_i\rangle|u_j\rangle$. This is a state composed of two subsystems $U = \sum_j c_j|u_j\rangle$ and $V = \sum_i c_i|v_i\rangle$. To each of these subsystems a Hilbert space $\mathcal{H}^{(U)}$ and $\mathcal{H}^{(V)}$ corresponds, respectively, and we assume that the sets $\{|u_j\rangle\}$ and $\{|v_i\rangle\}$ correspond to orthonormal bases that span $\mathcal{H}^{(U)}$ and $\mathcal{H}^{(V)}$, respectively. The Hilbert space of the total system $\mathcal{H}^{(S)}$ in which the ket $|\psi\rangle$ is defined is given by the tensor product $\mathcal{H}^{(U)} \otimes \mathcal{H}^{(V)}$ between the two subspaces. A system like S typically results when the systems U and V have interacted in the past, and the ket $|\psi\rangle$ expresses intrinsic correlations between the subsystems, as we will see in due course. In the jargon of QM, states like S are called *entangled*.

Let us now consider an observable A that we measure only on the subsystem V. The mean value of A with respect to the total system S is given by $\langle A_S\rangle = \sum_{i,j} c_{i,j}^* c_{i,j}\langle v_i|\langle u_j|A|v_i\rangle\big|u_j\rangle$, but since A is related only to V we can also write the expectation value as $\langle A_V\rangle = \sum_i c_i^* c_i\langle v_i|A|v_i\rangle = \sum_i p_i\langle v_i|A|v_i\rangle = \mathrm{Tr}(\rho_V A)$, with $\rho_V = \sum_i p_i|v_i\rangle\langle v_i|$. That is, this line of reasoning seems to open the possibility of representing the subsystem V in terms of a density matrix ρ_V and to consider it as a mixture—the three main properties of a density matrix reviewed above hold for ρ_V, and it is the case that $\rho_V \neq \rho_V^2$. We could run the same reasoning in the case of subsystem U and obtain a density matrix ρ_U, and the state that the system S is composed of two mixtures U and V.

But this maneuver does not work. Let us assume that the subsystems are indeed mixtures, and that U and V are weighted sums of subsystems in some of the pure states $|u_j\rangle$ and $|v_i\rangle$, respectively. We could then define a system S_i constituted by all the subsystems in U that are in the pure state $|u_j\rangle$ and all the subsystems in V which are in the pure state $|v_i\rangle$—so that the total system S could be conceived as the sum of all the possible S_i. There must be a density matrix to describe each of the S_i to provide the correct probabilities for the result of possible measurements defined by operators in $\mathcal{H}^{(U)}$ and in $\mathcal{H}^{(V)}$. Let us assume that in an arbitrary S_i the subsystems in the pure states $|u_j\rangle$ and $|v_i\rangle$ are eigenstates of a certain operator in the corresponding sub-space. We have that the only density matrix in $\mathcal{H}^{(S)}$ that gives the correct predictions for a measurement of those observables (probability one for the corresponding eigenstates) is given by $\rho_{i,j} = |u_j\rangle|v_i\rangle\langle u_j|\langle v_i|$. This is clearly a pure state which is also describable by the ket $|u_j\rangle|v_i\rangle$. Now, since we assumed that S is given by the union of all the S_i, it must follow that the system S is describable by a weighted sum of all the $\rho_{i,j}$. That is, it must be possible to depict the system S by a density matrix $\rho' = \sum_{i,j} p_{i,j}\rho_{i,j}$. In this case, S would be a mixture, of course. However, we have that S is described by the vector $|\psi\rangle = \sum_{i,j} c_{i,j}|v_i\rangle|u_j\rangle$, or, alternatively, by the density operator $|\psi\rangle\langle\psi|$, i.e., we started from the fact that S is a pure state, not a mixture, so that the assumption that S can be thought of as the union of two mixtures U and V leads to a contradiction[53]. As a matter of mathematics, the partial states ρ_U and ρ_V are mixed state, but we cannot physically interpret them as representing mixtures, that is, ensembles of pure states.

Moreover, there are measurements that allow to distinguish between the system S from any mixture made up of subsystems like S_i. The measurements that allow the distinction are measurements of the correlations between observables corresponding to the systems U and observables corresponding to V. That is, a mixture described by a density matrix like ρ' does not yield any such correlations in the probabilistic information contained in ρ'. Bernard d'Espagnat has coined the terms *proper mixture* and *improper*

[53] As Hughes clearly states: 'Consider a composite system $\mathbf{D}$ in the pure state $\mathbf{D}$, of which the component states are $\mathbf{D}^{\mathrm{A}}$ and $\mathbf{D}^{\mathrm{B}}$. for the sake of the argument, assume that $\mathbf{D}^{\mathrm{A}} = a_1\mathbf{P}_1^{\mathrm{A}} + a_2\mathbf{P}_2^{\mathrm{A}}$, while $\mathbf{D}^{\mathrm{B}} = b_1\mathbf{P}_1^{\mathrm{B}} + b_2\mathbf{P}_2^{\mathrm{B}}$, with $a_1 \neq a_2$ and $b_1 \neq b_2$, so that there are no problems of degeneracy. Then, according to the ignorance interpretation of $\mathbf{D}^{\mathrm{A}}$ and $\mathbf{D}^{\mathrm{B}}$, system A is really in one of the pure states $\mathbf{P}_1^{\mathrm{A}}$ or $\mathbf{P}_2^{\mathrm{A}}$, and system B is really in one of the pure states $\mathbf{P}_1^{\mathrm{B}}$ or $\mathbf{P}_2^{\mathrm{B}}$. These four states may also be represented by vectors $\mathbf{v}_1^{\mathrm{A}}$, $\mathbf{v}_2^{\mathrm{A}}$, $\mathbf{u}_1^{\mathrm{B}}$ and $\mathbf{u}_2^{\mathrm{A}}$, respectively, such that $\mathbf{P}_1^{\mathrm{A}}\mathbf{v}_1^{\mathrm{A}} = \mathbf{v}_1^{\mathrm{A}}$, and so on. But this would mean that the composite system is really in one of the four states $\mathbf{v}_1^{\mathrm{A}} \otimes \mathbf{u}_1^{\mathrm{B}}$, $\mathbf{v}_1^{\mathrm{A}} \otimes \mathbf{u}_2^{\mathrm{B}}$, $\mathbf{v}_2^{\mathrm{A}} \otimes \mathbf{u}_1^{\mathrm{B}}$ or $\mathbf{v}_2^{\mathrm{A}} \otimes \mathbf{u}_2^{\mathrm{B}}$, with probabilities a_1b_1, a_1b_2, a_2b_1, a_2b_2, respectively—in other words, that the composite system is in a *mixed state*. Since this contradicts our original assumption, the ignorance interpretation simply will not do' (Hughes 1989, 150).

mixtures to distinguish between 'real' mixed states and 'mixture-like' subsystems like U and V. As we will see, that subsystems like U and V cannot be consistently considered as proper mixtures is yet another kernel in the measurement problem.

3.2.5 The measurement problem

In classical physics a measurement is a generally passive process through which we can unravel and observe certain properties of an object or system, and then determine the value of such properties according to a certain metric system. A measurement device interacts with the measured object in a way such that the former shows us the state of the latter. A simple example is the measurement of the mass of an object. We put the object on a balance, and the interaction between them causes that the pointer of the balance indicates a certain value for the mass of the object. In this case the process of measurement is totally passive, the mass of the object is not changed by the interaction—a non-disturbing measurement like this is usually called *ideal*. There are non-ideal measurements in classical physics too, that is, interaction which do alter the property that is measured. However, the dynamical rules of the relevant theories tell us in what way the interaction alters the measured property. Therefore, even in a non-ideal measurement, the laws of classical physics allow us to know precisely the value of the property before the interaction. The important point is that, in classical physics, a measurement is a physical interaction just like any other, in the sense that it is governed by the same dynamical laws that govern all physical interactions.

The situation is different in SQM. Let us call the system being measured I and the system corresponding to the measuring device II. Let us assume that at time t before the measurement interaction the system I is given by the state vector $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\psi_+\rangle + |\psi_-\rangle)$, where the $|\psi_i\rangle$ are the eigenvectors of the operator A corresponding to the property $\mathcal{A}$ that we are measuring on I—we can assume that we are measuring the spin of $|\Psi\rangle$ in a certain direction, for example. The measuring apparatus II is in the state $|\phi_0\rangle$, a state in which the pointer of the apparatus is in the 'ready $= 0$' position. Thus, at time t the total system I + II is given by the tensor product $|\Psi\rangle|\phi_0\rangle = \frac{1}{\sqrt{2}}(|\psi_+\rangle|\phi_0\rangle + |\psi_-\rangle|\phi_0\rangle)$. If we want to know the state of the system I + II at the time t' after the measurement, we have to apply the unitary operator U, of course. By so doing we get $U|\Psi\rangle|\phi_0\rangle = \frac{1}{\sqrt{2}}(U|\psi_+\rangle|\phi_0\rangle + U|\psi_-\rangle|\phi_0\rangle)$.

Now, for the measuring device II to be effective, it must be the case that its interaction with a system in the eigenstate $|\psi_+\rangle$ will yield a resulting total state $|\psi_+\rangle|\phi_+\rangle$, and that its interaction with a system in the eigenstate $|\psi_-\rangle$ will yield a resulting total state $|\psi_-\rangle|\phi_-\rangle$. That the system II is in a state $|\phi_+\rangle$ or $|\phi_-\rangle$ means that its pointer indicates 'spin $(+)$' or 'spin $(-)$', respectively. In other words, if before the measurement we have a total system $|\psi_+\rangle|\phi_0\rangle$, after the measurement we get a total system $U|\psi_+\rangle|\phi_0\rangle = |\psi_+\rangle|\phi_+\rangle$—and by the same token, $U|\psi_-\rangle|\phi_0\rangle = |\psi_-\rangle|\phi_-\rangle$. Therefore, we have that, at time t', our system I + II is in the state $U|\Psi\rangle|\phi_0\rangle = \frac{1}{\sqrt{2}}(|\psi_+\rangle|\phi_+\rangle + |\psi_-\rangle|\phi_-\rangle)$. Hence, the measurement problem. If we stick only to the dynamics of postulate 5, we have that in the measurement interaction I + II we obtain $\frac{1}{\sqrt{2}}(|\phi_+\rangle + |\phi_-\rangle)$ for the measuring device with respect to the 'pointer position' observable, which does not represent a definite outcome for the measured property in I, of course. More generally, the continuous and deterministic time evolution of a quantum system as described by the time dependent Schrödinger equation cannot account for the discontinuous and indeterministic projection of $|\Psi\rangle|\phi_0\rangle$ onto one of the A measurement eigenstates. If before the measurement system I is in a superposed state $\sum_n c_n|\psi_n\rangle$ and system II is in the state $|\phi_0\rangle$, the after-measurement total system is given by the superposition $\sum_n c_n|\psi_n\rangle|\phi_n\rangle$, that cannot represent a definite outcome.

To put it in plain words, we have that the dynamical rules of the theory are governed by the Schrödinger equation through the unitary operator U. But, if we stick *only* to the dynamics, then the theory

tells us that even after measurements we can find systems that are in superposition states with respect to the measured property. However, experience shows that the outcomes of experiments always draw a definite, non-superposed state for the measured property. That is, unlike in classical physics, the standard formalism of QM seems to indicate that measurements cannot be considered as ordinary physical interactions governed by the dynamics—otherwise no definite outcomes obtain.

Von Neumann recognized this problem in his seminal book, and he proposed the following way out. Postulates 3 and 4 tell us that if we measure a quantity $\mathcal{A}$ we can get as result only the eigenvalues that constitute the spectrum of the operator A, and that the probability to obtain one of those eigenvalues is given by $\text{Prob}^{|\psi\rangle}(a_i) = \langle\psi|P_{\alpha_i}|\psi\rangle = |c_i|^2$. Now, a basic principle of physics is that measurements must be repeatable, and that if after a measurement is made on a certain system, if we immediately repeat such measurement we must find the same result—assuming that no tampering has occurred in the time interval between the measurements and that the natural evolution of the dynamics of the system have not changed its state (this is why we talk about immediate measurements). Von Neumann thus introduced yet another postulate to assure that this principle holds:

> *Postulate 6. Projection postulate*. If the system is in a state $|\psi\rangle \in \mathcal{H}$ and a measurement of the physical quantity $\mathcal{A}$ that corresponds to the operator A, and the outcome of the measurement is the eigenvalue $a_i \in \text{Spec } A$, the system, immediately after the measurement, is in the eigenstate corresponding to a_i. In other words, in a measurement the following transition occurs: $|\psi\rangle \rightsquigarrow \frac{P_{\alpha_i}|\psi\rangle}{\left\|P_{\alpha_i}|\psi\rangle\right\|}$.[54]

If we have a state vector $|\psi\rangle$ which is an eigenstate corresponding to the eigenvalue a_i of A, we know for sure that a measurement of $\mathcal{A}$ will yield the result a_i—and in this case the projection postulate is idle. However, if $|\psi\rangle$ is not an eigenstate of A, but, say, it is defined as $|\psi\rangle = c_i|\alpha_i\rangle + c_j|\alpha_j\rangle$ (we assume again that $\mathcal{H}$ is two-dimensional), and if a measurement draws the result a_i, postulate 6 states that the vector $|\psi\rangle$ has been *projected* onto the eigenvector $|\alpha_i\rangle$—and the denominator $\left\|P_{\alpha_i}|\psi\rangle\right\|$ in the above expression assures that the resulting vector is normalized. In the jargon of QM it is sometimes said that upon a measurement a state vector gets *collapsed* onto an eigenvector. Onto which of the eigenvectors of A the state vector is projected upon a measurement is just a matter of chance, whose probability is given by $\text{Prob}^{|\psi\rangle}(a_i) = \langle\psi|P_{\alpha_i}|\psi\rangle = |c_i|^2$, of course. That is, it is at this point where the indeterministic component which is normally associated to QM enters the picture[55].

The projection postulate was introduced by von Neumann as a *brute force* way out to the measurement problem. Postulate 6 *by fiat* fills the gap between the dynamics in postulate 5 and the postulates 3-4 by plainly telling us that measurements are an exception to the regular dynamics of the theory: when a measurement occurs the state vector of the system is projected onto one of the eigenvectors of the observable corresponding the measured property. Unlike ideal measurements in classical physics, in SQM measurements are intrinsically active interactions which determine the value of the property being measured, and unlike non-ideal measurements classic mechanics, in SQM the way in which measurements

---

[54] After a measurement of the property $\mathcal{A}$, the state of becomes $\rho' = \sum_i P_{\alpha_i}\rho P_{\alpha_i}$. This last expression describes the full measured ensemble, the particular subsystem ρ_i' for which the value a_i has been found, gets described by $\rho_i' = \frac{P_{\alpha_i}\rho P_{\alpha_i}}{\text{Tr}(\rho P_{\alpha_i})}$. This the adaptation of the projection postulate for the case of density matrices.

[55] As D. Albert puts it: 'the effect of measuring an observable must necessarily be to *change* the state vector of the measured system, to "collapse" it, to make it "jump" from whatever it may have been just prior to the measurement onto some eigenvector of the measured observable operator. Which particular such eigenvector it gets changed into is of course determined by the outcome of the measurement; and note that that outcome, in accordance with principle (D) [postulate 4], is a matter of probability. It's at this point then, and at no point other than this one, that an element of pure chance enters into the evolution of the state vector' (Albert 1992, 36).

determine the value of the measured property is not governed by the dynamics of the theory. As David Albert puts it:

> The dynamics and the postulate of collapse are flatly in contradiction with one another […]; and the postulate of collapse seems to be right about what happens when we make measurements, and the dynamics seems to be bizarrely *wrong* about what happens when we make measurements; and yet the dynamics seems to be *right* about what happens whenever we *aren't* making measurements; and so the whole thing is very confusing; and the problem of what to do about all this has come to be called "the problem of measurement" (Albert 1992, 79).

The solution provided by the projection postulate is a matter of brute force in the sense that it is difficult to find some physical rationale for it: what is so special about measurements that they do not respect the linear dynamics determined by the Schrödinger equation? Von Neumann himself and E. Wigner argued that it is the involvement of consciousness what assigns their peculiar nature to measurements. Consider again the after-measurement system $I + II$ in the state $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$. Wigner and von Neumann had it that the physical system is really in that superposed state until a conscious observer 'looks' at it, that is, it is the participation of consciousness what collapses the state $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$ into the state $|\psi_i\rangle|\phi_i\rangle$. This presupposes an ontological mind-matter dualism that we know—ever since Descartes own times—is highly problematic, especially considering the ontological framework of modern science.

Tim Maudlin (1995a) has offered a simple and clear formulation of the problem that grasps what has been said so far. He claims that the following three statements are mutually inconsistent: *i)* the wave function is complete, in the sense that it specifies all the physical properties of a system; *ii)* the wave function always evolves in accord with a linear dynamical equation, namely, Schrödinger's equation; and *iii)* measurements always have definite outcomes (eigenvalues of an observable operator). Statement *i)* corresponds to postulate 1 in the formulation of SQM, whereas statement *ii)* is connected to postulate 5, and statement *iii)* to postulate 3. As we will see, the measurement problem is one of the main motivations for the flourishing of different interpretations of SQM. Maudlin's formulation will be useful when we take a look at them, for some of them can be understood in terms of which of the mutually inconsistent statements they deny or reassess.

That some kind of *interpretation* of SQM that finds a solution for the measurement problem is needed can be seen in that the possibility of taking the easy way out provided by the ignorance interpretation of mixtures—a way out within the bare formalism of the theory—has been proven impossible. First, as d'Espagnat has shown (1999, 171-2), we can describe the measured system I by a density operator $\rho = |\Psi\rangle\langle\Psi|$, and interpret it as an ensemble of identical subsystems that after a measurement interaction with system II becomes a mixture described by $\rho' = \sum_i P_{\alpha_i}\rho P_{\alpha_i}$, where the P_{α_i} are the projectors of the observable A we are measuring[56]. Thus, we may interpret each of the separate terms in the sum in ρ' as corresponding to a pure state that is an eigenstate of A, and which is present in ρ' in a certain proportion p_i. That is, we may say that the measuring interaction between I and II changes the system I into a mixture composed by pure states that are eigenstates of A—so that we do have definite outcomes, but we only have a statistical representation of them (in the classical, ignorance sense). But this maneuver does not work. The total after-measurement system $I + II$ is an entangled pure state, and, accordingly, the density operator ρ' represents an improper mixture to which the ignorance interpretation cannot be ascribed—so we cannot conclude that the terms in the sum in ρ' are eigenstates of A.

A very general proof has been offered by Harvey Brown. This 'insolubility theorem' starts from an assumption that Brown calls 'real unitary evolution' (RUE). Let us consider an ensemble E given by $I +$

[56] As we saw in the previous section, this is the scheme for the transition of a density operator after a measurement occurs. It can be shown that even though $\rho = \rho^2$, $\rho' \neq \rho'^2$—this is a condition for the possibility of considering ρ' as a mixture, of course.

II at time t_0 that is represented by the density operator $W^0 = \sum_n p_n P_{\phi_n}$—we assume that $\sum_n p_n = 1$, that $W^0 \neq (W^0)^2$ and that the pure states $|\phi_n\rangle \in \mathcal{H}^{\mathrm{I}} \otimes \mathcal{H}^{\mathrm{II}}$ are not necessarily mutually orthogonal, that is, we can consider W^0 to be a mixture of pure states $|\phi_n\rangle$ weighted according to p_n. Moreover, we assume that W^0 is the *real* mixture that describes the ensemble E, meaning that we know for sure that the pure states $|\phi_n\rangle$ give an objective, real description of the subsystems in E, i.e., we avoid the ambiguity given by the possibility of different decompositions of W^0 that we discussed above. Under this assumption, it is natural to expect that if every element in E evolves freely during the interval $[t_0, \tau]$ according to the unitary operator U, then the final *real* mixture is given by $W^\tau = \sum_n p_n P_{U\phi_n}$. That is, the final mixture is composed by the pure states $U|\phi_n\rangle$, with corresponding weights p_n. The motivation for this RUE is simply assume that it is possible to express the after-measurement state $I + II$ by means of a density operator that can certainly be given the ignorance interpretation: we can state that E is an ensemble of pure states that really correspond to eigenstates the of the measured observable[57].

Granted RUE, Brown spells out two conditions for the ignorance interpretation way out of the measurement problem to succeed. First, we assume that the measuring system II is set out to measure the observable Q on I, where the spectral resolution of Q in $\mathcal{H}^{\mathrm{I}}$ is given by $Q = \sum_n \lambda_n P_n^Q$, and that the observable A in $\mathcal{H}^{\mathrm{II}}$ corresponding to the "pointer position" of the measuring instrument is given by the spectral decomposition $A = \sum_n \mu_n P_n^A$[58]. Let W_{I}^0 and W_{II}^0 represent the density operators for the systems I and II at time $t = 0$. Now we can define a $\langle Q, A, W_{\mathrm{II}}^0 \rangle$ *measurement* as an interaction ruled by the unitary operator U on $\mathcal{H}^{\mathrm{I}} \otimes \mathcal{H}^{\mathrm{II}}$, under the condition that whenever W_{I}^0 and $W_{\mathrm{I}}^{0\prime}$ are Q-distinguishable, then $U(W_I^0 \otimes W_{II}^0)U^{-1}$ and $U(W_{\mathrm{I}}^{0\prime} \otimes W_{\mathrm{II}}^0)U^{-1}$ are $\mathbf{1} \otimes A$-distinguishable[59]. In simple words, this measurement interaction condition establishes that the value that the pointer of II indicates after the measurement is determined by the value of Q in I before the interaction. The second condition is that the state given by $U(W_I^0 \otimes W_{II}^0)U^{-1}$ can be described by means of a density operator that can be given the ignorance interpretation. That is, the density operator must correspond to a mixture of $\mathbf{1} \otimes A$ eigenstates (definite outcomes for the measurement as expressed by the pointer of the device), and it must represent the *real* mixture of the final states in I + II (in the sense of RUE).

Brown's insolubility theorem goes as follows[60]. Assume that the real mixture in the initial ensemble II is over the pure states $\gamma_n \in \mathcal{H}^{\mathrm{II}}$ with corresponding weights p_n, that is, $W_{\mathrm{II}}^0 = \sum_n p_n P_{\gamma_n}$. Consider that ϕ_1 and ϕ_2 are eigenstates of Q in $\mathcal{H}^{\mathrm{I}}$, and choose three initial I ensembles (pure states) given by $W_{\mathrm{I}}^0 = P_{\phi_1}$, $W_{\mathrm{I}}^{0\prime} = P_{\phi_2}$ and $W_{\mathrm{I}}^{0\prime\prime} = P_{\phi}$, where $\phi = a_1\phi_1 + a_2\phi_2$, with $|a_1|^2 + |a_2|^2 = 1$ and $a_1, a_2 \neq 0$. Notice that these three ensembles are Q-distinguishable and that $W_{\mathrm{I}}^{0\prime\prime}$ is not an eigenstate of Q. Now, given the RUE principle, the after-measurement ensembles $I + II$ in the three cases will be real mixtures described by the following states in $\mathcal{H}^{\mathrm{I}} \otimes \mathcal{H}^{\mathrm{II}}$: $\beta_n = U(\phi_1 \otimes \gamma_n)$; $\beta_n' = U(\phi_2 \otimes \gamma_n)$ and $\beta_n'' = U(\phi \otimes \gamma_n)$—and the weight of the elements in β_n, β_n' and β_n'' is given by the same p_n that derives from the initial II mixture. The linearity of the operator U implies that $\beta_n'' = a_1\beta_n + a_2\beta_n'$. Recall that it is required that the final $I + II$ real mixture must be over eigenstates of $\mathbf{1} \otimes A$, so the γ_n in the three final mixtures must be eigenstates μ_n,

[57] In Brown's own words: 'the principle [RUE] simply rules out the possibility that any of the formally possible resolutions of W^τ, other than the that featuring the states $U\phi_n$, is interpreted as representing the real mixture of states in E at $t = \tau$' (1986, 860). This is a very reasonable premise in an insolubility theorem. The supporter of the ignorance interpretation as a way out of the measurement problem must assume that something that RUE holds, of course. Assuming RUE, Brown is saying that even if we put aside the objections against the ignorance interpretation of mixtures grounded on the ambiguity of the density operators, we can still show that a solution along this line is not possible.

[58] The eigenvectors of Q and A are given by $|\phi_n\rangle$ and $|\xi_n\rangle$, respectively. That is, $P_n^Q = |\phi_n\rangle\langle\phi_n|$ and $P_n^A = |\xi_n\rangle\langle\xi_n|$. The eigenvalues of Q and A are λ_n and μ_n, respectively, as the corresponding spectral decompositions show.

[59] Given two density operators W and W' and an observable $C = \sum_n a_n P_{\alpha_n}$, the states defined by the density operators are said to be C-distinguishable if $\mathrm{Tr}(WP_{\alpha_n}) \neq \mathrm{Tr}(W'P_{\alpha_n})$ for some n.

[60] There is a whole 'tradition' of formal arguments that intend to show that the after measurement system $I + II$ cannot be described by a density operator that represents a mixture of 'pointer-position' eigenstates. The one by Brown that I hereby review is an especially simple one. Brown (1986) comments and evaluates several other proofs.

μ'_n and μ''_n of A[61]. However, $\beta''_n = a_1\beta_n + a_2\beta'_n$ cannot be an eigenstate of $\mathbf{1} \otimes A$ unless $\mu_n = \mu'_n = \mu''_n$, but if this is so, the process described cannot be considered as a $\langle Q, A, W^0_{\mathrm{II}}\rangle$ measurement, for even though the initial ensembles are Q-distinguishable, the final ensembles $I + II$ are not $\mathbf{1} \otimes A$-distinguishable. The conclusion is thus

> *Insolubility theorem*. If U corresponds to a $\langle Q, A, W^0_{\mathrm{II}}\rangle$ measurement defined on the interval $[0, \tau]$, then the final state of the $I + II$ ensemble cannot be described as a real mixture of $\mathbf{1} \otimes A$ eigenstates for all initial ensembles I.

3.3 INTERPRETATIONS OF SQM

As we have seen, the principle of superposition along with the linear dynamics expressed in the Schrödinger equation lead to the measurement problem. Although the projection postulate intends to close the gap between after-measurement states like $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$ and the definite outcomes observed, there is nothing in the standard formalism to endow this postulate with a sound justification[62]. Given this situation, ever since the early days of the theory, physicists and philosophers have attempted to formulate a consistent interpretation of the standard formalism that clarifies the precise conceptual meaning of the theory and that avoids the measurement problem. We can now take a look to the most important interpretations that have been proposed.

3.3.1 Bohr's interpretation

Bohr's view on QM can be understood as involving some Kantian aspects. Bohr argued, mor or less in a Kantian fashion, that in order to have an intelligible representation of nature, knowing subjects must possess a certain conceptual scaffold that assigns coherence and unity to the subject's perception of the real world. The basic assumption in Bohr's interpretation of QM is that such a scaffold corresponds to the concepts of classical physics. The possibility of understanding and communicating in intelligible terms the results of experiments, for example, presupposes that these results can be expressed in terms of classical physics vocabulary. QM certainly violates many of the principles on which classical physics rests, and therefore it cannot be incorporated in the conceptual framework of classical physical theory. Nevertheless, Bohr considered that the experimental results predicted by quantum theory must be described in terms of the classical scaffold:

> It is decisive to recognize that, however far the phenomena transcend the scope of classical physical explanation, the account of all evidence must be expressed in classical terms. The argument is simply that by the word "experiment" we refer to a situation where we can tell others what we have done and what we have learned and that, therefore, the account of the experimental arrangement and of the results of the observations must be expressed in unambiguous language with suitable application of the terminology of classical physics. (Bohr 1958, 39).

[61] If $\mu_n \neq \mu'_n$, then the γ_n in β''_n would be superposed states of A, not eigenstates.

[62] There is the possibility of including the projection postulate as an essential part of the standard formalism of the theory, so that it would not need any justification. However, in this case the problem takes the form of a possible inconsistency. There would be two different forms of dynamical evolution, given by postulates 5 and 6 (projection). Besides, we cannot determine when exactly the wave collapse occurs, so there is no clarity of when and why the sui geris dynamical evolution of the projection postulate holds.

The support that Bohr offered for this position is also rather Kantian. In order to be intelligible, the sensorial data of the knowing subject must be arranged in a spatio-temporal continuum according to causal connections: 'the description of ordinary experience presupposes the unrestricted divisibility of the course of the phenomena in space and time and the linking of all steps in an unbroken chain of cause and effect' (Bohr 1963, 59)[63]. For Kant, though, this conceptual apparatus, which operates as a condition for the possibility of knowledge and experience, is immanent to a 'fixed' human mind and given *a priori*. Bohr, instead, conceived it as the result of human adaptation to the environment, so, in this aspect, he was closer to a naturalization of Kant's transcendental approach *à la* Konrad Lorenz than to Kant's original views[64].

The advent of QM made it clear that at the quantum level classical concepts broke down. In order to save the intelligibility of the natural world that these concepts provided, in spite of the challenge that QM posed to their applicability, Bohr introduced his famous concept of *complementarity*. In a nutshell, by complementarity Bohr understood the relation between kinematical (spatio-temporal) and dynamical (causal) concepts at the quantum level: these two sets of concepts cannot be simultaneously applied to the same physical (quantum) context. However, both are necessary for a complete comprehension of the phenomena. In Bohr's own words, the quantum of action

> forces us to adopt a new mode of description designated as *complementary* in the sense that any given application of classical concepts preclude the simultaneous use of other classical concepts which in a different connection are equally necessary for the elucidation of phenomena. (Bohr 1934, 10)
>
> The very nature of the quantum theory thus forces us to regard the space-time co-ordination and the claim of causality, the union of which characterizes the classical theories, as complementary but exclusive features of the description. (*ibid.*, 19).

Bohr founded the complementarity between kinematical and dynamical attributes in two features of quantum physics: *i)* measurements of properties corresponding to different types (kinematical or dynamical) require mutually exclusive experimental setups; and *ii)* the interaction between measured object and measuring instrument is indeterminable, so that the extrapolation of different measurement results to the complementary context is impossible. The paradigmatic examples of complementary properties and experimental setups are given by measurements of position (kinematic) and momentum (dynamic), on the one hand, and time (kinematic) and energy (dynamic), on the other. In the first case, we have that in order to make a precise position measuring, the measuring apparatus must be rigidly fixed with respect to the spatial reference frame, but this condition is incompatible with the possibility of a precise measurement of momentum. Momentum measurements invoke the law of conservation of momentum, so

[63] As Jane Faye clearly puts it, 'in order to separate the object from the subject itself, the experiential subject must be able to distinguish between the form and the content of his or her experiences. This is possible only if the subject uses causal and spatio-temporal concepts for describing the sensorial content, placing phenomena in causal connection in space and time, since it is the causal space-time description of our perceptions that constitutes the criterion of reality for them. Bohr therefore believed that what gives us the possibility of talking about an object and an objectively existing reality is the application of those necessary concepts, and that the physical equivalents of "space", "time", "causation", and "continuity" were the concepts "position", "time", "momentum", and "energy", which he referred to as *the classical concepts*. He also believed that the above basic concepts exist already as preconditions of unambiguous and meaningful communication, built in as rules of our ordinary language. So, in Bohr's opinion the conditions for an objective description of nature given by the concepts of classical concepts were merely a refinement of the preconditions of human knowledge' (Faye 2009, 6-7).

[64] As Faye interestingly points out, Bohr's commitment to the necessity of classical concepts resulted in that his mature comprehension of the correspondence principle became more philosophical and fundamental. Its original usefulness was heuristic in the development of the old theory. Once Bohr formulated his interpretation of the mature theory, 'the correspondence rule was based on the epistemological idea that classical concepts were indispensable for our understanding of physical reality, and it is only when classical phenomena and quantum phenomena are described in terms of the same classical concepts that we can compare different physical experiences. It was this broader sense of the correspondence rule that Bohr often had in mind later on' (Faye 2009, 10).

that it is necessary to know the momentum of the measuring device before and after the interaction with the measured object. But these can be known only if the instrument is not rigidly fixed to the spatial frame. If we allow the instrument to be loosely attached to the spatial frame, then a precise position measurement is not possible. This is an instance of statement *i)*[65].

Regarding statement *ii)*, we have that since measurements involve a physical interaction between instrument and object, there is usually an exchange of momentum and energy between them. In classical physics the value of the exchange can be made small enough as to be negligible, or at least determinable and controllable—so that its effects on the measurement can be precisely determined. However, on the quantum level, the values of the energy and momentum interchanges involved in the measuring interaction cannot be ignored because they are of the same order of the quantities that are being measured, and they cannot be made negligibly small given the minimum quantum of action expressed by h. Moreover, this non-negligible energy-momentum interaction cannot be precisely determined. As we just saw, in a position measurement, the energy and momentum that is gained by the measuring instrument cannot be determined given its rigid fixedness to the spatial reference frame[66].

One might think that the measurement energy-momentum interaction could be determined, for example, by measuring the momentum-energy of the instrument before and after the interaction and then to apply the conservation law—in that way, the measuring interaction could be corrected, just as in classical physics. But let us suppose that the instrument is a position-measuring device, then from what was said just above we quickly notice that in order to know its momentum or energy with precision the device could not work as an effective position-measuring instrument. The important point is that if we decide to apply a maneuver like this in order to determine the measurement interaction, then we have to consider the instrument not as an instrument, but as an object in the measuring process, and subject to quantum mechanics. But then, due to complementarity, we could not know in this case its dynamic and kinematic properties simultaneously. That is, the maneuver at issue pushes the problem one stage back, and it is clear that we could go on *ad infinitum* in our quest for the precise value of the measurement interaction. These remarks point out another essential feature of Bohr's view of QM: the *wholeness* of the conditions of observation. The indeterminacy of the measurement dynamical interaction entails that the object and the instrument cannot be considered as two separate entities in a context of measurement. In classical mechanics, the determined properties of a closed system can be measured. As we mentioned, the measuring interaction is negligible or controllable. In the quantum case, a system remains 'really' closed a long as it is not measured. The fact that the measuring interaction cannot be dodged entails that what we can consider as a closed system is the instrument *plus* the object, and any attempt to decompose or analyze this system in order to determine the measurement interaction leads to a new complementarity situation and to an alteration of the original phenomenon that was to be measured: 'the essential wholeness of a proper quantum phenomenon find indeed logical expression in the circumstance that any attempt at its

[65] In the case of energy-time complementarity: 'measurement of time requires a clock which is precisely synchronized with the process defining the temporal reference frame, and which is constructed in such a way that it is not appreciably affected by the process or event being timed or by the reading of the time registered. This being the case, the clock is incapable of functioning as an energy-measuring instrument: the very imperturbability of the time-measuring device renders it unsuitable for measuring energy, since an energy-measuring device *eo ipso* must be perturbable' (Murdoch 1987, 84).

[66] Bohr saw statements *i)* and *ii)* and the corresponding kinematic-dynamic complementarity, as reflected in the non-commutation relation $\mathbf{pq} - \mathbf{qp} = (\hbar/i)\mathbf{1}$, for this formula expresses that the result of a momentum measurement followed by a position measurement is different from the same measurements performed in the reverse order. That is, complementarity can be taken as the epistemological counterpart of the non-commutativity of observables and the associated uncertainty relations in the formalism of SQM. Bohr's understanding of complementarity was also ontological, in the sense that it is not only that we cannot know or measure complementary properties simultaneously, but that physical object cannot possess complementary properties simultaneously (see Murdoch 1987, chapter 7).

well-defined subdivision would require a change in the experimental arrangement incompatible with the appearance of the phenomenon itself' (Bohr 1958, 72).

The wholeness of measurement processes poses a challenge to the classic concept of observation and measurement, for in the classical picture a clear separation between measuring instrument and object observed is required. Since this distinction is a condition for the possibility of the expression of measurement outcomes in classical terms, Bohr concludes that the classical notion of observation is an idealization which is possible only under conditions in which the quantum of action can be made negligible. In the case of quantum observation, thus, in order to express the results in the classical language—which, as we saw, Bohr considered the only way to encompass the description of nature in an intelligible and communicable framework—a *cut* between instrument and object must be done. In the discussion about the indeterminateness of the measurement interaction we saw that such a cut can in principle lead to an infinite regress. Therefore, although the place of the cut that is required to frame quantum measurements within the classical picture is, in principle, arbitrary, each experimental context can suggest a suitable place. Usually, the microscopic-macroscopic dichotomy can be taken as the criterion to perform the cut, but not all experimental contexts can be treated in this way. Therefore, the issue of the distinction between instrument and object is one of the most problematic features of Bohr's view.

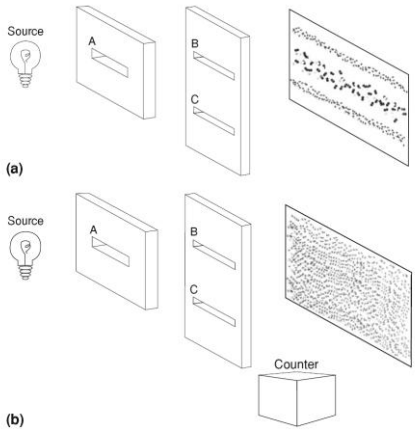
One last essential component of Bohr's view is his understanding of the meaning of the wave function as applied to systems in the absence of measurements. As Jan Faye explains, Bohr understood the formalism to represent an undisturbed state in a mere *symbolic* way. That is, Bohr would have accepted postulate 1 with the important proviso that the state vector does not really *describe* the state of a system in any sense, but it only codifies the statistical information corresponding to the performance of eventual measurements on the system. According to Bohr,

> the quantum mechanical formalism does not provide physicists with a 'pictorial' representation: the ψ-function does not, as Schrödinger had hoped, represent a new kind of reality. Instead, as Born suggested, the square of the absolute value of the ψ-function expresses a probability amplitude for the outcome of a measurement. Due to the fact that the wave equation involves an imaginary quantity this equation can only have a symbolic character, but the formalism may be used to predict the outcome of a measurement that establishes conditions under which concepts like position, momentum, time and energy apply to the phenomena. (Faye 2009, 14)[67]

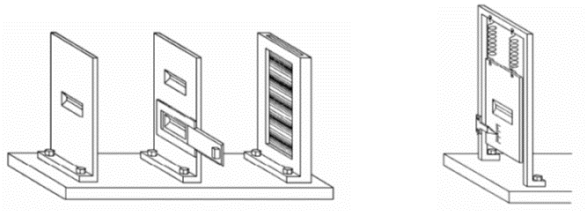
Bohr's views can be illustrated by a paradigmatic experiment of quantum physics, the double-slit experiment. The basic arrangement of this famous example is given by a source of monochromatic light or particles, a first diaphragm with one single slit at a certain distance of the source, a second diaphragm with two slits at a certain distance of the first one, and a detection screen. A large number of photons or electrons are emitted from the source and made to pass across the slits in the two diaphragms, so that they can be detected at the screen. Two different setups for the experiment can be considered. In the first one, the two slits in the second diaphragm are open, so that we do not know through which of them the particle passes on its way to the screen. In the second setup, some method to detect the path of the particle, such as a counter, is implemented. The interesting point is that the results obtained in each setup are different. In the first one a pattern of interference is observed in the screen, due to the interference of the wave functions that come out of the two slits in diaphragm 2, a pattern which is independent of the time

[67] One must be careful here. This symbolic interpretation of the wave function might sound as an expression of an instrumentailst epistemology. Heisenberg seems to have understood QM along a line like that. However, in the case of Bohr, the symbolic interpretation is coherent with his Kantian framework. In the absence of measurements, the wave function cannot be considered as descriptive because in that context *classical concepts completely breakdown*, not because what is *beyond observation is meaningless*—as an instrumentalist or positivist would claim. Bohr would not have been puzzled by the principle of superposition, for example, for wave functions corresponding to a state like that do not really describe an object or entity. To say that a system is in a superposed state is just a symbol to express the statistics of results in eventual experiments.

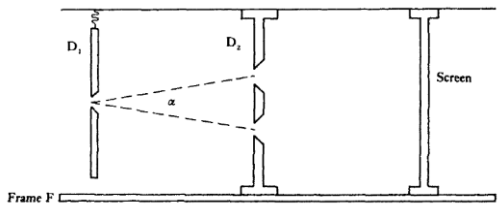
between the emissions and the distance between the slits—it is assumed that photons or electrons can be emitted one by one, in time intervals large enough so that each emission can take place after the previous photon or particle has reached the screen. In the second setup no interference pattern is observed:



That the decision of detecting the path of the particle results in a different outcome illustrate Bohr's considerations. First, the two setups are an example of complementary experimental contexts, related to the measurement of complementary properties. Suppose that the experiment in the first setup is being considered and that we want to know through which slit in the second diaphragm the particle goes through. This could be done, for example, by measuring the momentum interchange between the particle and diaphragm 1, if the latter suffers a jolt upwards, then the particle goes through the lowest slit in the second diaphragm, and *vice versa*. In order to perform this momentum measurement with the precision needed, the first diaphragm cannot be rigidly fixed, of course. However, if we know the momentum of the first diaphragm before and after the interaction, then our knowledge of its position becomes uncertain according to Heisenberg's formula. This uncertainty, in turn, entails an equal uncertainty in the positions of the fringes in the detection screen, so that the interference pattern cannot be observed[68]. Therefore, the measurement of the interference pattern and the detection of the slit through which the particles go require mutually exclusive experimental setups:



[68] The momentum transferred to the first diaphragm would be different depending on which of the two slits in the second diaphragm the particle passed through. If α is the angle between the paths through the slits, the difference is given by $hk\alpha$, wher k is the wave number (number of cycles per unit of length). An measurement of the momentum of the first diaphragm accurate enough would entail an uncertainty in its position of the order $\frac{1}{k\alpha}$, but un uncertainty of this order would imply an uncertainty in the position of the fringes on screen, for the number of fringes per unit length is given by $k\alpha$, so that the interference pattern would be lost:



This example also illustrates the wholeness of the experimental setup. Just as we saw above, any attempt in order to determine the value of the momentum-energy interaction between object and instrument results in a change of the experimental arrangement that is incompatible with the appearance of the phenomenon itself. In this case, to determine the value of the momentum interaction between the particle and the first diaphragm blurs the interference phenomenon.

Finally, a comment about Bohr's interpretation of the meaning of the wave function is in order. One might be tempted to think that, in the interference setup, the particle somehow passes through both slits, so that the two separate components of the wave function of the particle that come out of the second diaphragm overlap and produce the interference pattern. Bohr would not accept this interpretation. To say that a quantum particle, in the absence of positions measurements, behaves like a wave in its flight to the detecting screen suggests an understanding of the wave function in a literal way. However, the wave-like behavior of the particles is only *collectively* reflected in the measured interference pattern on the screen[69].

So far, so good; but what about Bohr's interpretation and the measurement problem? It is difficult to say, since he never directly addressed the issue in its canonical formulation. Murdoch argues that if we consider the wholeness of the measurement process, we can conjecture that he would have responded that the problem arises because the measuring instrument is being treated as an object—in the measurement problem the whole process is described in quantum mechanical terms—but this means that we cannot properly consider the process as a measurement, for in a measurement the distinction between instrument and object gets blurred. That is, to describe the process by means of a unitary transformation involving both the object and device means that we are not really talking about a measurement[70]. However, as Murdoch points out, even if this way out works, another puzzle arises: 'how is that the pointer

[69] Allow me to quote at length this illuminating passage from Plotnisky's treatment of this issue: 'First of all, we do not appear to be able—nobody has ever accomplished this thus far—to observe, through any instrument, the independent behavior, say, motion, (particle-like or wave-like), of quantum objects. We can only observe certain trace-like effects of this behavior manifested in these instruments, and we infere the existence of quantum objects from these effects […]. These effects, such as a trace on a silver bromide screen or a click in a detector, define, in Bohr's terms, *individual* quantum phenomena. These phenomena are always particle-*like* insofar as such individual traces, form contained, point-*like*, individual entities, always discrete relative to each other.

Accordingly, the statement "a photon passed through a slit" only means that a measuring device registered an event that is *analogous* to a certain classical physical event, say, that of the hitting of a screen by a small classical object that passed through an opening in some diaphragm on its way. The statement "a photon never passes through both slits" means that no events corresponding to such a statement can be observed or registered. We can never register an individual event simultaneously linked to both slits, say, by placing a detector near each slit. Only one of these detectors registers each individual event: the two detectors never click simultaneously […]. On the other hand, one could speak of a single photon as "passing through a single slit" in the sense that the corresponding event could be registered by a "which-path" measuring device, but only in this sense […].

In sum, either type of characterization—particle-*like* (which can be both individual and collective) and wave-*like* (which is only collective)—only relates to the behavior of quantum objects as concerns the effects of this behavior on measuring instruments, or phenomena in Bohr's sense, since we observe nothing else. Neither concept—that of "wave" or that of "particle"—applies as a physical concept to quantum objects and their behavior themselves. The individual phenomenal *effects* in question in the double-slit experiment and other quantum experiments may be seen as *particle-like* insofar as they are *similar* to the kind of traces classical particle-like objects colliding with the screen as well […]. In Bohr's ultimate interpretation of the situation, we cannot […] apply to quantum objects classical physical concepts associated with these properties [exact position, exact momentum, trajectory], such as those of motion, or even use words such as "happens" or "occurs". As Both Bohr and Heisenberg argue, such words can only apply at the level of observation, and not to what happens before an observation or between observations and hence not to quantum objects themselves' (Plotnisky 2013, 81-3).

[70] 'He would, I suggest, have responded along the following lines. The problem arises because the measurement process is described in terms of a unitary transformation on the state of the compound object, $O + M$. But to treat the process in this way is to elide the distinction between the object and the instrument: it is to treat the instrument as an object, an object which is an integral part of a larger object. Treating the instrument as an object, however, precludes our treating the interaction between the object and the instrument as a process of *measurement*. The measurement problem thus arises as a consequence of our treating the instrument as an object. But even if we confined our attention to the object alone, the measurement process would not be fully describable in quantum-mechanical terms. We might treat the evolution of the

of a macroscopic instrument may display a definite position, say, when the instrument is described in classical physical terms […], yet have no definite position when we treat it as a quantum-mechanical object' (1987, 114). That is, even if we accept that Bohr's interpretation avoids the measurement problem by negating that a measurement can be described in strict quantum mechanical terms, the equally serious problem of the connection between the quantum and classical description arises: how is it that our classical world emerges from the quantum world? Murdoch suggests that this question could be answered in a Bohrian fashion:

> The solution to the puzzle lies in Bohr's thesis that the ultimate instrument of measurement, i.e., the apparatus which the observer looks at or listens to, must be a comparatively massive macroscopic object which is describable in everyday physical terms […], the comparative massiveness of the instrument ensures that a definite result is observable. Why does it? Simply because the massiveness of the instrument guarantees that the 'interference' terms contained in the superposition state may for all practical purposes be ignored, since they have no observable effects: the superposition state of a sufficiently massive object is practically indistinguishable from a mixed state. (ibid., 114-5)

That the washing out of the interference terms allows us to consider the post-measurement description of the object-instrument system as a mixed state is usually understood as the result of *decoherence*. Therefore, this physical feature might then provide a solution of the measurement problem and work as the base of a consistent interpretation of QM along Bohr's line of thought. We can now consider this possibility.

3.3.2 Decoherence and consistent histories

To understand the essentials of decoherence it is useful to take a look at the quantum description of a measurement once again. We saw above that if before the measurement system I is in a superposed state $\sum_i c_i|\psi_i\rangle$ and system II is in the state $|\phi_0\rangle$, the after-measurement total system is given by $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$. For simplicity, let us assume that system I is given by the two-dimensional ket $|\Psi\rangle = c_m|m\rangle + c_n|n\rangle$, and that the eigenvectors $|\phi_m\rangle$ and $|\phi_n\rangle$ represent the 'pointer positions' of system II, so that the after-measurement system $I + II$ is given by $|\Theta\rangle_{\Psi\Phi} = c_m|m\rangle|\phi_m\rangle + c_n|n\rangle|\phi_n\rangle$. Thus, the corresponding density operator is $\rho_{\Psi\Phi} = |c_m|^2|m\rangle\langle m||\phi_m\rangle\langle\phi_m| + c_m c_n^*|m\rangle\langle n||\phi_m\rangle\langle\phi_n| + c_m^* c_n|n\rangle\langle m||\phi_n\rangle\langle\phi_m| + |c_n|^2|n\rangle\langle n||\phi_n\rangle\langle\phi_n|$, where the second and third terms in the sum correspond to the 'interference' terms expressing the entangled correlations between I and II. We have thus another aspect in the measurement problem. The final state is not only a superposition of definite outcomes, but the probabilities of possible results include 'interference effects'. These effects are observable in some cases, like the double-slit experiment—where the interference terms are responsible for the interference pattern, of course—but in most of the experiments in which macroscopic devices are involved these effects are never observed[71].

---

state of the object alone by means of an operation describing a perturbation on the Hamiltonian operator associated with the energy of the object, but only if we had knowledge of the perturbation, which of course we do not have, and cannot have if the measurement is to be made. Thus it is not surprising that the process of measurement cannot be properly described in terms of a unitary transformation. If we treat the instrument as an object, and describe the object-instrument interaction in quantum-mechanical terms, then measurement *is* impossible' (Murdoch 1987, 113).

[71] More precisely, the interference terms have observable effects only if we perform a measurement of an observable of the composite system I + II that does not commute with the observables whose eigenvectors constitute the basis in which the subsystems I and II are expressed. In our example, system I is expressed in the basis $\{|m\rangle, |n\rangle\}$ corresponding to the spectrum of observable $O \in \mathcal{H}^{\mathrm{I}}$, and system II is expressed in the basis $\{|\phi_m\rangle, |\phi_n\rangle\}$ corresponding to the spectrum of observable $M \in \mathcal{H}^{\mathrm{II}}$. The interference terms would have observable effects if an observable of the form $V_{\mathrm{I}} \otimes V_{\mathrm{II}} \in \mathcal{H}^{\mathrm{I}} \otimes \mathcal{H}^{\mathrm{II}}$, where V_{I} does not commute with O and V_{II} does not commute with M, is measured—otherwise the mutual orthogonality between $|m\rangle$ and $|n\rangle$ and between $|\phi_m\rangle$ and $|\phi_n\rangle$ cancels the interference terms. That is, the state of the system I + II, interference terms included, is an eigenstate of a suitable observable of the form $V_{\mathrm{I}} \otimes V_{\mathrm{II}}$. See (Bub 1997, section 8.1)

Let us now consider the environment E of the experiment as well, which we label with the ket $|e\rangle$. Assuming that before the experiment the environment state is given by $|e_0\rangle$, then the final state of $I + II + E$ is thus $|\Theta\rangle_{\Psi\Phi E} = c_m|m\rangle|\phi_m\rangle|e_m\rangle + c_n|n\rangle|\phi_n\rangle|e_n\rangle$. The physical effect known as decoherence consists simply in that the interaction between the systems I and II and the environment E—which is of course described by an enormous number of degrees of freedom given by the $|e_i\rangle$—implies that the $|e_i\rangle$ quickly approach mutual orthogonality; that is, $\langle e_m|e_n\rangle(t) \to \delta_{mn}$. This in turn means that if we take the density matrix $\rho_{\Psi\Phi E}$ corresponding to $|\Theta\rangle_{\Psi\Phi E}$, the fact that $\langle e_m|e_n\rangle = 0$ entails that the interference terms vanish. Therefore, we can take the *reduced density matrix* corresponding to $|\Theta\rangle_{\Psi\Phi}$, which is given by $\rho^r_{\Psi\Phi} = |c_m|^2|m\rangle\langle m||\phi_m\rangle\langle\phi_m| + |c_n|^2|n\rangle\langle n||\phi_n\rangle\langle\phi_n|$, so that the interference effects, at least in principle, no longer bother us[72]. Notice also that $\rho^r_{\Psi\Phi}$ is formally identical to a density matrix describing a mixture composed by the pure states $|m\rangle|\phi_m\rangle$ and $|n\rangle|\phi_n\rangle$, which suggests that a solution of the measurement problem is in sight.

The decoherence effect plays an essential role in an interpretational scheme known as the *consistent histories interpretation*, which is due to R. B. Griffiths, R. Omnès, M. Gell-Mann and J. B. Hartle. A history is a sequence of events corresponding to a system at different times, and is defined by a set of temporally successive projectors that evolve according to the unitary operator U, that is, $h = \{P^1_{\alpha_i}(t_1), P^2_{\alpha_i}(t_2), \dots, P^n_{\alpha_i}(t_n)\}$. For each t_i in the sequence an exhaustive set of mutually exclusive projectors $\mathcal{P}^i$ is defined such that in the history h, $P_i(t_i) \in \mathcal{P}^i$, $\sum_{\alpha_i} P^i_{\alpha_i}(t_i) = 1$ and $P^i_{\alpha_i}(t_i)P^i_{\beta_i}(t_i) = \delta_{\alpha_i\beta_i}P^i_{\alpha_i}(t_i)$. That is, a history h is given by choosing a projector for each t_i from the corresponding set $\mathcal{P}^i$. Histories which, like h, are specified by a set of 1-dimensional projectors onto eigenstates of a complete set of observables at all times in the interval are called *maximally grained histories*, whereas histories given by projectors which are sums of the projectors in a maximally grained history are called *coarse grained histories*. For example, if we take the maximally grained histories $h_1 = \{P^1_{\alpha_i}(t_1), P(t_2)\}$ and $h_2 = \{P^1_{\beta_i}(t_1), P(t_2)\}$, then $h' = \{P^1_{\alpha_i}(t_1) + P^1_{\beta_i}(t_1), P(t_2)\}$ is a coarse grained history. The natural interpretation of a history is that the system at issue is, at each instant t_i, in the state defined by the corresponding projector—and the set of all possible histories for a system within a time interval can be thus defined.

The probability of a history $h = \{P^1_{\alpha_i}(t_1), P^2_{\alpha_i}(t_2), \dots, P^n_{\alpha_i}(t_n)\}$ for a system described by a density matrix ρ, can be given, in principle, by a repeated application of the Born rule for the probability of individual events $\mathrm{Tr}(\rho P_{\alpha_i})$, and considering the time evolution of density matrices, that is, by $p_h = \mathrm{Tr}\left[P^n_{\alpha_i}(t_n) \dots P^1_{\alpha_i}(t_1)\rho P^1_{\alpha_i}(t_1) \dots P^n_{\alpha_i}(t_n)\right]$. A natural requirement for this expression for the probability of histories to be sound is that it must respect additivity, that is, that the probability of a coarse grained history that contains a combined projection operator $P^1_{\alpha_i}(t_1) + P^1_{\beta_i}(t_1)$ must be calculable from the sum of the two maximally grained histories containing each individual projector. However, if we consider the case of h_1, h_2 and h' we can readily see that

$$
\begin{aligned}
p_{h'} &= \mathrm{Tr}\left(P(t_2)\left[P^1_{\alpha_i}(t_1) + P^1_{\beta_i}(t_1)\right]\rho\left[P^1_{\alpha_i}(t_1) + P^1_{\beta_i}(t_1)\right]P(t_2)\right) \\
&= \mathrm{Tr}\left(P(t_2)P^1_{\alpha_i}(t_1)\rho P^1_{\alpha_i}(t_1)P(t_2)\right) + \mathrm{Tr}\left(P(t_2)P^1_{\beta_i}(t_1)\rho P^1_{\beta_i}(t_1)P(t_2)\right) \\
&+ \mathrm{Tr}\left(P(t_2)P^1_{\alpha_i}(t_1)\rho P^1_{\beta_i}(t_1)P(t_2)\right) + \mathrm{Tr}\left(P(t_2)P^1_{\beta_i}(t_1)\rho P^1_{\alpha_i}(t_1)P(t_2)\right) \\
&= p_{h_1} + p_{h_2} + \mathrm{Tr}\left(P(t_2)P^1_{\alpha_i}(t_1)\rho P^1_{\beta_i}(t_1)P(t_2)\right) + \mathrm{Tr}\left(P(t_2)P^1_{\beta_i}(t_1)\rho P^1_{\alpha_i}(t_1)P(t_2)\right)
\end{aligned}
$$

[72] That the $|e_i\rangle$ approach mutual orthogonality is grounded in the fact that the reduced density matrix of the system evolves according to the *master equation* $\frac{\partial\rho(x',x,t)}{\partial t} = \frac{1}{i\hbar}\langle x'|[H,\rho(t)]|x\rangle - \frac{\gamma}{2}(x'-x)\left(\frac{\partial}{\partial x'} - \frac{\partial}{\partial x}\right)\rho(x',x,t) - \frac{\gamma mkT}{\hbar^2}(x'-x)^2\rho(x',x,t)$, where H is the Hamiltonian of the system, γ is the 'dissipation' or 'relaxation' coefficient, k is Boltzmann's constant, and T is the temperature of the environment. As Zurek (2003,12) points out, the first term can be derived from the Schrödinger equation, the second describes dissipation, and the third is responsible for the decoherence effect. For concrete examples, see (Zurek 2002) and (Schlosshauer 2004).

The terms $\mathrm{Tr}\left(P(t_2)P^1_{\alpha_i}(t_1)\rho P^1_{\beta_i}(t_1)P(t_2)\right)$ and $\mathrm{Tr}\left(P(t_2)P^1_{\beta_i}(t_1)\rho P^1_{\alpha_i}(t_1)P(t_2)\right)$, which are responsible for the additivity violation, are the usual interference terms. Therefore, a condition to define the probability of a history in a way such that additivity is respected is that these terms should vanish, that is, that

$$\mathrm{Re}\{\mathrm{Tr}[P^n_{\alpha_i}(t_n)\dots P^i_{\alpha_i}(t_i)\dots P^1_{\alpha_i}(t_1)\rho P^1_{\alpha_i}(t_1)\dots P^i_{\beta_i}(t_i)\dots P^n_{\alpha_i}(t_n)]\} = 0 \qquad \text{if } \alpha \neq \beta$$

Since this probability consistency requirement means that interference terms must vanish, it is rather natural that the decoherence effect plays an essential role in this interpretative approach. If we define a *decoherence functional* $D(\alpha,\beta) = \mathrm{Tr}[P^n_{\alpha_i}(t_n)\dots P^1_{\alpha_i}(t_1)\rho P^1_{\beta_i}(t_1)\dots P^n_{\beta_i}(t_n)]$ then the requirement for the consistency of probabilities of histories becomes $\mathrm{Re}[D(\alpha,\beta)] = \delta_{\alpha\beta}D(\alpha,\alpha)$. The *consistent histories* that fulfill the formal consistency requirement satisfy it, from a physical point of view, because of the decoherence effect: environment-induced decoherence entails that the corresponding decoherence functional quickly approaches zero.

We saw above that the set of all possible histories for a system during a time interval can be defined. Now we can also define the set of all possible consistent histories given by the fulfillment of the requirement just mentioned. Since this set can be understood as a set of histories in which interference effects do not happen, then, at least in principle, consistent histories are the possible (quasi)classical histories of the evolution of a system. That is, decoherent-consistent histories seem to provide us with a picture of how classicality emerges from the quantum world in a precise way—through decoherence Bohr's epistemic argument about the primacy of classical concepts and the cut between object and instrument as the ground for the emergence of the classical picture may be replaced by a *physical process* that produces this emergence[73]. Besides, given their mutual exclusiveness, consistent histories can be used to extend the conceptual reach of Bohr's notion of complementarity. We saw above that this concept originally referred to experimental setups and measurement outcomes, but it can also hold for histories of systems even in the absence of observations. As Jeffrey Bub clearly explains,

> The concept of consistent histories provides an elegant way of conforming to the restrictions of the Copenhagen interpretation concerning what we can say about quantum systems in various experimental situations, without invoking the primacy of classical concepts, or a 'cut' between the observer and what is being observed. In fact, the appeal of the consistent histories approach to quantum mechanics is just that it purports to provide the basis for an 'observer-free' interpretation that can be applied to the universe as a closed system. (Bub 1997, 234)[74]

We can evaluate the fruitfulness of the consistent histories approach by examining it in connection with the measurement problem and the border and link between the quantum and classical worlds it

[73] As Bacciagaluppi points out: 'if we understand the theory of decoherence as pointing to how classical concepts might in fact emerge from quantum mechanics, this seems to undermine Bohr's basic position. Of course it would be a mistake to say that decoherence (a part of quantum theory) *contradicts* the Copenhagen approach (an interpretation of quantum theory). However, decoherence does suggest that one might want to adopt alternative interpretations, in which it is the quantum concepts that are prior to the classical ones, or, more precisely, that classical concepts at the everyday level emerge from quantum mechanics […].
On the other hand, Bohr's *intuition* that quantum mechanics as practiced requires a classical domain would in fact be *confirmed* by decoherence, if it turns out that decoherence is indeed the basis for the phenomenology of quantum mechanics' (2012, 30-1).

[74] That the consistent histories approach gives a purely physical underpinning to Bohr's ideas—no measurements and observers are needed to determine any cut or experimental setups—is reinforced by the following remark by J. J. Halliwell on the meaning of the probability of histories: 'This expression $[p_h]$ is a familiar one from quantum measurement theory, but the interpretation is different. Here, it is the probability for a sequence of alternatives for a closed system. The alternatives at each moment are characterized by projectors. The projectors are not generally associated with measurements, as they would be in the Copenhagen view of the formula $[p_h = \mathrm{Tr}[P^n_{\alpha_i}(t_n)\dots P^1_{\alpha_i}(t_1)\rho P^1_{\alpha_i}(t_1)\dots P^n_{\alpha_i}(t_n)]]$. They cannot because the system is closed' (Halliwell 1995, 729).

seems to draw. First, we saw above that the fact that decoherence allows us to take a reduced density matrix which is formally identical to the density matrix corresponding to a mixture of states corresponding to eigenstates of the observable measured in system I and eigenstates of the 'pointer position' observable in system II, could be taken as a solution of the measurement problem. However, this view would clearly be wrong. To say it in d'Espagnat's terms, the reduced density matrix is an improper mixture obtained from partially tracing out the entangled state of the system $I + II + E$, so it cannot be interpreted as describing a mixed state. Even though the fact that environment induced decoherence allows us to dodge the interference terms for most of practical purposes—only correlated and very complicated measurements display the interference effects—the reduced matrix partially describes a system that is in a superposed state, not a mixture[75]. Thus, we are left with a superposition of possible (quasi)classical outcomes—in the sense that interference effects have been suppressed for most of practical purposes—but neither decoherence by itself nor the consistent histories approach explains why only one particular outcome is observed in measurements[76]. Actually, there is a sense in which decoherence even exacerbates the problem, since the interaction with the environment entails that the post-measurement state vector gets even more entangled:

> Intuitively, if the environment is carrying out, without our intervention, lots of approximate position measurements, then the measurement problem ought to apply more widely, also to these spontaneously occurring measurements […]. The state of the object and the environment could be a superposition of zillions of very well localized terms, each with slightly different positions, and that are collectively spread over a *macroscopic distance*, even in the case of everyday objects […]. If everything is in interaction with everything else, everything is generically entangled with everything else, and that is a worse problem than measuring apparatuses being entangled with measurement systems. (Bacciagaluppi 2012, 14)[77]

Secondly, the simple picture of the emergence of classicality that decoherence and consistent histories in principle provide is not drawn in all its details. Bacciagaluppi (2012, section 4) states that whether or not decoherence can fully and adequately explain such emergence is not yet clear. The answer depends on how far the application of decoherence can be pushed—in the sense that it must be determined if decoherence applies in all physical contexts (quantum chaos, gravity, etc.), and if it applies always in a way such that the emergence of classicality results. Besides, Schlosshauer remarks that the consistency

---

[75] Another way to pinpoint the failure of the 'mixture way out' is that the ignorance interpretation of a reduced matrix is incompatible with the unitary evolution of the system from its original state: 'If an ensemble interpretation could be attached to a superposition, the latter would simply represent an ensemble of more fundamentally determined states, and based on the additional knowledge brought about by the results of measurements, we could simply choose a subensemble consisting of the definite pointer state obtained in the measurement. But then, since the time evolution has been strictly deterministic according to the Schrödinger equation, we could backtrack this subensemble in time and thus also specify the initial state more completely ("postselection"), and therefore this state necessarily could not be physically identical to the initially prepared state' (Schlosshauer 2004, 1270).

[76] Many decoherentists and supporters of the consistent history approach certainly acknowledge that no solution of the measurement problem has been achieved, but there is still a wide-spread otherwise belief. S. Adler (2003) and M. Schlosshauer (2004) provide many references to authors that endorse this belief.

[77] Put differently, the tracing out of the interference terms that decoherence allows only shifts the problem one stage further. If we assume that the state of the environment is written in the basis of the spectrum of an observable E given by the set of eigenvectors $\{|e_i\rangle\}$, we can define a suitable operator B of the form $V_{\mathrm{I}} \otimes V_{\mathrm{II}} \otimes V_E \in \mathcal{H}^{\mathrm{I}} \otimes \mathcal{H}^{\mathrm{II}} \otimes \mathcal{H}^{E}$—where V_{I} does not commute with O, V_{II} does not commute with A, and V_E does not commute with E—such that the state given by the non-reduced density operator

$$\rho_{\Psi\Phi E} = |c_m|^2|m\rangle\langle m||\phi_m\rangle\langle\phi_m||e_m\rangle\langle e_m| + c_m c_n^*|m\rangle\langle n||\phi_m\rangle\langle\phi_n||e_m\rangle\langle e_n| + c_m^* c_n|n\rangle\langle m||\phi_n\rangle\langle\phi_m||e_n\rangle\langle e_m| + |c_n|^2|n\rangle\langle n||\phi_n\rangle\langle\phi_n||e_n\rangle\langle e_n|$$

is an eigenstate of B, and therefore a measurement of B on $\rho_{\Psi\Phi E}$ would yield observable interference effects. To instantiate such a measurement would be virtually impossible, so that sometimes the view that decoherence solves the measurement problem *for all practical purposes* is maintained. However, it is clear that such a view presupposes that the reduced density matrix can be interpreted as describing a mixed state—presupposition which is completely unjustified, as we just saw.

requirement is neither a sufficient nor necessary condition for the (quasi)classicality of consistent histories. Accordingly, several authors have attempted to introduce additional criteria to select (quasi)classicality, but 'this approach intrinsically requires the notion of local, open systems and the split of the universe into subsystems, in contrast to the original aim of the consistent-histories approach to describe the evolution of a single closed, undivided system (typically the entire universe)' (Schlosshauer 2004, 1300). One might then conclude that the notion of a Bohrian 'cut' is still in some sense required to have an explanation for the emergence of classicality.

Summarizing, it is rather clear that in spite of providing a more formal, elaborated and precise framework for a Bohr-like interpretation of QM, the consistent histories interpretation armed with decoherence is not able to offer a solution for the measurement problem (neither is decoherence by itself), and, *a fortiori*, the solution that according to Murdoch is envisioned in Bohr's interpretation has proved unattainable. On the other hand, the explanation that the consistent histories interpretation proposes for the emergence of classicality is not completely spelled out, and the reference to open systems somewhat undermines the achievement of its original goal.

3.3.4 Everett's and Everettian interpretations

As we saw above, the measurement problem consists in that according to postulate 5, the unitary evolution of a compound system I + II given by $(\sum_i c_i|\psi_i\rangle)|\phi_0\rangle$ yields, after a measurement interaction, the superposed state $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$. States like this are not normally observed in measurement outcomes, thus the necessity of the projection postulate to obtain the state $|\psi_i\rangle|\phi_i\rangle$ with a probability given by $|c_i|^2$. The projection postulate is not an integrated part of the dynamics of the theory, so it is natural that either its suppression or a physical explanation for it should be included in a coherent interpretation of QM. Hugh Everett III proposed in 1957 that the first option can be taken, that is, to state that the dynamical evolution of physical systems always obeys the Schrödinger equation. Therefore, in Maudlin's formulation of the measurement problem, Everett's proposal basically consist in a denial, or at least in a reassessment, of statement *iii*), that measurements always have a definite outcome.

Everett's original motivation is highly reasonable. If the operation of the projection postulate is associated to measurements, then measuring devices (and observers) must be treated as *external* to the system that is being measured. Therefore, the dynamics of QM given by the Schrödinger equation could not be applied to the universe as a whole, for the latter contains all measuring devices and all observers[78]. Besides, observers and measuring devices are composed of subsystems of the same type as measured systems, so it is natural to expect that the dynamics of postulate 5 apply to them as well. Therefore, in Everett's view, after a measurement interaction, the state $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$ really obtains. The challenging point is then to explain why we experience definite outcomes of the type $|\psi_i\rangle|\phi_i\rangle$, and to assign a consistent meaning to the probabilities for such outcomes expressed by $|c_i|^2$—which are also observed in experience. Everett was not completely clear in providing such an explanation, but his no-collapse proposal has led to several interpretational variations on a theme.

Everett originally claimed that each of the superposed definite measurement readings given by the $|\phi_i\rangle$ in system II is *relative* to a corresponding state of the measured system $|\psi_i\rangle$ in system I—this is why his view is commonly known as the 'relative-state interpretation'. Now, since the correct description of the post-measurement state is given by $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$, he argued that we have the experience of definite outcomes of the form $|\psi_i\rangle|\phi_i\rangle$ because relative states correspond to relative observers, to whom it *subjectively* appears that a wave function collapse has occurred. However, the *objective* and full description is

[78] 'The probabilities of the various possible outcomes of the observation are prescribed exclusively by Process 1 [wave function collapse according to Born's rule]. Without that part of the formalism there is no means whatever to ascribe a physical interpretation to the conventional machinery. But Process 1 is out of the question for systems not subject to external observation' (from Everett 1957, quoted in Barrett 1999, 59).

given by the superposed state. Everett did not go much further than this in offering a precise account of the distinction between the subjective collapse and the objective superposition[79], and the different Everettian interpretations of QM can be characterized by their attempts at providing such an account[80].

One possibility is described by 'the bare theory'[81]. In this interpretation the subjective wave function is simply a delusion originated in the application of quantum dynamics to the observer's belief in a certain measurement record. If we ask an observer measuring the spin of a particle in a certain direction whether she believes that the system is in a spin-up (+) or spin-down (−) state, if the state vector is given by $\frac{1}{\sqrt{2}}(|+\rangle|\phi_+\rangle + |-\rangle|\phi_-\rangle)$ she would answer 'no'. Assuming that the observer interacts with a system in the state $|+\rangle|\phi_+\rangle$, the result would be $|+\rangle|\phi_+\rangle|b_+\rangle$ (we write it down in the 'spin-state-belief'basis), which means that the observer believes that the object system is in a state of spin-up. The same goes, mutatis mutandis, for the system $|-\rangle|\phi_-\rangle|b_-\rangle$. But if we have a superposition $\frac{1}{\sqrt{2}}(|+\rangle|\phi_+\rangle|b_+\rangle + |-\rangle|\phi_-\rangle|b_-\rangle)$, the observer does not have a *definite* spin-state belief. Now, if we ask if she believes that the system is in a determinate spin state—regardless of whether it is up or down—her (deluded) answer would be 'yes'. If the total system is described by $\frac{1}{\sqrt{2}}(|+\rangle|\phi_+\rangle|b_+\rangle + |-\rangle|\phi_-\rangle|b_-\rangle)$, the observer can be said to have a spin-belief, but a superposed one. Having this (superposed) *belief* that the system has a spin value is, according to the bare theory, the reason why the observer is deluded in that the object system *has* a definite spin-state.

In short, the bare theory states that 'the dynamical equations of motion together with the standard way of thinking about what it means to be in a superposition somehow flatly contradict what we unmistakably know to be true of our mental lives'(Albert 1992, 118-9). There are several problems with the bare theory[82], but I think that the most basic one is the following. We might take the theory as an illustration that our beliefs about the results of measurements are wrong, but it says nothing about our *perceptual experiences* of finding definite measuring records. Even if we consider these experiences as illusory—in the sense that they provide us with a false determinate-outcome picture of a truly superposed world—the illusions would be so radical that they would undermine any possible empirical reasons we had to accept QM and the bare-theory in the first place.

Another Everettian interpretation goes like this. We could still affirm that the after-measurement state is objectively and correctly given by $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$, or more generally, that the whole physical world, including observer's brains, evolve according to postulate 5—to include the observer's brain relative state we can write down $\sum_i c_i|\psi_i\rangle|\phi_i\rangle|\beta_i\rangle$. However, we can postulate that when a spin-measurement is performed several minds supervene on the observer's brain in this state, according to the projection postulate and the Born rule. That is, when a measurement of the spin direction of system I takes place, to the same brain β that is described by a superposition of relative states in $\sum_i c_i|\psi_i\rangle|\phi_i\rangle|\beta_i\rangle$ there correspond different mind states $|\psi_i\rangle|\phi_i\rangle|\beta_i\rangle|m_i\rangle$ distributed with a relative frequency given by $|c_i|^2$. Like in the bare theory, the description of the objective world strictly follows the dynamics of postulate 5, but this time, as a result of measurements, on the observer's brain several splitting minds supervene according to the usual statistics of QM. In each of this minds, the observer correctly reports to have a definite belief—based on a corresponding mental experience—regarding the spin-state of the particle measured. The most obvious problem with the many-minds theory is the psycho-physical dualism it presupposes. The physical world is governed by the unitary dynamics of the Schrödinger equation, whereas the mind

[79] See (Barrett 1999, chapter 3).

[80] Notice that in Everett's original proposal there is no splitting in different worlds. His idea seems have been more that quantum reality is relative in the sense explained. It is a mistake to directly identify Everett's view with a many-worlds interpretation—*relative state interpretation* is a better label.

[81] See (Albert 1992, chapter 6) and (Barrett 1999, chapter 4).

[82] See (Barrett 1999, section 4.5)

world is governed by the projection postulate. We need not rehearse here all the (arguably insurmountable) difficulties that this 'Cartesian' position implies, especially in connection with the basic ontological and epistemological assumptions of modern science[83]—the many-minds interpretation falls prey to the same objections as the von Neumann-Wigner view.

The most widely known Everettian variation is called the 'many-worlds interpretation', originally proposed by Bryce DeWitt and Neil Graham. The basic idea is that

> it makes sense to talk about a state vector for the whole universe. This state vector never collapses and hence reality as a whole is rigorously deterministic. This reality, which is described *jointly* by the dynamical variables and the state vector, is not the reality we customarily think of, but is a reality composed of many worlds. By virtue of the temporal development of the dynamical variables the state vector decomposes naturally into orthogonal vectors, reflecting a continual splitting of the universe into a multitude of mutually unobservable but equally real worlds, in each of which every good measurement has yielded a definite result and in most of which the familiar statistical quantum laws hold (from DeWitt & Graham 1973, quoted in Barrett 1999, 149).

For example, consider a system described by $\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$ on the x-direction spin basis. If a measurement of this property is performed, the I + II system evolves, according to the unitary dynamics, into the state $\frac{1}{\sqrt{2}}(|+\rangle|\phi_+\rangle + |-\rangle|\phi_-\rangle)$. According to the many-worlds interpretation, this means that upon the measurement interaction the universe has split in two worlds, where the measurement outcome in each world is given by $|+\rangle|\phi_+\rangle$ and $|-\rangle|\phi_-\rangle$. None of these worlds exhausts reality, they are simply branches in the state defined by the wave function of the universe, so they are equally real. This is a natural explanation for why we have experiences of definite outcomes in experiments while at the same time the evolution of states is strictly governed by the Schrödinger equation—the collapse of the wave function is only apparent in each of the splitting worlds. Observers of course do not experience the splitting, for the worlds are mutually inaccessible and mutually unobservable. Barrett clearly summarizes the theory in the following statements:

> 1. There is a state vector that represents the state of the entire universe.
> 2. The global state evolves according to the usual deterministic linear dynamics and never collapses (one assumes that there is something like a global Hamiltonian that determines this evolution).
> 3. The universe (physical reality) consists of many mutually unobservable but equally real worlds.
> 4. A complete description of physical reality requires one to specify the universal state vector and the dynamical variables.
> 5. The state vector representing the global state naturally decomposes into orthogonal vectors that represent the states of the various worlds. There is exactly one world corresponding to each term in the preferred decomposition of the *global* state and each term describes the *local* state of the corresponding world.
> 6. The natural decomposition of the global state vector is one where there is a determinate record (typically different in each world) of the result of every good measurement, and this is what explains our determinate measurement records. (Barrett 1999, 151)

The many-worlds interpretations is comparatively popular among philosophers and physicists, but its cogency is challenged by many conceptual difficulties. To begin, the ontology postulated is rather bizarre. It is many time criticized insofar as it violates Ockham's principle. Supporters of this view usually reply that the vast ontology of branching worlds gets compensated by the conceptual simplicity achieved in the formal interpretation of the theory: it accepts the first five postulates at face value and does not require the projection postulate—in this sense it is the simplest interpretation of QM. However, the weirdness of the ontology looks bad not only from a simplicity point of view. To postulate that the reality described

[83] The original 'many-minds interpretation' is presented in (Albert & Loewer 1988). See also (Albert 1992, chapter 6), and for a detailed and critical discussion (Barrett 1999, chapter 7).

by a scientific theory is in its most part essentially unobservable seems to result in that the theory becomes (quasi)metaphysical discourse. To make sense of the theory, the supporters of this view are obliged to assert that a crucial part of the content of the theory is immanently beyond empirical test.

Second, the splitting process itself is problematic. One way to understand the coexistence of the worlds is that they are defined within the same space-time manifold[84]. In this case, it is difficult to see how conservation of energy-mass is respected, for the amount of mass-energy in the universe in each splitting becomes at least twofold. Besides, how one is to make sense of the essential disconnection between the worlds if they are all defined within the same space-time. One way out of this difficulty is that space-time itself gets split along with each measurement interaction. This would explain the inherent disconnectedness and we could say that in each space-time branch all the laws of physics are respected. But then other difficulties come up. R. I. G. Hughes, following J. Earman, complains that a spin-measurement, for example, causing a bifurcation of space-time is very strange idea, and if the supporter of many-worlds replies that the space-time bifurcation should not be understood as *caused* by a measurement, then it is not clear in what terms it should be understood[85].

Third, it is difficult to see how the many-worlds interpretation can make sense of the *probabilities* of the prediction in QM. Unlike the one-world view, the many-worlds proposal cannot define probabilities as relative frequencies in the outcomes of experiments, for the simple reason that in this case *all* possible outcomes obtain after a measurement. This can be seen considering the example of a physicist gambling on the result of an experiment. If the unitary evolution gives him an after-measurement state like $\frac{1}{\sqrt{3}}|+\rangle + \frac{\sqrt{2}}{\sqrt{3}}|-\rangle$, in the one-world picture he knows that $|-\rangle$ is a better bet than $|+\rangle$, but in the many-worlds scenario it is difficult to see how the squared-coefficients $\frac{1}{3}$ and $\frac{2}{3}$ can have a probabilistic meaning, for the physicist knows that after the measurement there will be two equally real worlds in which each outcome obtains, so $|+\rangle$ is no worse bet than $|-\rangle$—actually, it is difficult to make sense at all of gambling on the outcomes of measurements based on the amplitude coefficients.

Finally, the formalism of QM tells us that the state vector of a system can be written on different orthogonal bases corresponding to different observables, and this of course holds as well for composite systems. In the many-worlds interpretation it is postulated that at every instant a preferred-basis in which the wave function of the universe is written exists, and that it is such that the local state of each world depicts definite outcomes for all measurement interactions. But recall Everett's concept of relative states and think of the measurement interactions responsible for world splitting. That the relative states determine the world-splitting, together with the possibility of writing wave functions of the same state on different bases, entail the preferred-basis problem in the many-worlds proposal. Suppose that a measurement is performed on a system whose state is given by $\sum_i c_i|\psi_i\rangle$ with a measuring device described by the state $|\phi_0\rangle$. We know that the after-measurement system is given by $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$, and this means that each splitting world is described by each of the $|\psi_i\rangle|\phi_i\rangle$. But both states could be written on a different basis, say $\sum_i c_i'|\psi_i'\rangle$ and $|\phi_0'\rangle$, so that the final system can also be written as $\sum_i c_i'|\psi_i'\rangle|\phi_i'\rangle$, and in this case the same measurement interaction would imply that each splitting world would be given by each of the $|\psi_i'\rangle|\phi_i'\rangle$[86]. In rough words, the preferred-basis problem in the many-worlds interpretation is thus to find a justified criterion that tells us why the preferred-basis is preferred, to find an answer to the

[84] So does Lev Vaidman, for example: 'The Many-Worlds Interpretation is an approach to quantum mechanics according to which, in addition to the world we are aware of directly, there are many other similar worlds which exist in parallel at the same space and time' (Vaidman 2009, 1).

[85] 'To make sure that the different branches cannot interact even in principle they must be made to lie on sheets of space-time that are topologically disconnected after measurement […]. I do not balk at giving up the notion, held sacred until now, that space-time is a Hausdorf manifold. But I do balk at trying to invent a casual mechanism by which a measurement of the spin of an electron causes a global bifurcation of space-time' (from Earman 1989, quoted in Hughes 1989, 291).

[86] This simply means that the Hilbert space $\mathcal{H} = \mathcal{H}^I \otimes \mathcal{H}^{II}$ has been 'rotated' and a different factorization defined in the rotated space $\mathcal{H}' = \mathcal{H}^{I'} \otimes \mathcal{H}^{II'}$ is defined.

question why the wave function unitarily evolves on the specific basis described in Barrett's statement 6 above[87].

3.3.5 Modal interpretations

Another heuristic line is known as the 'modal interpretation of QM'. The basic idea was originally proposed by B. van Fraassen in the early 70's. The modal view rejects the projection postulate and reassesses the meaning of postulate 1. The state vectors and density operators are complete regarding the *possible* states that a system can take—hence the name 'modal'—, but they do not tell us what are the physical properties that are actually defined in the system at the corresponding instant. The modal outlook states that a really complete description of a system is given by the *dynamical state* (which simply corresponds to the state vector or density matrix) plus the *value state* that specifies the well-defined properties of a system at an instant[88]. The dynamical state always evolves according to the unitary dynamics and never collapses, and the well-defined properties that the value state defines provide an account for why measurement draw definite outcomes—so the projection postulate is not needed. Therefore, this interpretative approach denies the *eigenstate-eigenvalue link*, which states that a system has a definite value for a property (an eigenvalue of an observable) if and only if the system is an eigenstate of the corresponding observable. The modal interpretation accepts that if a system is in an eigenstate then it has a definite value for the corresponding property, but it is not the case that *only* systems which are eigenstates have definite values. In this way the modal approach intends to solve the measurement problem.

As we will see below, the Kochen-Specker theorem forbids that a system can have definite values for all observables at the same time, so the possible value states of a system are restricted and must be determined in such a way that the predictions of QM obtain and that they account for the definite outcomes of experiments. Here is where the modal approach ramifies in several different proposals. Though they all share the dynamic-value state distinction, not all of them determine the properties that the value states define in the same way. One of these proposals is commonly known as the Kochen-Dieks interpretation. These authors determine the value states based on the biorthogonal decomposition theorem. Given a state vector $|\Psi\rangle$ in a Hilbert space $\mathcal{H}^{\mathrm{I}} \otimes \mathcal{H}^{\mathrm{II}}$, there exist orthogonal bases $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$ for $\mathcal{H}^{\mathrm{I}}$ and $\mathcal{H}^{\mathrm{II}}$, respectively, such that $|\Psi\rangle$ can be written as a linear combination of the form $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$; the theorem tells us that if the coefficients c_i are such that $|c_i|^2 \neq |c_j|^2$ if $i \neq j$, then the decomposition is unique. In the Kochen-Dieks interpretation, the bases for each subsystem that the total system picks up when the theorem applies determine what are the definite properties depicted by the value state. In our example, this means that the total system $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$ has definite values for the properties associated to the observables corresponding to the eigenbases $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$.

This gives us a simple account for ideal measurements. Suppose that the observed subsystem I and the measuring apparatus subsystem II are expressed on the bases $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$, and that the biorthogonal decomposition theorem holds for the total after-measurement state $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$. Assuming that the mentioned bases determine the value state of the system—system I has a definite value for the measured property associated with the observable to which the eigenvectors $\{|\psi_i\rangle\}$ correspond, and system II has a definite value for the 'pointer position' observable associated with the eigenvectors $\{|\phi_i\rangle\}$—then it is totally natural that eigenvalues of the mentioned observables are obtained upon the measurement.

---

[87] For a detailed exposition and defense of the many-worlds interpretation, see (Wallace 2012).

[88] The definite-value properties that the value state of a system determines may vary over time. Another important remark is that 'the dynamical state in general only tells us what is *possible.* An important point is that one should not consider this modality as arising from an incompleteness of the description, which is the aim of science to remove. The dynamical state provides us with possible physical properties of the system, and this is all the theory has to do' (Lombardi & Dieks 2012, 4). This is why this approach reassesses, but does not deny, postulate 1.

This proposal has two main shortcomings, it does not work in cases of degeneracy—when $|c_i| = |c_j|$ the theorem does not hold—and it seems to require a relational conception of properties. Consider the three-component system $\alpha\beta\gamma$, the biorthogonal decomposition theorem could be applied to the two-component system $\alpha(\beta\gamma)$, to the system $\beta(\alpha\gamma)$, or to the system $\gamma(\alpha\beta)$, in the first case obtaining that the system α has a definite value for property P, in the second the system β has a definite value for property Q, and in the third case the system $\alpha\beta$ has a definite value for property R. If one assumes that the definite properties are essentially relational this is no problem, for the properties are not possessed absolutely by a system, but only in relation to another system[89]. But if the relational conception is not assumed, then one could ask how the definite-value properties of systems α and β connect to the definite-value properties of system $\alpha\beta$, that is, we could ask whether the definite-value properties of $\alpha\beta$ are given by $P \wedge Q$, by R, or by both.

A solution for the first problem was developed by Dieks and P. Vermaas. Instead of the unique biorthogonal decomposition, now the definite-value properties are determined by the spectral decomposition of a density operator. The basic idea is that given a system in the after-measurement state $\sum_i c_i|\psi_i\rangle|\phi_i\rangle$, we can take the reduced density operators, in their spectral decomposition form, $\rho_{\mathrm{I}} = \sum_i |c_i|^2 |\psi_i\rangle\langle\psi_i| = \sum_i |c_i|^2 P_{\psi_i}$ and $\rho_{\mathrm{II}} = \sum_i |c_i|^2 |\phi_i\rangle\langle\phi_i| = \sum_i |c_i|^2 P_{\phi_i}$, corresponding to subsystems I and II, respectively. The definite-value properties are now determined by the projectors P_{ψ_i} and P_{ϕ_i}, more precisely, the observable associated with the eigenvectors defined by the projector P_{ψ_i} has a definite value in system I, and the 'pointer position' observable associated with the eigenvectors defined by the projector P_{ϕ_i} has a definite value in system II. That is, the bases in which the after-measurement state is expressed correspond to observables for which the state has a definite value—so that it is again rather natural to obtain a definite outcome in the measurement. In this approach the definition of definite-value properties in terms of the biorthogonal unique decomposition simply becomes a special case, and when degeneracy obtains and the theorem does not hold, the spectral decomposition states that the projector that defines one of the possible values for the definite-value property is not one-dimensional, it projects the state not onto a ray, but onto a plane or a subspace of higher dimensions.

However, the problem with the 'relativity' of the properties comes back with a vengeance. The Hilbert space $\mathcal{H} = \mathcal{H}^I \otimes \mathcal{H}^{\mathrm{II}}$ could be rotated and a decomposition defined in the rotated space $\mathcal{H}' = \mathcal{H}^{\mathrm{I}'} \otimes \mathcal{H}^{\mathrm{II}'}$ can be given. G. Bacciagaluppi (1995) has shown that if the spectral decomposition is taken in each possible rotation in order to define the definite-value properties of the system, then if we assume that the properties are not relational a violation of the Kochen-Specker theorem follows—the system gets definite values for all properties at the same time, because each possible rotation of the same Hilbert space associates with a possible observable defined in that space. The 'perspectival modal interpretation' was introduced by Bene and Dieks in order to solve this problem. Taking seriously the idea that properties are essentially relational, this proposal affirms that the definite-value properties of a system are well defined only with respect to another physical system that acts as a *reference system*—and the rule to define such properties is simply given by the spectral decomposition. The definite-value properties of a system S are defined by the spectral decomposition of its density matrix, which is denoted by ρ_R^S. If the measured system is a part of the larger system A, the matrix ρ_A^S is simply the reduced matrix of S—in which the degrees of freedom in A that do not belong to S are traced out. In the case where the reference system R and the measured system S coincide, the state ρ_S^S is called 'the state of the system S with respect to itself'.

The perspectival approach then affirms that ρ_U^U,the quantum state of the universe with respect to itself, evolves according to the Schrödinger equation, and that for any system S contained in ρ_U^U, its state with respect to itself ρ_S^S is given by the reduced matrix ρ_U^S, and the definite-value properties of the system with respect to itself are defined by the eigenvectors of the observable(s) that the projectors of the last matrix define. A natural consequence of the perspectival conception of definite-value properties is that from the

[89] Notice the similarity with Everett's concept of a relative state.

perspective of a certain observer (reference system), the system *S* looks to have a definite value for a certain property, but from the perspective of a different observer the definite-value properties are otherwise. In analogy with the status of coordinates attributed to events in different reference frames in special relativity, the relative definite-value properties are *all* objective and *all* physically real. No subjectivism is implied, since the relational states are always unambiguously defined by the formalism[90].

The modal approach offers a neat account for measurement interactions in the idealized case, that is, if it is assumed that the correlations between the eigenstates of the 'pointer-position' observable of system II and the eigenstates of the measured observable in system I are perfect. This kind of interactions are ideal precisely in the sense that they are not obtained in practice. The realistic picture is that the correlations are not perfect: non-ideal measurements are interactions in which eigenstates of the 'pointer-position' observable in system II are correlated with relative states of system I that are non-orthogonal linear superpositions of eigenstates of the measured observable, or interactions in which eigenstates of the measured observable in system I are correlated with relative states in system II that are non-orthogonal linear combinations of eigenstates of the 'pointer-position' observable[91]. The problem is that in non-ideal measurements both the biorthogonal decomposition method and the spectral decomposition method select value states associated with observables other than the 'pointer-position' one, and this is an unacceptable result, of course.

The decoherence effect offers some help in this case. It can be shown that, in non-ideal measurements, the system-environment interaction implies that the reduced matrix is very close to the one expected in the ideal-measurement case, so that the apparatus pointer is approximately selected as the value state. However, it has also been shown that this works neatly in the case in which the measuring device system II is defined in a finite-dimensional Hilbert space, but in the more realistic case of an infinite-dimensional measuring system, the approximation between the 'pointer-state' observable and the observable picked out as value state by the spectral decomposition is not good enough.

Apart from the non-ideal measurements problem, there are other conceptual challenges that the modal approach must face in order to succeed, such as the application to the relativistic framework—it is not so easy to see how the modal interpretations can draw Lorentz-invariant results. In spite of these difficulties, Lombardi and Dieks give a general evaluation of the modal interpretations along the following line:

> These and similar problems, and their proposed solutions, have arisen in the context of detailed technical investigations. This illustrates one of the advantages of the modal approach: it makes use of a precise set of rules that determine the set of definite-valued observables, and this makes it possible to derive rigorous

[90] Both the biorthogonal decomposition approach and the spectral decomposition approach determine the value state of a system in terms of the rules for the representation of states of SQM. There are other modal proposals in which the determination of the value state is based on not-so-formal and more-physical considerations, such as the *atomic* modal interpretation and the modal *Hamiltonian* interpretation (see Lombardi & Dieks 2012, sections 3 and 9).

[91] Non-ideal measurements are interaction where eigenstate of an observable get correlated with states that are slightly superposed. For a very clear conceptual treatment of non-ideal measurements, see (Bub 1997, section 5.3). there the author explains that 'a non-ideal measurement, then, measures 'irreducibly unsharp' observables [a 'sharp' observable is given by a one-dimensional projector onto an eigenstate of the observable, an 'unsharp' observable is given by a projector onto a linear combination of eigenstates of the observable] […]. What a quantum mechanical measurement does is to reproduce, more or less precisely, the probabilities associated with an observable of the measured system in the probabilities of the pointer readings of the instrument […]. The more accurately the distribution of pointer readings reproduces the probabilities of measured system observables, the closer the measurement is to an ideal measurement. In the limiting case where the distribution of pointer readings exactly matches the probabilities of a system observable, we can regard the observable as having a value […]. So we can regard the limiting case of an ideal measurement as a dynamical process in virtue of which a measured observable comes to have a determinate value.

The methodological significance of an ideal measurement in quantum mechanics is similar to that of other ideal elements introduced in physics; for example, an ideal gas, or the straight line motion of a body not under the action of any forces in a Newtonian universe. It is only relative to the theoretical possibility of an ideal measurement that we can understand a non-ideal measurement as a *measurement*—as reproducing, more or less precisely, the probabilities of an observable of the measured system, where a precise reproduction of the probabilities in an ideal measurement would yield a value of the measured observable correlated with the value of the pointer reading' (Bub 1997, 154-5).

> results […]. Whatever the merit of the modal ideas in the end, one can at least say that they have given rise to a serious and fruitful series of investigations into the nature of quantum theory. (Lombardi & Dieks 2012, 30).

3.4 THE HIDDEN VARIABLES APPROACH: MOTIVATIONS AND CONSTRAINTS

Now that the most important proposals for a consistent interpretation of SQM have been reviewed[92], we can move on and consider the essentials of Bohm's theory. Before I directly address the formalism of the theory, it is appropriate to take a look at some relevant 'background' issues, namely, the motivations for the formulation of a hidden variables theory (HVT) in connection with the Einstein-Podolsky-Rosen argument, the meaning and consequences of von Neumann's 'impossibility proof' of HVT, and the conceptual and formal constraints that Bell inequalities and the Kochen-Specker theorem impose on any conceivable HVT.

3.4.1 The EPR argument and the completeness of SQM

Even though he actively participated in the formulation of the theoretical cornerstones of QM, Einstein always had a critical attitude towards (the predominating interpretation of) the theory. He profoundly disliked the objective and non-epistemic views that were usually hold concerning the probabilistic features of the theory, and also the abandonment of a realist conception of science apparently implied in Bohr's interpretation: he could not accept that, in the absence of measurements, the conceptual framework of science cannot be meaningfully used to describe the physical world.

During the fifth Solvay conference of 1927—shortly after Bohr's first public exposition of the complementarity view—Einstein undertook a crusade against QM as understood by Bohr & co. He attempted to show that the theory was problematic, in the sense that if it is held that the wave function provides a complete description of physical systems, then a form of action at a distance follows. His argument was given by a thought experiment. If a particle (photon or electron) is emitted from a source and it is made to pass through a narrow slit in a diaphragm on its way to a detection screen, the wave function of the particle after going through the slit gets diffracted and 'smeared out' in space. Before the particle's arrival, the wave function determines a non-zero probability that the particle is located in every point in the area of the screen. When the particle 'hits' the screen, the wave function of the particle collapses and a definite position is recorded. If the particle is found, say, at point A in the detection screen, this obviously implies that the particle is not at any other point B. According to Einstein, the event 'finding the particle in A' implies a sort of action at a distance, for in any other point B the probability for the particle to be found there instantly vanishes—the collapse of the wave function in A entails that it instantly vanishes at any other point in the detection screen. This form of action at a distance, Einstein argued, constitutes a violation of special relativity, for a causal connection between space-like events seems to be the case.

[92] Usually, reviews of the different interpretations of QM include the Ghirardi-Rimini-Weber (GRW) theory. This proposal denies statement *ii)* in Maudlin's inconsistent triad: the dynamics are not given only by the Schrödinger equations, a stochastic non-linear expression is included, and it determines that superposed states can suffer a physical collapse onto position eigenstates—so there is no measurement problem. I do not include this proposal in my review because the GRW theory is not really an interpretation of SQM, but a rival theory that, in principle, entails different empirical predictions. These predictions cannot be tested in the current state of technology, but this is enough to see that the GRW theory is not en element involved in the case of EE and UD I am addressing. For a simple and clear explanation of this theory, including the possible context in which divergent predictions obtain, see (Ghirardi 2011).

The problem of action at a distance was avoided, Einstein argued, if the wave function is not considered as a complete description of an individual system, but as representing an ensemble of identical systems. The empirical predictions are in both views the same, but the corresponding interpretations of probability are very different. In the conception of the wave function as a complete description of individual systems, Born's rule depicts an objective feature of the physical word: physical reality itself is probabilistic. On the other hand, in the ensemble interpretation the probabilities are simply relative frequencies of the results of identical experiments. This basic difference opens the possibility for an epistemic conception of probability: the wave function is not a complete description of an individual physical systems, and it fails to represent objective features that may be causally responsible for the definite outcomes of experiments:

> The difference in the two interpretations of probability as used in quantum mechanics is extremely important in view of the following fact. Although it does not lead to *experimental consequences*—for in both cases the confirmation of predictions requires the performance of an ensemble of identical experiments—it does lead to *interpretative consequences*: the Einstein frequency interpretation opens the way to a hidden-parameter theory, which reduces quantum mechanics to a branch of statistical mechanics; the Bohr-Born probabilistic interpretation precludes this possibility. (Jammer 1974, 119-20)

Einstein attempted to show also that the 'complementarity view' of QM is inconsistent, for, in principle, some conceivable experimental setups would lead to a violation of Heisenberg's uncertainty relations. In the Solvay conference of 1927, Einstein used the double-slit experiment as a possible setup in which these violations could occur, but Bohr countered his arguments along the lines explained in section 3.3.1, thus winning support for the complementarity interpretation. At the sixth Solvay conference of 1930, Einstein attempted the same task with yet another thought experiment, the 'photon-box'. The basic idea was that by measuring the weight of a box containing a clock before and after the emission of a photon, and considering the mass-energy relation of special relativity, both the energy and the time of the emission could be calculated with an accuracy greater than the limit allowed by Heisenberg's relations. Once again, Bohr countered the argument—this time invoking Einstein's own general relativity: the gravitational time-dilation implied in the position shift of the box after the photon-emission precluded an exact determination of the emission time. Einstein was thus forced to admit that Bohr's view was indeed consistent, and the general feeling after the conference was that Bohr had succeeded in convincing the scientific community of the correctness of his complementarity interpretation[93].

In 1935, in a paper written in collaboration with Boris Podolsky and Nathan Rosen entitled *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?*, Einstein resumed his crusade, but this time with the completeness of SQM as target. The EPR argument was built on a sufficient condition for physical reality, consisting in that 'if without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this quantity' (Einstein, Podolsky & Rosen 1935, 777); and on a requirement for the completeness of a physical theory: 'every element of the physical reality must have a counterpart in the physical theory' (ibid). Two implicit premises that are also relevant in the argument are the separability and locality principles:

> If two dynamical systems are spatially [space-like] separated, then each system can be characterized by its own properties, independently of the properties of the other system. That is, each system separately has a 'being-thus', a characterization in terms of certain properties intrinsic to the system, insofar as the systems are separable as dynamical systems.
> The locality principle is the requirement that no influence on a dynamical system can directly affect another system that is spatially separated from the first system. In particular, measurement performed on a system

[93] For a vivid, first-person outline of the Einstein-Bohr debate, including the EPR argument, see Bohr (1949). Detailed analysis can be found in (Jammer 1974, chapters 5 and 6) and (Fine 1986).

> cannot alter any properties of another system that is spatially separated from the first system. (Bub 1997, 41).

According to the criterion of physical reality, the existence of an element of physical reality corresponds to a determinate value of a physical property in the system. The basic idea of the EPR argument is to apply the reality criterion to a pair of space-like separated systems S_1 and S_2 such that the measurement of a physical property in system S_2 allows, according to the standard formalism of QM, to predict with certainty the value of a property in system S_1. Since the systems are space-like separated, the separability principle entails that they both have their own independent elements of reality, and the locality principle entails that the elements of reality of system S_1 cannot be affected by measurements performed on S_2. Therefore, that we can predict with certainty the value of a property P in system S_1 through a measurement in system S_2 means that there is an element of reality corresponding to P in system S_1.

Two space-like separated systems can be correlated in this way if they interacted in the past and became entangled. That is, if S_1 and S_2 are subsystems of a larger entangled system in the pure state $|\Psi\rangle$. This entangled larger system is the result of an earlier interaction between S_1 and S_2, but at the instant considered the subsystems are space-like separated. The EPR argument considers a system $|\Psi\rangle$ that can be written in two alternative forms $\sum_i c_i|a_i\rangle|e_i\rangle = \sum_j d_j|b_j\rangle|f_j\rangle$. Suppose that we measure the quantity F with eigenvectors $|f_j\rangle$ and eigenvalues f_i in S_2 and obtain the value f_k. This means that the state vector $|\Psi\rangle$ has collapsed onto the state $|b_k\rangle|f_k\rangle$, so that the state of subsystem S_1 is given by $|b_k\rangle$. We could also measure the property E with eigenvectors $|e_i\rangle$ and eigenvalues e_i in S_2. If we obtain the value e_l then the state vector $|\Psi\rangle$ collapses onto the state $|a_l\rangle|e_l\rangle$, and the state of S_1 is given by $|a_l\rangle$. Thus system S_1 can be predicted with certainty to be in two different quantum states if two different measurements are performed on S_2. But according to Einstein, Podolsky and Rosen, given the space-like separation between the systems, the separability and locality principles entail that system S_1 being in these two quantum states cannot be determined by the measurements on S_2. Therefore, there must simultaneously be elements of reality in S_1—independently of the measurements in S_2—corresponding to the property-eigenstates $|a_l\rangle$ and $|b_k\rangle$. In other words, we can choose to measure either E or F in S_2, by doing so, we can predict with certainty that system S_1 is in the eigenstates $|a_l\rangle$ and $|b_k\rangle$, respectively. Now, since S_1 has its own independent elements of reality (separability principle), and its states cannot be affected by measurements performed on S_2 (locality principle), this means that there are elements of reality in S_2 corresponding to the eigenstates $|a_l\rangle$ and $|b_k\rangle$.

Now, the eigenstates $|a_l\rangle$ and $|b_k\rangle$ can be eigenstates of non-commuting observables A and B. In the original EPR argument, this is exemplified by position and momentum. Given the state vector $|\Psi\rangle$, the *difference* of the position coordinates and the *sum* of the momentum components between the subsystems S_1 and S_2 are well defined. This is possible because the 'position difference' and the 'momentum sum' operators do commute. Taking advantage of this, by measuring the momentum or the position of system S_2 the momentum or position of system S_1 can be predicted with certainty. Another simple example was proposed by Bohm in 1951. Instead of momentum and position, Bohm considered spin in two different directions (recall that the Pauli spin-matrices S_x and S_y do not commute). Take, for example, the state vector $|\Phi\rangle = \frac{1}{\sqrt{2}}(|x_+\rangle|x_-\rangle - |x_-\rangle|x_+\rangle) = \frac{1}{\sqrt{2}}(|y_+\rangle|y_-\rangle - |y_-\rangle|y_+\rangle)$. In this case, we can run the EPR reasoning and predict with certainty, by spin-measurements performed in S_2, the spin values of system S_1 in the x and y-directions, without in any way disturbing the system S_1. Both the examples in the original EPR paper and in Bohm's version of the argument intend to illustrate that if we assume the separability and locality principles, then quantities defined by non-commuting observables can have simultaneous reality.

Now we can formulate the core of the argument. Einstein, Podolsky and Rosen state that the following disjunction is true: 'either (a) the quantum mechanical description of reality given by the quantum state

is incomplete, or (b) quantities represented by non-commuting operators cannot have simultaneous reality'. To see why this is true, take the proposition ($\neg a \wedge \neg b$), that is, 'the quantum mechanical description given by the quantum state is complete and quantities represented by non-commuting operators have simultaneous reality'. The statement ($\neg a \wedge \neg b$) cannot be true: considering the completeness requirement of the EPR argument, for the statement to be true the quantum state should define determinate values for properties represented by non-commuting operators, but it does not. Therefore, the statement $\neg(\neg a \wedge \neg b)$ is true, and this proposition is logically equivalent to ($a \vee b$). Now the final conclusion of the EPR argument can be presented. Given the separability and locality principles, the analysis of the quantum states given by $|\Psi\rangle$ and $|\Phi\rangle$ shows that in S_1 there are simultaneous elements of reality corresponding to properties described by non-commuting operators, i.e., that $\neg b$, and since it is the case that ($a \vee b$), a follows: the quantum mechanical description of reality given by the quantum state is incomplete.

One last important remark that Einstein, Podolsky and Rosen made in their groundbreaking paper was that a stronger criterion of reality is rejected, namely, that 'two or more physical quantities can be regarded as simultaneous elements of reality *only when they can be simultaneously measured or predicted*' (ibid, 780). It is clear that the elements of reality in the examples given by $|\Psi\rangle$ and $|\Phi\rangle$ do not fulfill this requirement. The momentum and position of system S_1 cannot be simultaneously predicted because the momentum and position of S_2 cannot be simultaneously measured—the same holds, *mutatis mutandis*, for $|\Phi\rangle$. Einstein, Podolsky and Rosen argue that this criterion is not adequate, though. If it were applied in the two mentioned examples, it would follow that the reality and objectivity of the quantities predicted with certainty in S_1 would depend on what measurements are performed on S_2, and, they claim, 'no reasonable definition of reality could be expected to permit this' (ibid). In other words, the stronger criterion of reality implies, in cases like $|\Psi\rangle$ and $|\Phi\rangle$, a rejection of the separability and locality principles, and the authors consider that a conception of a non-separable or non-local reality is untenable.

As mentioned above, Einstein's crusade against the main stream interpretation of SQM—a crusade that reached its pinnacle with the EPR argument—was motivated by his dissatisfaction with the abandonment of physical realism and determinism implied in Bohr's view. Until the advent of QM, modern physics had been usually understood as an attempt to provide a scientific explanation of the objective, independently-existing world. In Bohr's interpretation, though, SQM describes the physical world only when measurement contexts are considered, but the 'world in itself' is beyond its reach—if a meaningful notion at all. This view is associated with an objective, non-epistemic interpretation of the probabilities given by the Born rule: they are not a measure of our ignorance of the microstates of systems, but an objective feature of the physical world—and this conception also purported a drastic change in the classical conception of the aims and goals of modern physics, where determinism plays an essential role. Now, if extra parameters not considered in the standard formalism of QM are postulated, parameters that are supposed to represent objective features of quantum systems, the hope for a restitution of determinism and realism grows[94].

Besides these epistemological issues, the formulation of an empirically adequate HVT would allow an explicit explanation for the definite outcomes of quantum experiments and a unified account of the physical world in which a border between the quantum and the classical world is not drawn. The hidden

[94] 'The main driving force toward a belief that hidden variables should exist, therefore, is in the religious belief that "nature must be deterministic", and that "everything happening in nature must be predetermined by previous happenings in the physical world", even where our own knowledge is too limited for grasping the physical causes of what is happening. It therefore is often said that one would want to believe in hidden variables even if it were fundamentally impossible ever to know their values in advance. In that case, the knowledge that they would predetermine the results of a measurement would, of course, be of little or no *practical* use to us; but it would satisfy our emotional need for believing that *there exist* predetermining causes for all that is happening, even if we cannot measure these causes' (Belinfante 1973, 18). In a footnote linked to this passage the author clarifies that 'we use the word "religious" in a rather wide sense. For instance, we would consider dogmatic atheism to be a religious belief'. From our discussion of the EPR argument, we can see that for Einstein determinism was not the main issue, his main dissatisfaction with SQM was that it seemed to imply an abandonment of scientific realism.

variables postulated by the theories can be considered as causally responsible for the specific outcomes of measurements according to deterministic dynamical laws, restoring the classical picture of the physical world. In turn, the accomplishment of this goal would also allow an invaluable achievement in the foundations of quantum theory, namely, the dissolution of the measurement problem:

> The first dissatisfaction about quantum theory comes from people who reason that "*if different members of E_ψ are found upon measurements to have different values A_i of an observable A, then these individual systems must have been in different microstates*".
> The differences between these states then must be more subtle than what can be described by ψ. One would need additional parameters, say ξ, to describe them. According to these people, *the ψ and ξ together would completely determine the state of an individual system*. By 'completely determine' they mean that, if on a thus completely determined physical system a measurement is made which is fully described by giving the complete orthonormal set of eigenfunctions $\{\phi_i\}$ that are the possible results of the measurement, then ψ and ξ together will select from the set $\{\phi_i\}$ unambiguously the state ϕ_n which represents the result of the measurement. This may be described by $\psi \to \phi_n$, with $n = n(\psi, \xi, \{\phi_i\})$. (Belinfante 1973, 8)

The EPR argument, at least at first sight, supported this view. The correlated measurement outcomes in space-like separated systems in the examples above can be taken as suggesting that a hidden variable in the total systems $|\Psi\rangle$ and $|\Phi\rangle$, not considered in the respective representations of the quantum states, determines the correlations at the moment of the interaction between the systems S_1 and S_2. This view would of course comply with the separability and locality principles, and would naturally explain why there is an element of reality corresponding to $|a_l\rangle$ and $|b_k\rangle$ in system S_1. That is, the EPR argument provided a specific motivation for the goal of formulating a *completion* of QM. Einstein, Podolsky and Rosen explicitly stated that 'while we have thus shown that the wave function does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible' (780)[95]. However, as we will see below, it turned out that a HVT that is consistent with the empirical predictions of SQM cannot respect the locality principle—but this fact, of course, does not deny that the EPR argument played a catalyst role for the HVT approach. Jammer clearly describes the conceptual framework according to which a HVT would provide a fulfillment of all the desiderata just mentioned:

> Whereas classical particle mechanics […], as a theory about the behavior of individual systems logically as well as historically preceded its generalization to ensembles of systems—that is, classical statistics mechanics—the situation in quantum physics was the reverse: one had to construct a theory to explain the behavior of individual systems from the statistics of their ensembles, a task obviously much more complicated than its reverse.
> Such a theory, it was hoped, would not only restore determinism and causality to the realm of microphysics, it would also dispense with the peculiar dichotomy of physics into classical and quantum phenomena and reestablish a unitary account of the physical world, a prospect of sometimes greater incitement than the desire for determinism. A third, more specific motivation to search for such a "completion" of the theory was the problem raised by the Einstein-Podolsky-Rosen argument. The correlation between the two measurement results obtained at separated locations suggested that these results were actually determined in advance, when the two systems still interacted with each other, by certain dynamical variables also correlating the states of the systems after their separation. If these variables, though hidden from our sight and beyond our control, thus correlate the states, it could be understood that the outcome of one measurement makes it possible to predict that of the other without the need of assuming that the very performance of the first measurement influences causally the outcome of the second. (Jammer 1974, 253-4)

[95] What exactly Einstein conceived as a completion of QM is a controversial matter, though. As Bub summarizes: 'the Einstein-Podolsky-Rosen argument appears to show that the state descriptions of quantum mechanics are incomplete. This raises the question of what would count as a 'completion' of the theory. Several authors, notably Fine (1986, chapter 4) and Jammer (1974, p. 254) have suggested that Einstein had something other than hidden variables in mind here. Einstein's negative reaction to Bohm's hidden variable theory […] is often cited in support of this view […]. But Einstein's lack of enthusiasm for Bohm's theory, as this theory is usually formulated, should not be construed as a blanket rejection of 'completions' of quantum mechanics in the sense of hidden variable reconstructions of quantum statistics' (1997, 45).

3.4.2 Von Neumann's 'impossibility proof'

In his epoch-making book of 1932, John von Neumann obtained a formal result that would strongly determine the attitude of the scientific community towards the HVT approach that I have just outlined. During the following three decades, and mainly as a consequence of von Neumann's so called 'impossibility proof', it was generally thought that a 'completion' of SQM in terms of hidden parameters was unattainable. In the 1960s, though, the groundbreaking work of J. S. Bell made it clear that the scope of von Neumann's theorem was rather limited: it did not really show that HVTs, in general, would necessarily be empirically inadequate, but only that a very narrow and not-so-interesting class of HVTs were predictively doomed. Anyhow, the damage was done. The prevalence of the Copenhagen interpretation and the intellectual authority of von Neumann had already established a general spirit in the scientific community for which the HVT approach was either wrong or idle. Actually, Bohm's theory was introduced as early as 1952, but at that time it was scornfully received. A somewhat more positive reception and consideration came—though still only as an alternative-to-the-main-stream-view—after Bell's papers of 1964 and 1966.

Von Neumann's own proof is rather technical, but the basic idea is simple. He settled the question between the completeness and incompleteness of SQM by means of two possibilities of interpretation for a quantum system: as an ensemble of different individual systems, whose differences are given by a variable not represented by the state vector, or as an ensemble of identical systems completely described by the state vector:

> I. The individual systems […] of our ensemble can be in different states, so that the ensemble […] is defined by their relative frequencies. The fact that we do not obtain sharp values for the physical quantities in this case is caused by our lack of information: we do not know in which state we are measuring, and therefore cannot predict the results.
> II. All individual systems […] are in the same state, but the laws of nature are not causal. Then the cause of the dispersions is not our lack of information, but is nature itself, which has disregarded the 'principle of sufficient cause'. (Von Neumann 1955, 302)

If I is the correct interpretation of quantum states, then the formalism can in principle be completed by hidden variables that represent the difference between the individual systems in the ensemble, and these variables, in turn, may allow to deterministically predict a definite outcome in measurements. Now, If deterministic predictions in a measurement of a physical property corresponding to a certain observable are to be obtained, it must be the case that the quantum state represents a *dispersion-free* ensemble, and an ensemble is dispersion-free if and only if the expectation value of the squared operator acting on the corresponding state vector is equal to the square of the expectation value of the operator acting on the state vector, that is, it must hold that $\langle A\rangle_\Psi^2 = \langle A^2\rangle_\Psi$. Consider a system described by the state vector $|\Psi\rangle$ and an observable $A = \sum_i a_i|\alpha_i\rangle\langle\alpha_i|$. For simplicity, we can assume that A is defined in a two-dimensional Hilbert space, so that the state of the system can be expressed as $|\Psi\rangle = c_1|\alpha_1\rangle + c_2|\alpha_2\rangle$. The expectation or mean value of the observable A for system in state $|\Psi\rangle$ is $\langle A_\Psi\rangle = \langle\Psi|A|\Psi\rangle = c_1^2a_1 + c_2^2a_2$. The dispersion-free criterion does not obtain in our example, for $\langle A\rangle_\Psi^2 = (c_1^2a_1 + c_2^2a_2)^2$, but $\langle A^2\rangle_\Psi = c_1^2a_1^2 + c_2^2a_2^2$ [96]. But, following the hidden variable approach, let us suppose that a parameter ξ, not represented by $|\Psi\rangle$, determines a partition of the individual systems in two sub-ensembles $|\Psi_{a_1}\rangle$ and $|\Psi_{a_2}\rangle$. We can suppose that the value of the hidden parameter is $\xi > n$ and $\xi < n$ in each case, and that these values, respectively, determine that in the first sub-ensemble the measurements of A yield the eigenvalue a_1 with probability one, and that in the second sub-ensemble measurements yield the outcome a_2 with probability one. In this case we have that $\langle\Psi_{a_1}|A|\Psi_{a_1}\rangle = a_1$ and that $\langle A^2|\Psi_{a_1}|A^2\rangle = a_1^2$, so it follows that $\langle A\rangle_{\Psi_{a_1}}^2 = \langle A^2\rangle_{\Psi_{a_1}} =$

[96] If in the orthonormal basis $\{|\alpha_1\rangle, |\alpha_2\rangle\}$, $A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$, then it is obvious that $A^2 = \begin{pmatrix} a_1^2 & 0 \\ 0 & a_2^2 \end{pmatrix}$.

a_1^2; and by the same token, it holds that $\langle A\rangle_{\Psi_{a_2}}^2 = \langle A^2\rangle_{\Psi_{a_2}} = a_2^2$. When the hidden parameter is considered, thus, the sub-ensembles are dispersion-free, and a definite outcome is deterministically predicted in each case.

So far, so good. But von Neumann introduced a formal assumption known as the 'additivity postulate'. It can be proven in SQM that for all observable operators $A, B, ...$, regardless of whether they commute or not, the expectation value of an operator defined by $(pA + qB+\ldots)$, where $p, q, ...$ are real numbers, given any quantum state $|\Phi\rangle$ is such that $\langle\Phi|pA + qB + \cdots|\Phi\rangle = p\langle\Phi|A|\Phi\rangle + q\langle\Psi|B|\Phi\rangle + \cdots$. Von Neumann then considered a measurement on $|\Psi\rangle$ of another property associated with an observable $B = \sum_j b_j|\beta_j\rangle\langle\beta_j|$, which is defined in the same two-dimensional Hilbert space. We assume that, according to suitable values of ξ, we can define the sub-ensembles $|\Psi_{a_i b_j}\rangle$. We pick the sub-ensemble $|\Psi_{a_1 b_1}\rangle$, for which we know that measurements of A or B will yield the eigenvalues a_1 or b_1, respectively. We consider now the observable defined by $A + B$. Von Neumann showed that the additivity postulate entails that $\langle A + B\rangle_{\Psi_{a_1 b_1}} = \langle A\rangle_{\Psi_{a_1 b_1}} + \langle B\rangle_{\Psi_{a_1 b_1}}$, but we know that this is equal to $\langle A + B\rangle_{\Psi_{a_1 b_1}} = a_1 + b_1$. The problem with this last expression is that it means that the *eigenvalues* of $A + B$ obey the additivity postulate as well, but this is certainly not the case when the observables are non-commuting. Suppose that A and B are the non-commuting Pauli density matrices S_x and S_y, that we know have eigenvalues ± 1. The spin operator corresponding to the direction given by the bisector between the x and y axes is defined by $\frac{S_x+S_y}{\sqrt{2}}$, and its eigenvalues are also ± 1; but if the additivity postulate applies to eigenvalues as well, they should be $(\pm 1 \pm 1)/\sqrt{2}$, which is of course contrary to experience. In short, von Neumann's argument tells us that if we introduce hidden variables to obtain dispersion-free ensembles for which definite outcomes can be deterministically predicted, then the additivity postulate entails that the eigenvalues of observables defined by sums of non-commuting observables are also additive. Since this is refuted by experience, the conclusion is that HVT cannot reproduce the (confirmed) predictions of SQM. In von Neumann's own words: 'it is therefore not, as is often assumed, a question of reinterpretation of quantum mechanics—the present system of quantum mechanics would have to be objectively false, in order that another description of the elementary processes than the statistical one be possible' (1955, 325).

As mentioned above, von Neumann's result led to a widespread opinion within the scientific community according to which the HVT approach was a non-starter. The general feeling that in the Bohr-Einstein debate the former was the winner, plus von Neumann's impossibility proof, determined a very hostile general attitude towards hidden variables endeavors[97]. However, one of the few 'mavericks' that questioned the proof was the German mathematician-philosopher Grete Hermann. In a 1935 essay she challenged that the observance of the additivity postulate for expectation values was a reasonable and justified constraint for a HVT. It is possible and expected, she argued, that the introduction of a parameter ξ, that allows the partition of the ensembles into dispersion-free subensembles, also entails that for the latter subensembles the additivity postulate for expectation values does not hold. Hermann affirmed that, in the context of a HVT, retaining the additivity postulate for individual systems is equivalent to assuming that such individual systems cannot be differentiated by the introduction of further parameters that may determine specific outcomes in measurements of a property—and so von Neumann's proof is question begging[98]. Hermann put her finger on the weakest feature of von Neumann's argument, but it was largely ignored.

---

[97] 'There were always a few mavericks who questioned the validity of the proof [...]. But most physicists were only too pleased to accept the word of such a famous mathematician—when, in any case, it told them what they wanted to hear. Cynically, I don't even believe that many of them studied the proof in detail, and just missed the problems. I suspect they perhaps glanced through it, or more likely didn't even bother to do that! They believed that everything had come together to support a rather unholy Bohr/von Neumann axis, which should not be questioned, and did not really have to be understood—just accepted' (Whitaker 1996, 200).

[98] As Jammer explains: 'Grete Hermann now pointed out that since an arbitrary ensemble is a mixture of pure cases it presumably suffices to claim the additivity only for ensembles all elements of which are described by the same (pure state)

J. S. Bell (1966) leveled a similar criticism, but instead of charging von Neumann's argument with a question begging accusation, he stated that, in the context of a HVT, the additivity postulate for dispersion-free subensembles determined by certain ξ-values is a requirement completely beside the point. If A and B are non-commuting operators, he argued, the non-additivity of the eigenvalues of $A + B$ is naturally explained and expected, for a measurement of the associated property requires a quite distinct experiment than the ones required to measure A and B—so the property value associated to $A + B$ cannot be obtained by trivially combining outcomes of the experiments set out to measure A and B. In the spin example above, measuring the observable defined by $\frac{S_x+S_y}{\sqrt{2}}$ requires a Stern-Gerlach device oriented in the corresponding direction. Therefore, it is rather obvious that in a viable HVT the additivity of eigenvalues should not hold. In turn, this obviousness also indicates that the observance of additivity of expectation values in the case of dispersion-free (sub)ensembles is not a reasonable requirement for a viable HVT either. Even if the expectation values of dispersion-free subensembles do not observe the additivity postulate, it is still possible that in the mean—when the expectation values of all the dispersion-free subensembles are averaged over in the total ensemble—the additivity condition is met anyway. In other words, Bell shows that what von Neumann's argument achieves is only to discard a not-so-interesting and clearly nonviable type of HVTs:

> This explanation of the nonadditivity of allowed values also establishes the nontriviality of expectation values. The latter is a quite peculiar property of quantum mechanical states, not to be expected *a priori*. There is no reason to demand it individually of the hypothetical dispersion free states, whose function is to reproduce the *measurable* peculiarities of quantum mechanics *when averaged over*. (Bell 1966, 449).

Bell's analysis has certainly changed the general appraisal of von Neumann's argument. Actually, the criticized additivity assumption has been harshly characterized by Bell himself and others[99]. Though it is clear that the proof does not achieve the goal that has been normally attributed to it, there are some authors that argue that it does draw an interesting conclusio. Jammer, for example, criticizes Hermann's charge of circularity. The conclusion in von Neumann's argument is that dispersion-free ensembles violate the additivity assumption when non-commuting observables are considered because their eigenvalues are not additive. Jammer states that the additivity assumption in the proof is not stipulated as valid for non-commuting observables alone—if it were the argument would be certainly circular, for the non-additivity of eigenvalues of non-commuting operators would imply that the additivity assumption automatically rules out dispersion-free ensembles. The assumption, in the case of commuting operators, does not automatically rule out dispersion-free ensembles, so the conclusion is not logically contained in

---

wave function. For such ensembles, however, von Neumann, according to Hermann, resorted to the mathematical formalism $(\varphi, (R + S)\varphi) = (\varphi, R\varphi) + (\varphi, S\varphi)$ $[\langle\varphi|R + S|\varphi\rangle = \langle\varphi|R|\varphi\rangle + \langle\varphi|S|\varphi\rangle]$ as a valid relation irrespective of whether R and S commute. Hermann now objected that, as long as the possibility of hidden variables has not yet been disproved, $(\varphi, R\varphi)$ denotes the expectation value of R for such ensembles E alone whose elements are described by φ. This does not imply that also subensembles E_1 of E, defined by perhaps not yet available criteria (hidden variables), have the same expectation value of R, nor that the latter satisfies the additivity condition. Thus an important step in von Neumann's proof is lacking. But to retain this assumption, as von Neumann did, is tantamount to assuming that the elements of an ensemble, described by φ, cannot be further differentiated by any criteria on which the result of an R measurement may depend. Since the denial of the existence of such criteria is precisely the thesis that has to be proved, Hermann concluded that von Neumann's proof is circular' (Jammer 1974, 273).

[99] 'Von Neumann's no-hidden variables proof was based on an assumption that can only be described as silly—so silly, in fact, that one is led to wonder whether the proof was ever studied by either the students or those who appealed to it to rescue them from speculative adventures' (Mermin 1993, 805-6). Mermin also quotes a 1988 interview with John Bell where he states that 'yet the von Neumann proof, if you actually come to grips with it, falls apart in your hands! There is *nothing* to it. It's not just flawed, it's *silly*! [...]. When you translate [his assumptions] into terms of physical disposition, they're nonsense. You may quote me on that: The proof of von Neumann is not merely false but *foolish*!' (quoted in Mermin 1993, 805, footnote 8).

the additivity assumption. Jammer concludes that the right criticism is the severe and unjustified restriction of the class of conceivable dispersion-free ensembles: those for which the additivity assumption holds (this is what Bell said, of course). Now, even though the proof does not establish the impossibility of HVT, it does show, according to Jammer, that the *standard formalism* of quantum mechanics does not allow the interpretation given by the possibility I that von Neumann defined (see above):

> We agree with Grete Hermann's criticism that the proof did not achieve its declared objective of demonstrating that quantum mechanical ensembles cannot be decomposed into *any* kind of dispersion-free subensembles [...]. But we do not dismiss the proof as nugatory. True, in view of von Neumann's excessively restrictive assumptions it is not an *impossibility proof* of any conceivable class of hidden variables, but it is a *completeness proof*, in this respect, of von Neumann's axiomatics (with the inclusion of [the additivity assumtpion]), since it shows that this formalism does not admit nonquantum mechanical ensembles. It may even be regarded as a *consistency proof* of this formalism with its usual interpretation. (Jammer 1974, 274, footnote 45)

More recently, Jeffrey Bub (2010) has offered a similar analysis. Though Bub does not explicitly mention Jammer's view, his proposal can be considered as an elaboration on it. He argues that the interesting point in von Neumann's proof (of which its author was aware, Bub claims) is that it clarifies that an empirically adequate HVT must be such that the association between physical properties and Hermitian operators in Hilbert space—essential in the standard formalism—does not hold:

> According to Bell, von Neumann proved only the impossibility of hidden deterministic states that assign values to a sum of physical quantities, $\mathcal{R} + \mathcal{S}$, that are the sums of the values assigned to quantities $\mathcal{R}$ and $\mathcal{S}$, even when $\mathcal{R}$ and $\mathcal{S}$ cannot be measured simultaneously. As we saw, von Neumann regarded a sum of physical quantities that cannot be measured simultaneously as implicitly defined by the statistics [*via* $\langle\Phi|R+S|\Phi\rangle = \langle\Phi|R|\Phi\rangle + \langle\Phi|S|\Phi\rangle$], and he drew the conclusion that such an implicitly defined physical quantity cannot be represented by the operator sum in a hidden variable theory. (Bub 2010, 1338)[100]

In the standard formalism, the expectation values, expressed in terms of Hermitian operators applied to quantum states, implicitly define physical properties like $\mathcal{R} + \mathcal{S}$. Bub claims that what von Neumann's proof really tells us is that, in a HVT, the trace formula $\langle R+S\rangle_\rho = \mathrm{Tr}(\rho(R+S))$ generates the measured probabilities corresponding to a property $\mathcal{R} + \mathcal{S}$ requires further explanation, for the latter property cannot be implicitly defined by the expectation value of the Hermitian operator $R + S$—otherwise eigenvalue additivity follows for non-commuting operators. Actually, he continues, this is the case in Bohm's HVT, where a disturbance theory of measurement generates the statistics. As mentioned above, Bub claims that von Neumann's proof's lesson is that in a HVT the link between physical properties and Hermitian operators breaks down:

> Von Neumann's proof establishes that if the physical quantities of quantum mechanics [...] are characterized by the [property-Hermitian operator, additivity assumption] conditions, then dispersion-free states are excluded and the quantum pure states are the appropriate extremal states for quantum probability distributions. So in a hidden variable theory in which dispersion free (deterministic) states are the extremal states for quantum probability distributions, the quantum probabilities could not reflect the distribution of pre-measurement values of beables, but would have to be derived in some other way, e.g., as in Bohm's theory, where the probabilities are an artefact of a dynamical process that is not in fact a measurement of

[100] To support his view that this is what von Neumann really had in mind, Bub quotes the following passage: 'It should be noted that we need not go any further into the mechanism of the 'hidden parameters', since we now know that the established results of quantum mechanics can never be re-derived with their help. In fact, we have even ascertained that it is impossible that the same physical quantities exist with the same function connections [property-Hermitian operator] if other variables [hidden parameters] should exist in addition to the wave function.
Nor would it help if there existed other, as yet undiscovered, physical quantities, in addition to those represented by the operators in quantum mechanics, because the relations assumed by quantum mechanics [property-Hermitian operator] would have to fail already for the by now known quantities' (von Neumann 1955, 324-5).

> any beable[101] of the system. What von Neumann's proof excludes, then, is the class of hidden variable theories in which (i) dispersion free (deterministic) states are extremal states, and (ii) the beables of the hidden variable theory correspond to the physical quantities represented by the Hermitian operators of quantum mechanics. (ibid, 1339-40)

Bub's reappraisal of von Neumann's proof is quite interesting. First, it indicates that the formulation of an empirically viable HVT cannot be a mere completion of the standard formalism of QM (in this sense Bub's reasoning is an elaboration of Jammer's proposal that the proof is a completeness proof). If the property-Hermitian operator link is assumed, along with the additivity assumption, then the introduction of a hidden parameter would lead to wrong predictions. Second, the fact that in an empirically viable HVT the property-Hermitian operator connection must break down is an indication that such a theory would rather likely introduce a different ontology with respect to SQM. Since physical properties would necessarily be represented by conceptual tools other than Hermitian operators, it is expectable that such properties will not be like the properties represented in the standard formalism. Both remarks suggest that the formulation of a viable HVT would likely result in a *rival* theory rather than in an *alternative interpretation* of SQM, a result which is indeed verified in Bohm's theory.

3.4.3 The Kochen-Specker theorem

Another formal result that is highly relevant for the possibility of HVTs is given by a theorem due to S. B. Kochen and E. P. Specker (1967). In a word, the Kochen-Specker (KS) theorem states that 'it is, in general, impossible to ascribe to an individual quantum system a definite value for each of a set of observables not all of which necessarily commute' (Mermin 1990, 3373). That is, given a suitable set of observables defined in an n-dimensional Hilbert space with $n > 2$, a quantum system cannot possess, at the same time, definite values for all the properties associated with the observables in the set. This, of course, yields an important limitation for possible HVTs. The HVT approach was originally thought as a way to recover the classical picture in the quantum world—by means of a conceptual framework analogous to statistical mechanics—but the KS theorem shows that an essential feature of the classical picture is unachievable: a viable HVT must renounce to the aim of representing quantum systems in a way such that all of their properties have measurement-independent definite values.

The proof of the theorem goes like this. Consider the Hermitian operators corresponding to spin-1 particles in the x, y, and z-directions defined in a three-dimensional Hilbert space:

$$\mathrm{S}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad \mathrm{S}_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \qquad \mathrm{S}_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The operators corresponding to the 'squared-spin' observable are thus given by:

$$\mathrm{S}_x^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathrm{S}_y^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathrm{S}_z^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Unlike the spin operators, these squared-spin operators mutually commute. As it can be seen, they have eigenvalues 0 and 1, which of course correspond to the possible values of the corresponding property. It is also easy to see that their sum is given by $\mathrm{S}_x^2 + \mathrm{S}_y^2 + \mathrm{S}_z^2 = 2(\mathbf{1})$, where $\mathbf{1}$ is the unity or identity

[101] The term '*be*able' was introduced by John Bell to denote the element of reality that a (hidden) parameter represents (in contrast to '*observ*able').

operator. This in turn means that the eigenvalues of the squared-spin-sum operator are 2, so a measurement of the corresponding property will always yield a value of 2. Therefore, in the sum $\mathcal{S}_x^2 + \mathcal{S}_y^2 + \mathcal{S}_z^2 = 2$ (the terms in this sum are the values of the properties $\mathcal{S}_i^2$ that correspond to the observable operators S_i^2), two terms should have the value 1 and the other should have the value 0[102]. This holds for any sum $\mathcal{S}_\alpha + \mathcal{S}_\beta + \mathcal{S}_\gamma$, provided that α, β and γ represent mutually orthogonal directions. Notice also that *all* vectors defined in our three-dimensional Hilbert space are eigenvectors of the operator $2(\mathbf{1}) = \mathrm{S}_x^2 + \mathrm{S}_y^2 + \mathrm{S}_z^2$, so that these features hold for any arbitrary quantum system defined in this space.

Let us now pick an arbitrary state vector $|\Phi\rangle$ defined in the three-dimensional Hilbert space we are considering. We can define a sphere corresponding to that state for which we may expect that assumption (i) holds: for all possible orthogonal triple of points on its surface (each triple defined by three orthogonal vectors from the sphere's center), just one point receives a value 0 and the other two receive the value 1—let us call this assignment an A-assignement. That is, the set of all orthogonal triples of points on the sphere represents the set of the values of properties $\mathcal{S}_\alpha^2, \mathcal{S}_\beta^2, \mathcal{S}_\gamma^2$ in $|\Phi\rangle$, defined for all possible (mutually orthogonal) squared-spin directions—and (i) assumes that all these values respect the functional relations defined in the previous paragraph. In other words, (i) states that the system represented by $|\Phi\rangle$ has definite values for all the properties $\mathcal{S}_\alpha^2, \mathcal{S}_\beta^2, \mathcal{S}_\gamma^2$ at the same time. Now, the K-S theorem tells us that the following statements are true for our sphere: (ii) there is an angle β such that if any point p on the sphere receives a value 0 on an A-assignement, then so does any point q at angular distance β from p; and (iii) if one point on the sphere receives value 0 on an A-assignment, then, from (ii), so do all others. It is clear that (iii) contradicts (i), therefore, no A-assignement exists. That an A-assignement does not exist means that for the set of the squared-spin operators $\mathrm{S}_\alpha^2 + \mathrm{S}_\beta^2 + \mathrm{S}_\gamma^2$ defined in all possible mutually orthogonal directions, there is no consistent eigenvalue assignment such that the system described by $|\Phi\rangle$ can have definite values for all the corresponding properties $\mathcal{S}_\alpha^2, \mathcal{S}_\beta^2, \mathcal{S}_\gamma^2$.

To establish (ii), consider an arbitrary orthonormal triple of vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ from the center of the sphere. From this triple we can generate two more triples of mutually orthogonal vectors, $\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}, \mathbf{z}\}$ and $\{\mathbf{x}+\mathbf{z}, \mathbf{y}, \mathbf{x}-\mathbf{z}\}$. There is no A-assignment such that $v(\mathbf{x}+\mathbf{y}) = v(\mathbf{x}-\mathbf{y}) = 1$ and $v(\mathbf{x}+\mathbf{z}) = v(\mathbf{x}-\mathbf{z}) = 1$, for this would entail that $v(\mathbf{z}) = v(\mathbf{y}) = 0$. Consider now the vectors $(\mathbf{y}+\mathbf{z})-\mathbf{x}$ and $(\mathbf{y}+\mathbf{z})+\mathbf{x}$, for which it holds that $(\mathbf{x}+\mathbf{y}) \perp [(\mathbf{y}+\mathbf{z})-\mathbf{x}] \perp (\mathbf{x}+\mathbf{z})$ and $(\mathbf{x}-\mathbf{y}) \perp [(\mathbf{y}+\mathbf{z})+\mathbf{x}] \perp (\mathbf{x}-\mathbf{z})$, respectively. There is no A-assignment in which $v[(\mathbf{y}+\mathbf{z})-\mathbf{x}] = v[(\mathbf{y}+\mathbf{z})+\mathbf{x}] = 0$, for this would imply that $v(\mathbf{x}+\mathbf{y}) = v(\mathbf{x}-\mathbf{y}) = v(\mathbf{x}+\mathbf{z}) = v(\mathbf{x}-\mathbf{z}) = 1$, and we just saw that this cannot be the case. We have found then two vectors, $(\mathbf{y}+\mathbf{z})-\mathbf{x}$ and $(\mathbf{y}+\mathbf{z})+\mathbf{x}$, that cannot be both assigned a value 0. Calculating their inner product, we find that the angle α between these vectors is $\cos^{-1}\frac{1}{3} \sim 70°$. Since the basis $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ was arbitrarily chosen, then no two points on the sphere that are separated by an angle α can be both assigned a value 0 in an A-assignment. Now consider a vector $\mathbf{w}$ that lies on the x-y plane and that makes an angle α with $\mathbf{y}$, so it follows that $\mathbf{w}$ makes an angle $\beta = 90° - \alpha$ with $\mathbf{x}$. Suppose an A-assignment such that $v(\mathbf{w}) = 0$, therefore $v(\mathbf{y}) = 1$. Since $\mathbf{w} \perp \mathbf{z}$, then $v(\mathbf{z}) = 1$, and it follows that $v(\mathbf{x}) = 0$. Again, since the basis $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ was arbitrarily chosen, if two points p and q on the sphere are separated by an angular distance β, then for any A-assignmenet in which $v(p) = 0$ it follows that $v(q) = 0$. That is, (ii) holds. To establish (iii), we simply note that for any two different points p and q on the sphere there is a finite sequence of points $\langle p_1, p_2, \ldots, p_n \rangle$, with $n \geq 2$, such that $p_1 = p$, $p_n = q$ and that the angular separation between successive points is β. Therefore, from (ii), it follows that any A-assignment 0 to p also assigns 0 to q, and since *any* two different points p and q can be part of a sequence like the one just decribed, any

[102] Notice that this means that the functional relations that hold between the squared-spin operators also hold for their eigenvalues, and, *a fortiori*, for the values of the corresponding properties. That is, we are assuming that the additivity postulate holds for eigenvalues. In this case, the operators at issue commute, so it does not seem to be a problematic assumption.

A-assignment 0 to p assigns 0 to all other points in the sphere. Therefore, A-assignments are impossible—this concludes the proof[103].

A remarkably simple alternative proof (in a four-dimensional Hilbert space $\mathcal{H}^2 \otimes \mathcal{H}^2$) was offered by Mermin (1993). Consider the Pauli spin matrices for two independent spin-$\frac{1}{2}$ particles, namely, σ_μ^1 and σ_ν^2. The relevant properties of these operators are the following: the eigenvalues of each operator (in each possible directions) are ± 1; any σ_μ^1 commutes with any σ_ν^2; when μ and ν specify orthogonal directions, σ_μ^i and σ_ν^i with $i = 1,2$ (i.e., $\sigma_\mu^i \sigma_\nu^i = -\sigma_\nu^i \sigma_\mu^i$, so $[\sigma_\mu^i, \sigma_\nu^i] = -[\sigma_\nu^i, \sigma_\mu^i]$); and $\sigma_x^i \sigma_y^i = i\sigma_z^i$ for $i = 1,2$. Once again, we assume in the proof that the functional relations between (commuting) operators are 'inherited' by their eigenvalues and the corresponding property values. Consider now the following table of observables:

	(a)	(b)	(c)
(i)	σ_x^1	σ_x^2	$\sigma_x^1 \sigma_x^2$
(ii)	σ_y^2	σ_y^1	$\sigma_y^1 \sigma_y^2$
(iii)	$\sigma_x^1 \sigma_y^2$	$\sigma_x^2 \sigma_y^1$	$\sigma_z^1 \sigma_z^2$

From the properties of spin operators just mentioned and the 'inherited' functional relations it follows that 1) the operators in each row and in each column are mutually commuting; 2) the product of the three observables in column (c) is $-\mathbf{1}$, whereas the product of the three observables in the other two columns and in all three rows is $\mathbf{1}$; and 3) the product of the property values assigned by the three observables in column (c) is -1, whereas the product of the property values assigned by the observables in the other two columns and in all three rows is 1. However, there is no consistent assignment of property values such that 3) holds, for the row equalities imply that the product among all nine property values is 1, but the column equalities imply that the product of all nine property values is -1. Therefore, for the set of 9 observables in the table, there is no consistent assignment such that a quantum system can have definite values for all the corresponding properties.

The K-S theorem is highly relevant for the possibility of HVTs. One of the expected consequences of the introduction of a parameter (linked to a beable) not represented by a state vector or a density operator in the standard formalism, a hidden variable, would be that it determines an objective (measurement-independent), definite value for all physical properties in the corresponding quantum system. This result, the K-S theorem tells us, is logically impossible—assuming that the functional relations between commuting operators hold for their eigenvalues too, which, unlike the additivity postulate for non-commuting operators, is a rather natural and justified constraint for an empirically viable HVT.

However, the theorem does not imply a complete disavowal of the HVT approach. Commenting on a formal result that is equivalent to the K-S theorem (a corollary of Gleason's theorem), John Bell pointed out that 'it was tacitly assumed that measurement of an observable must yield the same value independently of what other measurements may be made simultaneously' (1966, 451). The same assumption

[103] The argument just sketched holds for a three-dimensional Hilbert space. To generalize it to n-dimensional spaces with $n > 2$, an analogous proof can be given in which instead of angular distance between points on a sphere, angles defined by the inner product between vectors are considered. A result equivalent to the KS theorem follows as a corollary of a theorem proved by A. Gleason in 1957. John Bell (1966) explicitly considered this corollary in connection with the possibility of HVTs.

A possible objection against this specific proof (and also against Bell's version) is that if the assumption that to any orthogonal basis on the sphere an observable corresponds is dropped, it might be the case that A-assignments are inconsistent only if the bases that do not correspond to any observable are considered—so that A-assignments may exist for all possible observables. In their original proof, though, Kochen & Specker showed that for a system of orthohelium in its lower orbital state there is a set of 117 vectors, corresponding to different coordinate decompositions of spin, for which A-assignments are impossible. That is, all the bases involved in that proof certainly are bases of operators that clearly correspond to observables.

is made in both the reviewed proofs. In the first proof, consider a property $\mathcal{S}_z^2$ associated to the operator S_z^2. The tacit assumption is that the value we find for this property when we measure $a\mathcal{S}_x^2 + b\mathcal{S}_y^2 + c\mathcal{S}_z^2$ must be the same than the value we find when we measure $a\mathcal{S}_{x'}^2 + b\mathcal{S}_y^2 + c\mathcal{S}_z^2$ (notice that in each case, the three directions considered are mutually orthogonal, but the primed and unprimed directions are not: S_z^2 mutually commutes either with S_x^2 and S_y^2, and with $S_{x'}^2$ and $S_{y'}^2$; but S_x^2 and $S_{x'}^2$ and S_y^2 and $S_{y'}^2$, are non-commuting operators). If this 'non-contextuality' for the values of properties is assumed, then it turns out that there is no consistent assignment such that the system has definite values for all the properties determined by the set of observables considered.

If one were to expect that a HVT restates the classical conception of fully objective properties with determined values even in the absence of measurements, then 'non-contextuality' would be a natural requirement. The KS theorem shows that non-contextual HVTs are not possible, but contextual ones are still conceivable. In such a theory, the measurement of an observable may yield a value that depends on what other measurements are performed simultaneously. In the context of the Mermin's proof, for example, we can conceive a HVT such that the measurement of the observable $\sigma_x^1\sigma_x^2$ draws different values depending on whether a simultaneous measurement of the other two observables in column (c) or a simultaneous measurement of the other two observables in row (i) is performed. Even though contextuality is a feature that certainly moves off from an essential aspect of classical physics, there is a justification for it in a *quantum* HVT. A measurement $\sigma_x^1\sigma_x^2$ along with the observables in column (c) requires a different (complementary!) experimental setup than a measurement of $\sigma_x^1\sigma_x^2$ together with the observables in row (i) because, even though the observables in the row are mutually commuting and the observables in the column are mutually commuting, it is clear that there are observables in the former that do not commute with observables in the latter. In Bell's words,

> these different possibilities require different experimental arrangements; there is no *a priori* reason to believe that the results for $[\sigma_x^1\sigma_x^2]$ should be the same. The result of an observation may reasonably depend not only on the state of the system (including hidden variables) but also on the complete disposition of the apparatus. (ibid).[104]

3.4.4 Bell's theorem

As we just saw, Bell argued that the non-contextuality that the KS theorem rules out for consistent and empirically viable HVTs is not an unsurmountable problem, for there are physical arguments that make contextuality a plausible and expected feature in a quantum HVT. However, Bell also noticed that there are certain cases in which a specific form of non-contextuality may naturally be expected in a quantum HVT—even if the theory is, in general, contextual. Consider the following case. An observable A is pairwise compatible with observables B and C, but B and C are non-commuting. Thus, we can define an observable X that is a function of A and B, and an observable Y that is a function of A and C, such that X and Y are non-commuting. Suppose that A is an observable of a system S, and B and C are incompatible observables of a system S' space-like separated from S. Discarding the mediation of an action at a distance—respecting the spirit of special relativity—we may naturally expect that the value we find for the property $\mathcal{A}$ upon a measurement of X will be the same as the value we find upon a measurement of Y. In other words, we may expect that the locality principle holds. Moreover, the EPR argument strongly suggested that, from a HVT point of view, it is natural to conceive that the correlations between the

[104] Bell actually quotes Bohr in order to support the contextuality of a HVT: '[non-locality demands] are seen to be quite unreasonable when one remembers with Bohr "the impossibility of any sharp distinction between the behavior of atomic objects and the interaction with the measuring instruments which serve to define the conditions under which the phenomena appear"' (ibid, 477). Bell's invocation of Bohr in order to justify contextuality in a HVT, to whom hidden-variables would have been anathema, is qualified by A. Shimony and D. Mermin as "a judo-like maneuver" (See Mermin 1993, 811, footnote 23).

measurement results on the space-like separated subsystems in the singlet state are grounded on the value of a hidden parameter determining the state of the total system when the interaction took place. That is, the argument wrought the hope that a HVT would reject any form of non-locality for the correlations that the standard formalism predicts for entangled states.

John Bell became originally interested in the possibility and features of HVTs through Bohm's 1952 proposal. This HVT is explicitly non-local, so he wondered whether *any* consistent and empirically viable HVT must be so:

> It must be stressed that, to the present writer's knowledge, there is no *proof* that *any* hidden variable account *must* have this extraordinary character. It would therefore be interesting, perhaps, to pursue some further "impossibility proofs", replacing the arbitrary axioms objected above [additivity postulate, general non-contextuality] by some condition of locality, or of separability of distant systems. (Bell 1966, 452)

In his groundbreaking 1964 paper entitled *On the Einstein Podolsky Rosen Paradox*[105] Bell formulated this proof. In the following decades, an enormous variety of different proofs of Bell's theorem have been introduced. I shall now reconstruct a very simple one due to E. Wigner (1970). Consider once again the singlet state of a pair of particles that we already met in the context of the EPR argument, but now we consider three different spin directions α, β and γ, so that we can write $|\phi\rangle = \frac{1}{\sqrt{2}}(|\alpha_+\rangle_1|\alpha_-\rangle_2 - |\alpha_-\rangle_1|\alpha_+\rangle_2)|\phi\rangle = \frac{1}{\sqrt{2}}(|\beta_+\rangle_1|\beta_-\rangle_2 - |\beta_-\rangle_1|\beta_+\rangle_2) = |\phi\rangle = \frac{1}{\sqrt{2}}(|\gamma_+\rangle_1|\gamma_-\rangle_2 - |\gamma_-\rangle_1|\gamma_+\rangle_2)$. Let us presume that an element of reality corresponds to the spin components along each of the three directions: a *beable* represented by a hidden parameter ξ determines, at the moment of interaction, the correlations between the spin values of particles 1 and 2 along each direction, so that each particle has a definite spin value for all three directions that cannot be affected by measurements performed in the other (space-like separated) particle. That is, we assume locality. Accordingly, we have eight possible spin value assignments that respect the correlations, so that they may have a non-zero probability:

$$\begin{aligned} p(1) &= p(\alpha_{+_1}, \beta_{+_1}, \gamma_{+_1}, \alpha_{-_2}, \beta_{-_2}, \gamma_{-_2}) & \qquad p(5) &= p(\alpha_{-_1}, \beta_{+_1}, \gamma_{+_1}, \alpha_{+_2}, \beta_{-_2}, \gamma_{-_2}) \\ p(2) &= p(\alpha_{+_1}, \beta_{+_1}, \gamma_{-_1}, \alpha_{-_2}, \beta_{-_2}, \gamma_{+_2}) & \qquad p(6) &= p(\alpha_{-_1}, \beta_{+_1}, \gamma_{-_1}, \alpha_{+_2}, \beta_{-_2}, \gamma_{+_2}) \\ p(3) &= p(\alpha_{+_1}, \beta_{-_1}, \gamma_{+_1}, \alpha_{-_2}, \beta_{+_2}, \gamma_{-_2}) & \qquad p(7) &= p(\alpha_{-_1}, \beta_{-_1}, \gamma_{+_1}, \alpha_{+_2}, \beta_{+_2}, \gamma_{-_2}) \\ p(4) &= p(\alpha_{+_1}, \beta_{-_1}, \gamma_{-_1}, \alpha_{-_2}, \beta_{+_2}, \gamma_{+_2}) & \qquad p(8) &= p(\alpha_{-_1}, \beta_{-_1}, \gamma_{-_1}, \alpha_{+_2}, \beta_{+_2}, \gamma_{+_2}) \end{aligned}$$

The probability of a state in which the spin values that correspond to $|\alpha_+\rangle_1$ and $|\beta_+\rangle_2$ is given by $p(\alpha_{+_1}, \beta_{+_2}) = p(3) + p(4)$. Analogously, it holds that $p(\beta_{+_1}, \gamma_{+_2}) = p(2) + p(6)$ and $p(\alpha_{+_1}, \gamma_{+_2}) = p(2) + p(4)$. It is easy to see that the following inequality commonly known as 'the Bell-Wigner inequality', follows:

$$p(\alpha_{+_1}, \gamma_{+_2}) \leq p(\alpha_{+_1}, \beta_{+_2}) + p(\beta_{+_1}, \gamma_{+_2})$$

Now, the formula that quantum mechanics gives us for joint (+) probabilities along two spin directions j and k in the singlet state is $p(j_{+_1}, k_{+_2}) = \frac{1}{2}\sin^2\frac{\theta}{2}$, where θ is the angle subtended by j and k[106]. Let us

[105] The 1966 paper I have referred to, *On the Problem of Hidden Variables in Quantum Mechanics*, was written by Bell in 1964, but due to an unfortunate course of events, it got misfiled and published only in 1966. Bell obtained the proof he speculates about in the 1966 paper shortly after he wrote it.

[106] This formula follows from simple vector-spaces considerations. The operator for a basis θ-angle rotation in a two-dimensional space has the form $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. In the case of spin-$\frac{1}{2}$, we have the constraint that, if the angle $\widehat{jk}$ is $\pi = 180°$ and the spin value of the state vector in the j-direction basis is $+1$ (in units of $\frac{1}{2}\hbar$), then the spin value of the state vector in

suppose that α, β and γ lie on the same plane and that $\widehat{\alpha\beta} = \widehat{\beta\gamma} = \frac{\pi}{3}$ and that $\widehat{\alpha\gamma} = \frac{2\pi}{3}$. Then, according to the Bell-Wigner inequality that we obtained under the locality assumption, it follows that $\frac{1}{2}\sin^2\frac{\pi}{3} \leq 2\left(\frac{1}{2}\sin^2\frac{\pi}{6}\right)$, or that $\frac{3}{8} \leq \frac{1}{4}$, which is absurd. That is, if locality is assumed, then the spin correlations in the table above and the Bell-Wigner inequality follow, but this last result is incompatible with the predictions of quantum theory[107]. In simple words, SQM violates the Bell-Wigner inequality, but a local HVT respects it. In turn, this means that (i) the decision between a local HVT and SQM can be done in terms of empirical evidence, and that (ii) any HVT that is capable to reproduce the empirical predictions of SQM must be non-local. Regarding (i), there is general agreement in that the 'cascade emission' experiments performed by A. Aspect's team in 1982 ruled out local HVTs. Concerning (ii), as already mentioned, in 1952 David Bohm formulated a HVT that is EE to SQM and that is immune to the KS theorem and to Bell's theorem.

We have dealt with the historical and conceptual contexts that determined the HVT project, and we have also made explicit and analyzed what are the formal and physical requirements that any theoretically and empirically and viable HVT must observe. Now we can move on to the specific HVT that interests us: Bohm's quantum theory.

3.5 BOHM'S QUANTUM THEORY

David Bohm introduced his quantum theory in 1952 (Bohm 1952). From the last two sections we know that any viable HVT must be contextual and non-local. Both features are, at least from the point of view of classical physics, rather weird. Therefore, it is clear that if something is gained with the formulation of a HVT, it will not be so by means of a full restoration of the classical picture in the quantum realm. In their 1993 book, *The Undivided Universe*, D. Bohm and B. J. Hiley explain what, in their opinion, constitutes the most important virtue of Bohm's quantum theory—this is not its *only* virtue, of course, but it certainly helps to appreciate the essential *spirit* of the theory. Bohm and Hiley evaluate the HVT at issue in explicit contrast with the dominant (Copenhagen) interpretation of the standard theory. According to these authors, under the Copenhagen interpretation SQM provides an ontologically determinate picture of the physical world only in measurement contexts, but it remains silent in their absence (or provides only a

---

the k-direction basis must be -1. Therefore, the suitable rotation operator for a spin-$\frac{1}{2}$ state vector is $\begin{pmatrix} \cos\theta/2 & \sin\theta/2 \\ -\sin\theta/2 & \cos\theta/2 \end{pmatrix}$. Let us consider the singlet state $|\varphi\rangle = \frac{1}{\sqrt{2}}(|j_+\rangle_1|j_-\rangle_2 - |j_-\rangle_1|j_+\rangle_2)$. If we want to know the spin correlations between particle 1 along direction j and particle 2 along direction k, we simply use the rotation operator just defined to obtain $|j_+\rangle_2 = \cos\frac{\theta}{2}|k_+\rangle_2 - \sin\frac{\theta}{2}|k_-\rangle_2$ and $|j_-\rangle_2 = \sin\frac{\theta}{2}|k_+\rangle_2 + \cos\frac{\theta}{2}|k_-\rangle_2$. Plugging these two expressions in $|\varphi\rangle$, one obtains $|\varphi\rangle = \frac{1}{\sqrt{2}}\sin\frac{\theta}{2}|j_+\rangle_1|k_+\rangle_2 + \frac{1}{\sqrt{2}}\cos\frac{\theta}{2}|j_+\rangle_1|k_-\rangle_2 - \frac{1}{\sqrt{2}}\cos\frac{\theta}{2}|j_-\rangle_1|k_+\rangle_2 + \frac{1}{\sqrt{2}}\sin\frac{\theta}{2}|j_-\rangle_1|k_-\rangle_2$. From this expression is quite clear that $p(j_{+_1}, k_{+_1}) = \frac{1}{2}\sin^2\frac{\theta}{2}$.

[107] As to the conceptual relation between Bell's and the KS theorem, Bub writes that 'it appears that Bell's theorem is a stronger version of the Kochen and Specker theorem: any hidden variable theory satisfying the Kochen and Specker constraint [contextuality] will necessarily satisfy the locality condition, but not conversely. Putting it differently, the Kochen and Specker theorem shows that the value assigned to an observable A in a hidden variable theory will in general have to depend on what other observables are measured together with A. That is, the value of A will depend on the complete experimental arrangement in which A is measured, or on the measurement context. Bell's theorem establishes that the value assigned to A must depend on the complete experimental arrangement, even when two alternative arrangements differ only in a region space-like separated from the region in which A is measured' (Bub 1997, 78-9). These remarks get clearly illustrated in a proof of both theorems by means of the same set of ten spin and spin-function observables for three particles in an eight-dimensional Hilbert space due to N. Mermin (1990, 1993) (based on a previous proof of Bell's theorem introduced by Greenberger, Horne and Zelinger (1989)).

rather indeterminate picture). The standard formalism, Bohm and Hiley state, provides us with an algorithm to calculate the probabilities of experimental results, and in this sense, it gives us statistical knowledge of how our measuring instruments work when they interact with quantum systems—it is a theory that is concerned more with our knowledge than with reality itself:

> To put it in more philosophical terms, it may be said that quantum theory is primarily directed towards *epistemology* which is the study that focuses on the question of how we obtain our knowledge (and possibly on what we can do with it).
> It follows from this that quantum mechanics can say little or nothing about reality itself. In philosophical terminology, it does not give what can be called an *ontology* for a quantum system. Ontology is primarily concerned with that which *is* and only secondarily with how we obtain our knowledge about this. (Bohm & Hiley 1993, 2)

This sketchy characterization of the Copenhagen interpretation may not be totally fair and accurate[108], but it gives us a good on the essential spirit of Bohm's theory. As we will see, the latter provides us with a *definite ontology* of the realm of physical reality it applies to, an ontology in which measurements do not play any *sui generis* role. Bohm's own characterization in his 1952 original paper states that his theory

> permits us to conceive of each individual system as being in a precisely definable state, whose changes with time are determined by definite laws, analogous to (but not identical with) the classical equations of motion. Quantum-mechanical probabilities are regarded (like their counterparts in classical statistical mechanics) as only practical necessity and not as a manifestation of an inherent lack of complete determination in the properties of matter at the quantum level. (Bohm 1952, 166)

3.5.1 Formalism, ontology, and the meaning of probability

The formalism of BQT) results from a simple but clever manipulation of the Schrödinger equation $i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi$ (this time I write it in three dimensions using the Laplacian operator $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$). The general polar form of the solutions of this equation is given by $\Psi = Re^{(iS/\hbar)}$, where R is the amplitude of the wave function and S its phase. Plugging this expression on each side of the Schrödinger equation, and then equating the real and imaginary parts, Bohm obtained

$$\frac{\partial S}{\partial t} = -\frac{(\nabla S)^2}{2m} - V + \frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} \qquad \frac{\partial R}{\partial t} = \frac{1}{2m}(R\nabla^2 S + 2\nabla R\cdot\nabla S)$$ [109]

Concerning the first formula, we have that the main equation of classical mechanics in its Hamilton-Jacobi formulation is $\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V = 0$, where S is a field function and V is the potential acting on the particle. This equation determines the trajectories of moving particles[110]. Comparison readily suggests, Bohm noticed, that an analogous quantum equation can be posited. Defining a 'quantum potential'

---

[108] Besides, the *modal* interpretations of the standard formalism do provide a (comparatively more) definite ontology of the quantum realm.

[109] First, we plug $Re^{iS/\hbar}$ in the time dependent side of the Schrödinger equation to obtain $i\hbar\frac{\partial\Psi}{\partial t} = i\hbar e^{iS/\hbar}\frac{\partial R}{\partial t} - Re^{iS/\hbar}\frac{\partial S}{\partial t}$. Doing the same in the time independent side, we get $-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = -\frac{e^{iS/\hbar}}{2m}[\hbar^2\nabla^2 R - R(\nabla S)^2 + 2i\hbar\nabla R\nabla S + i\hbar R\nabla^2 S] + VRe^{iS/\hbar}$. The next step is to equate the imaginary and real parts in the right hand side of both equations to obtain $i\hbar e^{iS/\hbar}\frac{\partial R}{\partial t} = -e^{iS/\hbar}(2i\hbar\nabla R\nabla S + i\hbar R\nabla^2 S)$ and $-Re^{iS/\hbar}\frac{\partial S}{\partial t} = -\frac{e^{iS/\hbar}}{2m}[\hbar^2\nabla^2 R - R(\nabla S)^2] + VRe^{iS/\hbar}$. Solving these last two equations for $\frac{\partial R}{\partial t}$ and $\frac{\partial S}{\partial t}$, respectively, we obtain Bohm's formulas above.

[110] Momentum is defined as $p = \partial S/\partial q$, which means that at each instant the direction of the momentum of a particle with position coordinates q is orthogonal to S (defined at that instant). Accordingly, the velocity vector is given by $\mathbf{v}(\mathbf{x}) =$

$$Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}$$

Bohm obtained

$$\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V + Q = 0$$

Just as the Hamilton-Jacobi equation determines the trajectory of a classical particle, the equation that Bohm obtained could be interpreted as determining the trajectory of a quantum particle. The concept of a particle trajectory is foreign to SQM. Therefore, and in contrast, we can already appreciate the 'definite ontology spirit' of Bohm's theory: the position and momentum of quantum particles are defined and determinate at all times. The quantum Hamilton-Jacobi equation, however, implies an important modification of the classical picture, for the motion of the particle is not only affected by the classical potential V, but also by the quantum potential Q:

> The first step in developing this interpretation in a more explicit way is to associate with each electron a particle having precisely definable and continuously varying values of position and momentum. The solution of the modified Hamilton-Jacobi equation defines an ensemble of possible trajectories for this particle, which can be obtained from the Hamilton-Jacobi function, $S(\mathbf{x})$, by integrating the velocity, $\mathbf{v}(\mathbf{x}) = \nabla S(\mathbf{x})/m$. The equation for S implies, however, that the particles move under the action of a force which is not entirely derivable from the classical potential, $V(\mathbf{x})$, but which also obtains a contribution from the "quantum mechanical" potential $[Q](\mathbf{x}) = (-\hbar^2/2m) \times \nabla^2 R/R$. The function, $R(\mathbf{x})$, is not completely arbitrary, but is partially determined in terms of $S(\mathbf{x})$ by the differential equation $\left[\frac{\partial R}{\partial t} = \frac{1}{2m}(R\nabla^2 S + 2\nabla R \cdot \nabla S)\right]$. Thus R and S can be said to codetermine each other. (Bohm 1952, 170)

Now, since the force determining the motion of a particle is given by Q, it depends on R (and also on S, for R and S codetermine each other). In turn, since both R and S are given by Ψ, it follows that, besides the particle, the ontology of the theory considers Ψ as representing an objectively real field. Therefore, Q is the quantum potential corresponding to such a field. The basic ontological furniture of Bohm's theory is thus given by the particle and a quantum field whose potential 'guides' the particle trajectory. In this sense, Bohm's proposal is rather close to the theory that de Broglie sketched in the 1920s:

> Since the force on a particle now depends on a function of the absolute value, $R(\mathbf{x})$, of the wave function, $\Psi(\mathbf{x})$, evaluated at the actual location of the particle, we have effectively been led to regard the wave function of an individual electron as a mathematical representation of an objectively real field. This field exerts a force on the particle in a way that is analogous to, but not identical with, the way in which an electromagnetic field exerts a force on a charge, and a meson exerts a force on a nucleon. (Bohm 1952, 170)

It is important to spell out some essential properties of the 'guiding wave' and its corresponding quantum potential in order to appreciate their peculiar nature—especially from the point of view of waves and potentials as conceived in classical mechanics and electrodynamics. As Bohm points out, the quantum wave field is analogous to usual fields—such as electromagnetic ones—in the sense that it partially determines the forces that affect the trajectory of a particle. Besides, the time dependent Schrödinger equation determines the value of the field at all times, so that, given the initial position and momentum of a particle, we can calculate its entire trajectory. However, the analogy with usual wave fields does not go further. The quantum wave field does not have any sources, for the Schrödinger equation does not

$\nabla S(\mathbf{x})/m$. By means of these expressions the trajectory of a particle gets defined. For a detailed presentation and assessment of the Hamilton-Jacobi version of classical mechanics, see (Holland 1993, chapter 2).

contain any corresponding terms. Moreover, the field is not in any way affected by the particle it guides, and in this sense it is not subject to Newton's second law.

Another difference between the classical case and Bohm's theory consists in that the quantum potential depends on the form, not on the amplitude, of the quantum wave field. The reason is simple. The expression $Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}$ contains the amplitude function R both in the numerator and denominator, so that if Ψ is multiplied by an arbitrary constant, this factor gets canceled out. Consequently, the way in which the quantum potential affects the trajectory of particles does not depend on its intensity, but only on its form. Therefore, the action of a quantum potential does not diminish or vanish with increasing distance. In classical mechanics, for example, a cork in a water pool bobs proportionally to the distance to the center of the wave (provoked by a falling stone, for example). If the wave is a Bohmian quantum field, the cork can bob with full strength even if it is far away from the place where the stone falls.

In order to make sense of this feature of the quantum potential, Bohm and Hiley (1993, section 3.2) propose an interpretation of the wave field and the quantum potential in terms of information. Consider a ship moving in automatic pilot whose trajectory is determined by radar waves. The trajectory of the ship is determined by the form of the incoming radar waves, not by their intensity. Actually, the energy and intensity of the radar waves may be negligible compared to the kinetic energy of the ship, so it is not the radar wave 'pushing' on the ship what determines its trajectory, the ship moves under its own energy. However, it is clear that the trajectory of the ship is indeed determined by the form of the incoming radar waves. Bohm states that an analogy with radar waves provides an intelligible picture of the quantum field and the quantum potential. The interaction between the potential and the particles cannot be classically described: the wave is not affected back by the particles, and its effect on trajectories depends on its form, not on its intensity. However, the form of the quantum wave field defines the 'active information' by means of which the quantum potential determines the trajectory of its associated particle. But this is just an analogy, of course. We can readily notice that in the case of the ship and the radar waves, the former finds out the form of the latter by means of a classical interaction—the ship 'classically detects' the incoming radar wave and so determines its shape. This is not so in the case of the particle and the quantum field, the way in which the particle is 'informed' by the quantum potential cannot be described in terms of a classical interaction.

Now that the basic ontology of the theory has been addressed, we can take a look at how probabilities enter in Bohm's theory. We know that the wave field and quantum potential derived from the Schrödinger equation deterministically determine the trajectories of quantum particles given certain initial conditions (position and momentum). However, these initial conditions (because of reasons that I will consider when I deal with measurements in Bohm's theory) cannot be precisely determined. That is, while the wave field and quantum potential determine the possible trajectories that particles in that field can take, which precise trajectory is the case depends on the specific initial position and momentum of the associated particle. But since we cannot precisely know these initial conditions, we only have epistemic access to the probability that a particle is in a region in which the value of the wave function Ψ is not 0 or undefined. It is important to underscore that in this theory the probability refers to a particle to *be* in a specific place, rather than to the possibility of *finding* a particle in a specific place upon a *measurement*.

In Bohm's theory, this strictly epistemic probability is grounded on the relative frequencies according to which an ensemble of particles corresponding to a single wave function Ψ are spatially distributed over the regions in which the value of Ψ is not 0 or undefined. Consider again the formula $\frac{\partial R}{\partial t} = \frac{1}{2m}(R\nabla^2 S + 2\nabla R \cdot \nabla S)$. By setting $R^2 = P$, this equation can be expressed as

$$\frac{\partial P}{\partial t} + \nabla \cdot \left(P\frac{\nabla S}{m}\right) = 0$$

This formula can be interpreted as stating conservation of probability in analogy with the Hamilton-Jacobi formulation of classical mechanics[111]. In this theory, an ensemble of particles can be associated with S, and the velocities and momentum of such particles can be determined *via* $\mathbf{v}(\mathbf{x}) = \frac{\nabla S(\mathbf{x})}{m}$. That is, if we know the initial position of a particle, we can determine its full trajectory. However, if we cannot precisely determine those initial conditions (like in statistical mechanics), we can still define a function $\rho(\mathbf{x})$ that represents the relative frequency with which we find particles in the region defined. We can consider a volume Ω fixed in space, so that the total number of particles contained in this volume is given by $\int_{\Omega}\rho \mathrm{d}^3x$ [112]. It holds that in the Hamilton-Jacobi theory the formula $\frac{\partial \rho}{\partial t} + \nabla \cdot \rho\mathbf{v} = 0$ can be derived, and it clearly represents the conservation of probability (as relative frequency of location distribution), that is, if at an initial time the distribution function $\rho_0(\boldsymbol{x})$ is specified, the same function defines the location (probability) distribution at any other time.

We can apply the same line of reasoning in Bohm's theory. If we interpret $R^2 = P = |\Psi|^2$ as the function that expresses the location (probability) distribution of particles belonging to an ensemble associated to a single wave function Ψ, then $\frac{\partial P}{\partial t} + \nabla \cdot \left(P\frac{\nabla S}{m}\right) = 0$ expresses conservation of probability. If for the ensemble of particles the probability distribution P holds at an initial time, it holds at any other time. Notice that, just as in the Hamilton-Jacobi formulation of classical mechanics, it is assumed that $\mathbf{v}(\mathbf{x}) = \frac{\nabla S(\mathbf{x})}{m}$. Now, if in Bohm's theory the probability density P so described is equal to $|\Psi|^2$ [113], it follows that the empirical predictions of the theory are equivalent to the empirical predictions in SQM[114]. However, the meaning of the probability involved is quite different. In SQM (at least in Bohr's and von Neumann's views) probability is an objective feature of physical reality, whereas in Bohm's theory is only a matter of ignorance:

> Let us now consider the meaning of the assumption of a statistical ensemble of particles with a probability density equal to $P(\mathbf{x}) = R^2(\mathbf{x}) = |\Psi(\boldsymbol{x})|^2$. From Eq. $\left[\frac{\partial P}{\partial t} + \nabla \cdot \left(P\frac{\nabla S}{m}\right) = 0\right]$, it follows that this assumption is consistent, provided that Ψ satisfies Schrödinger's equation, and $\mathbf{v} = \nabla S(\mathbf{x})/m$. This probability density is numerically equal to the probability density of particles obtained in the usual [Copenhagen] interpretation. In the usual interpretation, however, the need for a probability description is regarded as inherent in the very structure of matter [...], whereas in our interpretation, it arises [...] because from one measurement to the next, we cannot in practice predict or control the precise location of a particle, as a result of corresponding unpredictable and uncontrollable disturbances introduced by the measuring apparatus. Thus, in our interpretation, the use of a statistical ensemble is (as in the case of classical statistical mechanics) only a practical necessity, and not a reflection of an inherent limitation on the precision with which it is correct for us to conceive of the variables defining the state of the system. (Bohm 1952, 171)

---

[111] See (Holland 1993, section 2.5)

[112] This presupposes an actual ensemble of many particles, of course. If we want to discuss the situation as involving one particle, then $\int_{\Omega}\rho \mathrm{d}^3x$ represents the averaged number obtained after repeated measurements, and $\rho(\mathbf{x})$ represents probability of finding the single particle in the corresponding region.

[113] Since the quantum potential depends on the form and not on the amplitude of the wave function, there is no problem in normalizing Ψ, so that $\int P(\mathbf{x})\mathrm{d}^3x = 1$.

[114] Holland (1993, sections 3.5 and 3.6.4) proves that assuming $R^2 = P = |\Psi|^2$ in Bohm's theory, then the expectation value of an observable, just as in SQM, is given by $\langle A\rangle_W = \mathrm{Tr}(WA)$, both in the pure case $W = |\Psi\rangle\langle\Psi|$ and in the mixed state $W = \sum_i p_i|\Psi_i\rangle\langle\Psi_i|$. To stress out that the position representation has a fundamental meaning in Bohm's theory, the pure and mixed cases can be represented as $W(\mathbf{x},\mathbf{x}') = \Psi(\mathbf{x})\Psi^*(\mathbf{x}')$ and $W(\mathbf{x},\mathbf{x}') = \sum_i p_i\Psi_i(\mathbf{x})\Psi_i^*(\mathbf{x}')$, respectively. Regarding the (proper) mixed state case it is interesting to note that in Bohm's theory 'the density matrix formalism is therefore a particular type of *statistical mechanics of waves and [particles]*. Each element in the ensemble of individual systems comprises a wave Ψ_i and an associated particle whose momentum is given by $\mathbf{p}_i = \nabla S_i$. Because $[W]$ represents a fictitious ensemble all the waves may be considered to occupy, simultaneously, overlapping regions of space without interfering. Only one wave and one particle is present in any one trial. The density matrix describes both an 'ensemble of ensembles' of particles, and an ensemble of waves and particles' (1993, 104).

A question that naturally comes up at this point is why the probability density function (as location distribution) P must be equal to $R^2 = |\Psi|^2$. There is no *a priori* reason for this 'distribution postulate' or 'quantum equilibrium' hypothesis to hold. The answer that Bohm himself proposed (Bohm 1953; Bohm & Hiley 1993, chapter 9) consists in that given an ensemble of particles associated to a wave function Ψ, 'under typical chaotic conditions that prevail in most situations an arbitrary probability distribution P, will approach and remain equal to $|\Psi|^2$, the latter being an equilibrium distribution. The relationship between P and $|\Psi|^2$ is in this way seen to be contingent' (Bohm & Hiley 1993, 41). The basic idea can be explained by means of an example. Consider a box 'filled' with a linear combination of stationary wave functions, with each wave function associated with an ensemble of particles. Bohm and Hiley (1993, section 9.2) argue that the chaotic conditions determining the system cause that the location distribution of the particles quickly approaches $|\Psi|^2$. Now, if we open the box and let the particles get out, and make them pass through a collimator and velocity selector, we will obtain ensembles of particles all of which correspond to the same wave function. Since R^2 is conserved, and since P has already approached $|\Psi|^2$, we know that the probability distribution of the particles is $P = |\Psi|^2$. This is of course the case for a pure state. If the particles are selected by a method that does not allow us to specify a single associated wave function, we obtain a mixed state. And whereas in the pure case the expectation value of an observable is calculated by integrating over all the positions of the particle in the ensemble, in the mixed case the expectation value has to be calculated by integrating over the distribution of wave functions, in addition to averaging over the particle positions for each wave function. Bohm and Hiley (1993, section 9.3) show that this operation leads to the expression $\mathrm{Tr}(WA)$, as we may expect.

Now that both the ontology of the theory and the meaning of probability have been addressed, we can consider a list of basic postulates (Holland 1993, section 3.1) that specify the essentials of Bohm's proposal:

> *Postulate 1*. An individual physical system comprises a wave propagating in space and time together with a point particle which moves continuously under the guidance of the wave.
>
> *Postulate 2*. The wave is mathematically described by $\Psi(\mathbf{x})$, a solution of the Schrödinger wave equation.
>
> *Postulate 3*. The particle motion is obtained through the equation $\mathbf{v}(\mathbf{x}) = (1/m)\nabla S(\mathbf{x})$, with $\mathbf{x} = \mathbf{x}(t)$ and where S is the phase function of Ψ. To solve this equation we have to specify the initial condition $\mathbf{x}(0) = \mathbf{x}_0$. This specification constitutes the only extra information introduced by the theory that is not contained in $\Psi(\mathbf{x})$. An ensemble of possible motions associated with one single wave function is generated by varying $\mathbf{x}_0$.
>
> *Postulate 4*. The probability that a particle in the ensemble is located between the points $\mathbf{x}$ and $\mathbf{x} + \mathrm{d}\mathbf{x}$ at a time t is given by $P(\mathbf{x}, t)\mathrm{d}^3\mathbf{x}$, where $P = R^2 = |\Psi|^2$.[115]

3.5.2 Contextuality

So far we have seen that in Bohm's theory quantum particles possess a definite position and momentum at all times, so the concept of a particle trajectory is well-defined. The formalism of the theory, derived from the Schrödinger equation, allows to deterministically predict a particle trajectory, provided that we know its initial position. However, we cannot know with full precision what is the location of a

[115] Notice that postulates 1-3 already give us a consistent quantum theory of motion. Postulate 4 ensures that the motions of particles in an ensemble are in agreement with the predictions of SQM.

particle, for, as we will see in due time, the nature of measurement interactions preclude it. Therefore, the formalism of Bohm's theory provides us with statistical knowledge—we can only know the probability that a particle is at a certain location, at a certain time. The way in which probabilities enter the theory simply represents our ignorance of the precise initial conditions, not an inherent feature of the physical world. All of this sounds very classical. Actually, we may think that Bohm's theory is to SQM as statistical mechanics is to thermodynamics. This is true up to a certain extent, but there are essential features in BQT which are blatantly non-classical. We have already seen that the quantum wave and its associated quantum potential are entities that are radically different to their classical counterparts. Now I will deal with two other essential features of BQT that are clearly non-classical: contextuality and non-locality. Both properties are quite expected, of course. We saw in section 3.4 that Bell's theorem and the K-S theorem are proofs that any empirically viable HVT must necessarily be both contextual and non-local.

To clarify the way in which contextuality holds in Bohm's theory[116], consider the squared-spin observables for spin-1 particles that we already met in section 3.4.3, namely, the squared-spin (mutually commuting) operators S_x^2, S_y^2, S_z^2. Let $[S_z^2]_{H_{xyz}}$ to represent the outcome obtained for the squared-spin property in the z-direction of a system if measured via $H_{xyz} = aS_x^2 + bS_y^2 + cS_z^2$, and let $[S_z^2]_{H_{x'y'z}}$ to represent the outcome obtained for the same property when measured via $H_{xyz} = aS_x^2 + bS_y^2 + cS_z^2$, with a, b, c distinct real numbers, and where x, y, z and x', y', z define mutually orthogonal directions. For simplicity, let us assume that $a = 1$, $b = -1$, and $c = 0$, so that we can write down $H = S_x^2 - S_y^2$ and $H' = S_{x'}^2 - S_{y'}^2$. Now, since for any given orthogonal triad $\{x, y, z\}$ it holds that $\left(S_x^2 - S_y^2\right)^2 = S_z^2$[117], it follows that $H^2 = H'^2 = S_z^2$. That is, we can measure the value of the spin-squared property in the z-direction via an H-measurement or via an H'-measurement—notice that H and H' are non-commuting operators. Now, a non-contextuality condition is that it must always be the case that $[S_z^2]_H = [S_z^2]_{H'}$. However, since we know that the K-S theorem imposes a contextuality constraint on any viable HVT, in BQT it must be possible that $[S_z^2]_H \neq [S_z^2]_{H'}$.

Property $\mathcal{H}$ can be measured by passing the particle through a suitable inhomogeneous magnetic field—so a device similar to a Stern-Gerlach apparatus can be used. To describe the measurement, we denote the initial composite system, before the particle-magnetic field interaction and with the particle in position representation, by $\psi(q, 0) = \phi(q) \sum_n c_n |H = n\rangle$, where $\phi(q)$ is a narrow wave packet symmetric about $q = 0$, and $|H = n\rangle$ is an eigenstate of H with eigenvalue $n = -1, 0,$ or 1. The measurement interaction between the magnetic field and the particle, that occurs during the time interval $[0, T]$, is described by the application of a Hamiltonian of the form $H_{int} = gi^{-1}(\partial/\partial q)H$, where $g(t)$ is a coupling constant with value nonzero only during the interaction, and q is the position component of the particle that is associated to the value of H. This application gives us the during-measurement state $\psi(q, t) = \sum_n c_n \phi(q - gnt)|H = n\rangle$, with $t \in [0, T]$. With a suitable choice for the value of $g(t)$, at the time T (the end of the interaction) gT will be significantly larger than the width of the wave packet $\phi(q)$, so that the overlap between the adjacent wave packets $\phi(q - gnT)$ is negligible, and the H-value of the particle after its post-measurement deflection is discernible—a definite (no interference effects present) H-value of the particle gets effectively correlated with the particle's position q. Finally, from Bohm's probability conservation formula the equation of motion $\frac{dq}{dt} = \frac{J}{P}$ can be derived, where $P = |\psi(q, t)|^2$ is the probability distribution and $J = \psi^*(q, t) g H \psi(q, t)$, so that $\frac{dq}{dt} = \frac{g \sum_n n |c_n|^2 \phi(q - gnt)^2}{\sum_n |c_n|^2 \phi(q - gnt)^2}$. This equation can be solved for the different values of the initial position q of the particle. That is, for every allowed trajectory there is a corresponding definite H-value for the particle and a corresponding probability $|c_n|^2$. Consequently, the probability to obtain, for example, a value -1 is given by the sum of the probabilities of all the trajectories

[116] In this exposition I follow (Pagonis & Clifton 1995).

[117] This can be easily confirmed by inspecting the squared-spin operators presented in section 3.4.3.

to which $n = -1$ correspond. It is clear that the value obtained in the H-measurement is deterministically defined by the initial position of the particle.

In the case of an H' measurement, this time applying a suitable interaction Hamiltonian H'_{int}, the same analysis can be run in order to obtain the state $\psi(q', t) = \sum_n c_n{}'\phi(q' - gnt)|H' = n\rangle$, and an equation of motion $\frac{dq'}{dt} = \frac{g\sum_n n|c_n{}'|^2\phi(q'-gnt)^2}{\sum_n|c_n{}'|^2\phi(q'-gnt)^2}$, where the $c_n{}'$ are coefficients in the expansion of the same initial state $\psi(q, 0)$—but this time over H' eigenstates instead—, and q' is the position coordinate that gets correlated with the H'-value during the measurement interaction. This result presupposes that ψ has the same form in the q direction as in the q' direction—i.e., ϕ—and that the H' measurement is of the same strength g and of the same duration T as the H-measurement.

Since the equation of motion $\frac{dq}{dt} = \frac{J}{P}$ is deterministic, the same position coordinate cannot belong to two different trajectories[118]. Consequently, different possible trajectories cannot cross the q-axis in an H-measurement, and different allowed trajectories cannot cross the q' axis in an H'-measurement. This means that in an H-measurement, after a time $t \geq T$, the trajectories that end up at one of the three possible regions of final q-positions $-gt, 0, gt$ (for $n = -1,0,1$) started in one of the three possible q-regions in the initial wave packet $\phi(q)$, ordered from negative to positive values of q. By the same token, in an H'-measurement, after a time $t \geq T$, the trajectories that end up at one of the three possible regions of final q'-positions $-gt, 0, gt$ started in one of the three possible q'-regions in the initial wave packet $\phi(q')$, ordered from negative to positive values of q'.

Now we can clearly appreciate the contextuality of Bohm's theory. The $|c_n|^2$ give us the probability of finding a certain H-value in an H-measurement given ψ, and the $|c_n{}'|^2$ give us the probability of finding a certain H'-value in an H'-measurement given the same ψ. If the theory were non-contextual, the value of property $\mathcal{S}_z^2$ should be the same regardless of whether we measure it via an H- measurement or via an H'-measurement. However, since $[H, H'] \neq 0$, it holds that, in general, the c_n will differ from the $c_n{}'$. This means that the fractions $|c_{-1}|^2, |c_0|^2, |c_{+1}|^2$ of the initial positions in the q-direction that end up in the final positions $q = -gt, q = 0, q = gt$ (corresponding to measurement results $H = -1, H = 0, H = 1$), are, in general, different from the fractions$|c_{-1}{}'|^2, |c_0{}'|^2, |c_{+1}{}'|^2$ of the initial positions in the q'-direction that end up in the final positions $q' = -gt, q' = 0, q' = gt$ (corresponding to measurement results $H' = -1, H' = 0, H' = 1$). Therefore, given a particle associated to ψ in a certain initial position, the value we find for the property $\mathcal{S}_z^2$ may be different depending on whether we measure it via H or via H'. As Pagonis and Clifton put it: 'the ψ-field will be affected differently (more precisely, bifurcate differently) in each measurement context due to the differing measurement interaction Hamiltonians needed to measure H and H', and so will, in general, affect a given initial position differently' (1995, 294).

In Bohm's theory, the only properties that neatly correspond to the notion of a (categorical) classical property are position and momentum (we may add properties that get completely specified by a function of momentum and position, though, such as kinetic energy). On the other hand, the physical properties that are subject to contextuality in Bohm's theory can be interpreted, with respect to their ontological status, in two different ways. According to the first view, as Jeffrey Bub states,

> What we call a 'measurement' of a spin component of an electron in a Stern-Gerlach measurement, say, simply catalogues a certain characteristic movement of a particle in the presence of an inhomogeneous magnetic field, which reflects a particular way in which a certain sort of ψ-field can evolve in configuration

[118] Since velocity changes only as a function of position, if two trajectories intersect at a point they have a matching position—and therefore, velocity—and by determinism, the rest of both trajectories should also match. That is, they are one single trajectory.

> space and guide the motion of the particles. This suffices to account for all quantum phenomena, insofar as these phenomena can be characterized by changes in the positions of particles. (1997, 168-7).[119]

The second interpretation assigns spin a status analogous to position and momentum, that is, as an observable that always has a definite value: 'a spin-component observable can take determinate values that are not eigenvalues of spin in states that are not spin eigenstates, but the account of spin measurements shows that the measured value of spin is always an eigenvalue of spin' (Bub 1997, 169)[120]. This analysis can be extended to other contextual properties, of course (see Bohm & Hiley 1993, chapter 10; Holland 1993, chapters 9-10).

3.5.3 Non-locality

So far I have restricted the exposition of Bohm's theory to the one-particle case. As we will see now, the formalism of the theory applied to the many-body case clearly results in non-locality. The Schrödinger equation for an n-partciles system is given by $i\hbar\frac{\partial\Psi}{\partial t} = \left[\sum_{i=1}^{n}\left(\frac{-\hbar^2}{2m_i}\right)\nabla_i^2 + V(\mathbf{x}_1, \dots, \mathbf{x}_n, t)\right]\Psi$. Using $\Psi = Re^{iS/\hbar}$ (this time $R(\mathbf{x}_1, \dots, \mathbf{x}_n, t)$ and $S(\mathbf{x}_1, \dots, \mathbf{x}_n, t)$) we get

$$\frac{\partial S}{\partial t} + \sum_{i=1}^{n}\frac{(\nabla_i S)^2}{2m_i} + V + Q = 0 \qquad\qquad \frac{\partial P}{\partial t} + \sum_{i=1}^{n}\nabla_i \cdot \left(P\frac{\nabla_i S}{m_i}\right) = 0,$$

where $P = R^2$ is again interpreted as the probability distribution[121], and

$$Q = \sum_{i=1}^{n} -\frac{\hbar^2}{2m_i}\frac{\nabla_i^2 R}{R}$$

It follows that, according to the 'Hamilton-Jacobi analogy', the momentum of the i^{th} particle in the system is given by $\mathbf{p}_i = \nabla_i S(\mathbf{x}_1, \dots, \mathbf{x}_n, t)$, so that the equation of motion is now $\frac{d\mathbf{x}_i}{dt} = \mathbf{v}_i(\mathbf{x}_1, \dots, \mathbf{x}_n, t) =$

---

[119] Pagonis and Clifton elaborate on this view and characterize contextual properties as *dispositional*, under a reductionist interpretation of dispositional properties. That is, they conceive a contextual property, such as $\mathcal{S}_z^2$, as a dispositional property that ultimately reduces to a categorical property (position) in concomitance with a specific measurement context (the Hamiltonian interaction corresponding to H or H'): 'the property that "grounds" the measurement result is not the one "measured" but the categorical property to which (in addition to the context) the dispositional property is reducible to' (Pagonis & Clifton 1995, 286).

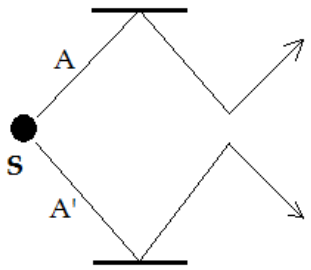
[120] More precisely, 'According to this approach a particle in a 'spin up' state with respect to some z axis has a spin vector which points along this z direction; the x and y components of the spin vector are not undetermined but actually zero. However, if the spin component is measured using a Stern-Gerlach apparatus oriented, for example, along the x direction the result 'up x' and 'down x' will be found with equal frequency in this state. A calculation shows that in this case the beam bifurcates along a central plane perpendicular to the analyzing direction as two separating 'up' and 'down' beams are formed. The outcome in a particular case is determined by the uncontrollable actual position of the particle relative to this bifurcation plane at the entrance slit to the field. The particle enters one beam or the other as a result of the action of a spin-dependent 'quantum force' and as the beams separate a 'quantum torque' rotates the spin vector to lie either along or opposed to the direction of the analysing field. In this way the quantum phenomena associated with spin can be understood in a manner closer, in some ways, to our customary forms of description than is usually the case, which highlights the essential differences between quantum and classical phenomena. Such a description is possible since the particle and the spinor wave are assumed to have equal ontological status' (Dewdney, Holland & Kiprianidis 1987, 4717).

[121] In the many-body case, the 'distribution postulate' (postulate 4) states that the probability that at time t particle 1 lies in the volume element d^3x_1 around the point $\mathbf{x}_1$, particle 2 lies in the volume element d^3x_2 around the point $\mathbf{x}_2$, etc., is given by $R^2(\mathbf{x}_1, \dots, \mathbf{x}_n)d^3x_1 \dots d^3x_2$. Equation $\frac{\partial P}{\partial t} + \sum_{i=1}^{n}\nabla_i \cdot \left(P\frac{\nabla_i S}{m_i}\right) = 0$ ensures that if this probability distribution holds at a certain time, then it holds at all times. The meaning of probability in the many-body case is, of course, the same as the meaning in the one-body case.

$\frac{1}{m_i}\nabla_i S(\mathbf{x}_1, \ldots, \mathbf{x}_n, t)$. Consequently, the trajectories of the particles get defined in a $3n$-dimensional configuration space. We can now clearly notice the non-local character of the theory through the fact that in order to determine the position $\mathbf{x}_i(t)$ of any particle at time t, we need to specify the positions of all the other particles in the system at time t. A different choice of initial position in one particle in the system entails a different subsequent motion for all the other particles.

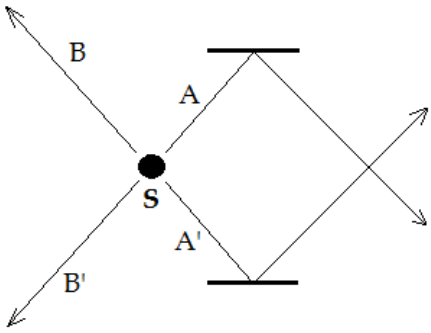
That the particles are guided in their trajectories in a correlated way becomes even more clear when we look at the many-body quantum potential Q. Consider a system of two particles, so that the corresponding quantum potential is given by $Q = -\frac{\hbar^2}{2m}\frac{(\nabla_1^2+\nabla_2^2)R}{R}$. Even if the classical potential V is zero, that is, if there is no classical interaction between the particles, their motion is still determined in a correlated way by Q, for R is a function of the positions of *both* particles. Moreover, since, as we saw above, the quantum potential depends on the form, not on the amplitude, of the wave field, there is no spatial limit for the correlation that the potential determines. That is, if a system of two particles is associated to a wave function Ψ, and the particles are correlated by a quantum potential Q, then, no matter how distant the particles may be—say, particle 1 in Santiago de Chile and particle 2 in the Andromeda galaxy—their trajectories are mutually correlated. This kind of correlation, in turn, has a notable consequence. Given an n-particle system correlated by a quantum potential Q, if we disturb one particle in the system, localized in a certain region of three-dimensional space, then the configuration space wave as a whole will be altered and all the other particles in the system will be affected instantaneously, no matter how far they are in three-dimensional space from the particle we disturbed. Thus, the non-locality of Bohm's theory is due to three closely related features, namely, the spatially unlimited extension of the action of the quantum potential, the dependence of the instantaneous position of a particle on the positions of all the other particles in the system at that same instant, and the response of the whole system to localized disturbances.

A nice and simple geometric illustration of how non-local effects manifest in Bohm's theory has been offered by D. Rice (1997). Consider the state $\psi = \frac{1}{\sqrt{2}}(|A\rangle + |A'\rangle)$ expressed in momentum basis, so that A and A' can be considered as the two possible trajectories for the corresponding particle. As mentioned above, given the determinism of Bohm's equation of motion, the trajectories cannot cross. Suppose that the particle is emitted from a source S. Two mirrors are set in a way such that the trajectories A and A' get 'reflected'. Since the trajectories cannot cross, they seem to 'bounce' off each other as depicted in the figure:

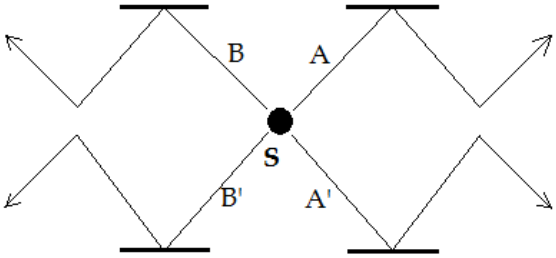


Consider now a two-particle system in the state $\varphi = \frac{1}{\sqrt{2}}(|A\rangle|B\rangle + |A'\rangle|B'\rangle)$. This is an entangled state, if particle 1 takes the A trajectory, then particle 2 takes the B trajectory (the same holds for A' and B', of course). This time the trajectories are defined in a 6-dimensional configuration space, so the no-crossing condition holds in configuration space. This means that it cannot be the case that *both* A an A', and B and B', cross in position space. If only the former cross, the different coordinates of particle 2 along B and B' still determine two different trajectories in configuration space. But if both A and A', and B and B', cross,

then the trajectories do have a point in common in the 6-dimensional configuration space they are defined in. Consider now a setup in which only A and A' are reflected by mirrors. In this case, trajectories A and A' cross in position space, as we can see in the figure:



But now we change the experimental setup and include reflecting mirrors along B and B'. This means that both B and B', and A and A', will be reflected towards each other, but we know that the trajectories cannot cross. Therefore, the change of the setup has a non-local effect on trajectories A and A': the inclusion of mirrors implies that trajectories A and A' now bounce off each other. We can make the non-locality even more apparent by setting the experiment in a way such that the events corresponding to the reflection of B and B' are space-like separated from the event corresponding to the bouncing off of A and A', for example[122]:



The analysis of non-locality in Bohm's theory leads us to two further strange features regarding the quantum wave field. We saw above that its ontological status is rather peculiar. The quantum wave guides the particle(s) associated to it, but it cannot be affected back by the particles, and the guidance is not the outcome of a classical interaction. Now we have seen that in the n-particles case, the wave field is defined in a $3n$-configuration space, not in our familiar 3-dimensional space. This reinforces the idea that it is a very non-classical entity. The quantum field *in configuration space* cannot be regarded as a usual field that carries energy and momentum, part of which is transferred to the particles, determining their trajectories. But then what kind of entity is it and how does it guide the particles? There is certainly no similar entity in the ontology of any other physical theory. Even if we accept Bohm's analogy with the 'information radar wave', the analogy does not give us a more detailed answer. All we can say is that it is a guiding wave defined in configuration space, period.

A second peculiar feature concerning the ontology of the theory and the quantum wave field is given by the essential holism inherent in Bohm's theory. Both in the one-body and in the many-body case, the quantum potential Q is determined by the quantum state Ψ, and it is not a preassigned function of the position coordinates of the associates particle(s). In classical physics, in contrast, an interparticle potential

[122] Since Rice's example intends to be a geometric illustrative example, all the dynamic features and assumptions are omitted. F. M. Toyama and K. Matsuura (2006) elaborate on the same example and make explicit all its dynamical aspects.

$V(\mathbf{x}_i, \mathbf{x}_j)$ connecting two particles with coordinates $\mathbf{x}_i, \mathbf{x}_j$ is uniquely specified by the properties of the particles under consideration. In Bohm's theory, the quantum potential $Q(\mathbf{x}_i, \mathbf{x}_j)$ is not uniquely specified by the position coordinates of the particles[123]. The conclusion that follows from these remarks is that 'we may say that whereas in classical dynamics the whole is the sum of the parts and their interactions, in quantum mechanics the whole is prior to the parts (particles) and its properties cannot be explained by a kind of superposition of the properties of the parts' (Holland 1993, 282)[124]. But if the whole is more than the sum of the parts, and if the whole is ontologically more fundamental than the parts, the question rises of how is it possible that we can have a scientific description of the world in terms of independent physical systems:

> the above –described feature [wholeness] should, in principle, apply to the entire universe. At first sight this might suggest that we could never disentangle one part of the universe from the rest, so that there would be no way to do science as we know it or even to obtain knowledge by the traditional method of finding systems that can be regarded as at least approximately isolated from their surroundings'. (Bohm & Hiley 1993, 59)

The answer to the question is given by a consideration of the conditions for 'non-locality'. Consider a two-particle system that can be expressed as the product of two wave functions, that is, $\psi(\mathbf{x}_1, \mathbf{x}_2) = \psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2)$. If this is the case, we call the wave function ψ 'strictly factorizable'. If the wave function associated to a two-particle system is strictly factorizable, the particles are independent, there are no non-local features involved. This is so because in a strictly factorizable wave function like ψ the phase and amplitude functions are obviously given by $S(\boldsymbol{x}_1, \boldsymbol{x}_2) = S_A(\mathbf{x}_1) + S_B(\mathbf{x}_2)$ and $R(\boldsymbol{x}_1, \boldsymbol{x}_2) = R_A(\boldsymbol{x}_1)R_B(\boldsymbol{x}_2)$, respectively. It follows that the quantum potential is $Q(\boldsymbol{x}_1, \boldsymbol{x}_2) = Q_A(\boldsymbol{x}_1) + Q_B(\boldsymbol{x}_2)$, where $Q_A = -\frac{\hbar^2}{2m_1}\frac{\nabla_1^2 R_A}{R_A}$ and $Q_B = -\frac{\hbar^2}{2m_2}\frac{\nabla_2^2 R_B}{R_B}$; and that $\mathbf{v}_1(\mathbf{x}_1) = \frac{\nabla_1 S_A}{m_1}$ and $\mathbf{v}_1(\mathbf{x}_1) = \frac{\nabla_2 S_B}{m_2}$. This means that particle 1 is affected only by Q_A and its trajectory is independent of, and can be determined without reference to, the trajectory of particle 2—and the other way around, of course.

Consider now the wave function $\varphi(\mathbf{x}_1, \mathbf{x}_2) = N[\varphi_A(\mathbf{x}_1)\varphi_B(\mathbf{x}_2) + \varphi_C(\boldsymbol{x}_1)\varphi_D(\boldsymbol{x}_2)]$, where N is a normalization constant. If the summands do not overlap (it is sufficient that either φ_A and φ_C, or φ_B and φ_D, have no points in common, no 'common support'), then φ is called 'effectively factorizable', for it behaves as if the wave is either $\varphi = \varphi_A(\mathbf{x}_1)\varphi_B(\mathbf{x}_2)$ or $\varphi = \varphi_C(\mathbf{x}_1)\varphi_D(\mathbf{x}_2)$. That is, the wave function behaves as a mixture: if measurements are performed one or the other summand obtains with a certain probability. However, if the particles described by ψ classically interact, the time evolution of the Schrödinger equation will, in general, turn ψ into a wave function of the form φ in which the summands do overlap (that is, $\varphi_A \cap \phi_C \neq \emptyset$ and $\varphi_B \cap \phi_D \neq \emptyset$), so that the resulting φ is not factorizable. In this case the probability

---

[123] Actually, in the one-particle case, it is this property—that the quantum potential Q is not uniquely determined by the particle position coordinates—that allows us to speak of an ensemble of particles associated to one single wave function, as we did above. The same holds, *mutatis mutandis*, in the many-boy case. An ensemble of correlated pairs of particles with different coordinate positions at a certain instant can correspond to the same wave function.

[124] Bohm and Hiley underscore the same point in a more philosophical vein: 'the relationship between parts of a system [...] implies a new quality of *wholeness* of the entire system going beyond anything that can be specified solely in terms of the actual spatial relations of all the particles. This is indeed the feature which makes the quantum theory go beyond mechanism of any kind. For it is the essence of mechanism to say that basic reality consists of the parts of a system which are in a preassigned interaction. The concept of the whole, then, has only a secondary significance, in the sense that it is only a way of looking at certain overall aspects of what is in reality the behavior of the parts. In our interpretation of the quantum theory, we see that the interaction of parts is determined by something that cannot be described solely in terms of these parts and their preassigned relationships. Rather it depends on the many-body wave function (which, in the usual interpretation, is said to determine the quantum state of the system). This many-body wave function evolves according to Schrödinger's equation. Something with this kind of dynamical significance that refers directly to the whole system is thus playing a key role in the theory. We emphasize that *this is the most fundamentally new aspect* of the quantum theory' (Bohm & Hiley 1993, 58-9).

distribution is given by $R^2 = N^2\{R_A^2R_B^2 + R_C^2R_D^2 + 2R_AR_BR_CR_D\cos[(S_A + S_B - S_C - S_D)/\hbar]\}$, where the third 'interference' term is responsible for the nonlocal effects—in an effectively factorizable wave function, the term vanishes. When φ is not factorizable, thus, we cannot say that particle 1(2) is associated only either with φ_A or φ_C (φ_B or φ_D). That is, the two-particle system is holistically associated to $\varphi(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and the trajectories of the particles are non-locally correlated. Hence, the condition for non-locality is non-factorizability. On the other hand, when a system in a non-factorizable state interacts with its environment, the interaction generally leads to a suppression (at least for all practical purposes) of the interference terms and to effective factorizability. In other words, decoherence plays a similar role in Bohm's theory as in SQM.

3.5.4 Bohm's proposal as a *rival* theory

Bohm's seminal paper is entitled *A Suggested Interpretation of Quantum Theory in Terms of "Hidden" Variables*. The fact that the expression *interpretation of quantum theory* appears in the title may suggest that we are dealing not with a rival EE theory to SQM, but with yet another interpretation that we should classify along with the different proposals revised in section 3.3. Actually, the expression *Bohm's interpretation of QM* is common both in the philosophical and physical literature. However, most of the times this is just an expression (motivated maybe by Bohm's own choice of words), for it is usually considered (explicitly or implicitly) that Bohm's work constitutes an alternative theory, not just an interpretation of SQM. To mention just a symptomatic example, Jeremy Butterfield comments on the expression at issue that 'I fear its name is off-putting: it has such a wealth of physical ideas and formalism that it deserves the name 'causal theory'. But who am I to re-name another person's creation?' (1992, 60).

But if we accept that Bohm's proposal constitutes a rival theory then we may ask for a criterion that distinguishes it as such with respect to SQM. The main ground for such a distinction is the introduction of further conceptual content—which in turn refers to ontological components—that is not present in the standard formalism. That is, the introduction of 'hidden variables' is enough to determine that we are dealing with two different theories. I think that the rationale of this feature as a criterion that establishes that we have two theories instead of two interpretations is not merely formal. The relevant point is that the introduction of a hidden variable commits us to an ontology that is different from the one 'depicted' by SQM. This rivalry criterion relies on the fact that, from the Bohmian point of view, the formalism of SQM is an incomplete representation of physical systems. Whereas all the interpretations of such a formalism agree in that nothing else is needed to represent a physical state apart from a Hilbert space vector or a density operator, Bohm's theory tells us that something else is needed, namely, the location of a quantum particle associated to a wave-function.

This additional element results in a clearly different ontology than the one that corresponds to any of the interpretations of the standard formalism. More particularly, we have already seen that in Bohm's theory particles have well defined trajectories at all times, trajectories that are in turn determined by the quantum wave and its corresponding potential. In SQM, the concept of a particle's trajectory is not well defined in all cases, as we saw in the examples of Heisenberg's reflections on the electron cloud chamber and of Bohr's account of the double-slit experiment[125]. As Richard Healey states,

> There is a more basic reason why I cannot accept a hidden variable theory as an interpretation of quantum mechanics. A hidden variable theory is, fundamentally, a separate and distinct theory from quantum mechanics. To offer such a theory is not to present an interpretation of quantum mechanics but to change the

[125] I invite the reader to ponder the contents of section 3.6.1 below also as spelling out this feature as a criterion to determine that Bohm's is a rival theory to SQM.

> subject [...]. A hidden variable theory incorporates quantities additional to the quantum dynamical variables. (1989, 24)[126]

The predictive equivalence between SQM and BQT is assured by the distribution postulate $P = R^2 = |\Psi|^2$, as we saw above. On the other hand, that SQM and BQT are rival theories is determined by the fact that the latter is a HVT, that is, because introduces some conceptual machinery, not included in the former, that refers to fundamental components of the ontology purported. Thus, we have set up a case that illustrates the problem of EE in UD in real life physics[127].

3.6 BOHM'S THEORY VERSUS SQM

Now that the essentials of Bohm's theory have been presented, we can move on to a comparative evaluation with SQM. Since the latter theory is clearly the dominant one in the scientific community, I will proceed by assuming that the burden of persuasion lies on Bohm's proposal. That is, I will approach the comparison by focusing on its virtues and flaws in order to find out whether Bohm's theory is a viable alternative to SQM. I have deliberately omitted the analysis of two important foundational issues of Bohm's theory in the previous section, namely, the account of measurement processes and the classical limit. The reason is that there are good reasons to think that Bohm's approach scores better than SQM with respect to such features, so they determine two important virtues that are relevant in the comparative evaluation. Therefore, I will address both issue in the present section.

3.6.1 Definite ontology and understanding

As mentioned above, Bohm's own account of the basic features and virtues of his theory is that it represents the quantum realm by means of a clear and determinate ontology of particles and waves. On

[126] Healey considers two further reasons to conceive Bohm's as a different theory from SQM. The first one is that 'hidden variable theories are held to underlie quantum mechanics in a way similar to that in which classical mechanics underlies the *distinct* theory of statistical mechanics' (1989, 24). I think this is not a really good reason. On the one hand, it may be more correct to say that BQT underlies SQM as statistical mechanics underlies phenomenological thermodynamics. On the other hand, Healey's view does not grasp the *rivalry* between the theories—neither the view that the analogy is to statistical mechanics and thermodynamics. Healey's second reason is that 'a hidden variable (at least typically) is held to be empirically equivalent to quantum mechanics only with respect to a restricted range of conceivable experiments, while leading to conflicting predictions concerning a range of possible further experiments which may, indeed, be extremely hard to actualize' (ibid., 25). Healey is not explicit about it, but I think he refers to Bohm's own comments on the possibility of the hidden variable as leading to diverging predictions in the 'fundamental length' of the order $10^{-13}\ cm$ (see Bohm 1952, 168-9). Bohm's comments on predictive divergence at this scale have not played a central role in the philosophical and physical discussion of the theory—maybe because BQT has not been developed in a way such there is some clarity regarding what the theory has (may have) to say in the order of $10^{-13}\ cm$ —, but Healey does have a point in that the very possibility of diverging predictions in certain contexts (just recall Laudan and Leplin's argument about the role of auxiliary assumptions and new measurement instruments) shows that BQT and SQM are two different theories.

[127] This line of thought may actually lead us to state that the different interpretations of SQM are also *rival theories*. Anyways, this rivalry is of a different nature. I agree with F. Muller (2013) in that these interpretations consist in modifications of the meaning of certain terms, and in the addition of other postulates, with respect to what he calls *minimal quantum mechanics* (QM_0)—which roughly consists in the postulates 1-5 reviewed in section 3.2. QM_0 fails to provide us with an essential element that we expect from scientific theories: a clear (physical) description that tells us that the world is thus and thus. QM_0 is thus not a fully constituted theory, hence the need for interpretation. Then, the rivalry between the different interpretations is given by a debate about what is the (best) theory that can be obtained from QM_0. On the other hand, BQT is not a completion of QM_0, it stands by itself as a fully constituted theory. Thus, we could have two levels of rivalry between quantum theories. First, the rivalry between the interpretations which consists on the debate about which is the (best) theory that can be obtained out of QM_0, and the rivalry between these theories—the interpretations of SQM—and BQT.

the other hand, SQM renounces the possibility of such a representation. Recall the 'symbolic' meaning that Bohr attributed to the wave function as a description of a state in the absence of measurements—quantum theory does not provide us with an ontological description of the physical processes that the wave function represents, unless we consider a certain measurement context that allows us to use classical concepts. All we have is a formalism that assigns probabilities for possible experimental outcomes[128]. Even if we take a more literal stance toward the meaning of the wave function, that is, if we take it as not only *representing* but also as *describing* physical systems—this seems to be von Neumann's approach—we find that the description provided is ontologically indefinite. Just recall the principle of superposition. If we accept that a superposed state objectively describes a physical system for a certain physical property, then we cannot have a definite ontological picture of that system with respect to that property. Furthermore, in Bohr's view—and also in the 'literal von Neumann' approach—the wave function is a complete representation of physical reality. That is, a detailed and definite description of quantum phenomena according to the standards of classical physics is just not possible. There is nothing else to say about physical systems beyond the information that the wave function provides us with. In Bohr's own words, 'in quantum mechanics we are not dealing with an arbitrary renunciation of a more detailed analysis of atomic phenomena, but with a recognition that such an analysis is *in principle* excluded' (1949, p.)

Bohr's view about the impossibility of a more detailed description is rather dogmatic. Anyhow, whatever the arguments for this position may be, the very existence of Bohm's theory is a clear refutation. As Bohm himself stated, the usual [Bohr-von Neumann] interpretation of quantum theory

> requires us to give up the possibility of even conceiving precisely what might determine the behavior of an individual system at the quantum level, without providing adequate proof that such a renunciation is necessary. The usual interpretation is admittedly consistent; but the mere demonstration of such consistency does not exclude the possibility of other equally consistent interpretations, which would involve additional elements or parameters permitting a detailed causal and continuous description of all processes, and not requiring us to forego the possibility of conceiving the quantum level in precise terms. (Bohm 1952, 168)

The ontologically definite, casual and continuous description of quantum phenomena that Bohm's theory provides is clearly illustrated in its account of the double-slit experiment. Recall that in Bohr's interpretation, what occurs between the electron source and the detection screen cannot be ontologically described, but only symbolically represented by the wave-function[129]. If we try to determine the spatio-temporal behavior of the electron, we need to use a different, mutually exclusive experimental setup in which the interference pattern cannot be observed. In the case of a literal interpretation of the wave function, we have that the system between the emission and detection is described by a superposed state, so that we cannot conceive the electron as having a determinate ontology regarding position and momentum. In neither case a continuous spatio-temporal description can be proposed—the concept of a trajectory is simply meaningless in SQM.

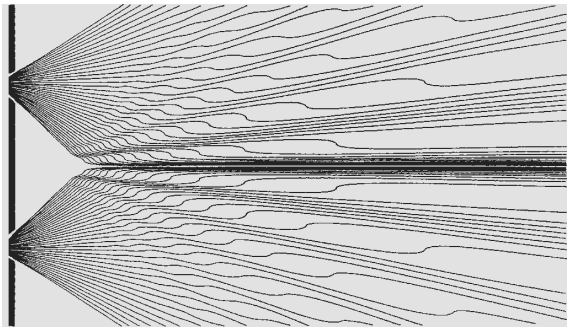
In Bohm's theory, the explanation of the interference effect includes a clear, determinate and continuous spatio-temporal description of the whole process. Consider an ensemble of electrons associated to a wave function ψ. The quantum wave travels from the emission source to the detection screen and gets separated in two parts as it goes through the two slits in the diaphragm. As a result, the two parts of ψ interfere with each other and the resulting quantum potential determines a set of possible trajectories

[128] Recall that these assertions are not meant as expressing an instrumentalist philosophy.

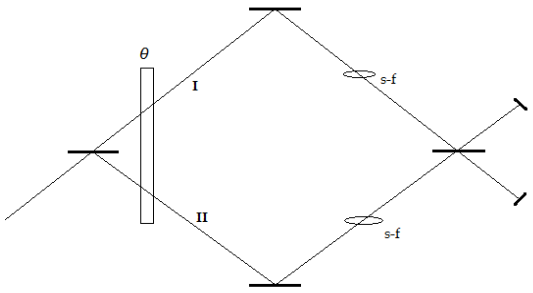
[129] One may think that in Bohr's view, what happens between electron emission and detection in an interference effects setup can be described by means of a wave-picture. As I explained above, I agree with Plotnisky (2013, Ch. 6) in that such a view is inconsistent with Bohr's symbolic conception of the meaning of the wave function in the absence of measurements. The wave-like behavior of the electron manifests, collectively, *in the detection screen* through the interference pattern exhibited, not in the electron's 'flight'. The same holds for the particle-like behavior of the electron that manifests in the detection screen in a complementary experimental setup (like measuring momentum interchange in the diaphragm).

that exhibits the interference pattern on the detection screen[130]. As explained above, the precise position of each electron in the ensemble completely determines which of the possible trajectories the electron actually follows. We do not know the exact position, but since $P = R^2$ represents the location distribution of the electrons in the ensemble at all times, we know that the probability distribution of the electrons on the detection screen is given by $|\psi|^2$, just as in SQM. In simple words, electrons pass one at a time accompanied by their quantum guiding wave. Each electron goes through one slit, but the wave passes through both. Given the recombination of the two parts of the wave, the quantum potential determines a collection of trajectories that reflect the interference effect.

Since we are considering a one-body example, the possible trajectories cannot cross in position space. This is clearly reflected in the figure, there is an axis of symmetry that the trajectories cannot cross. In turn, this means that if a particle is detected in the zone below (above) this axis, then it went through the lower (upper) slit in the diaphragm. That is, unlike in SQM, we do have the 'which slit information' even in a case where an interference pattern is present—because 'trajectory' is a well-defined concept in BQT. Finally, if one of the slits is closed, or if the diaphragm is loosely fixed in order to measure a momentum interaction, the resulting quantum potential will have a different form so that no interference effect would be exhibited in the possible trajectories it determines[131].



There is another experimental context in which a comparative evaluation between the explanations provided by SQM and BQT is most interesting, namely, neutron interferometry. Consider a beam of spin-up polarized neutrons in the z-direction described by the wave function $\varphi = |+_z\rangle$. The beam enters the interferometer and is split by a crystal in two partial beams φ_I and φ_{II}. Each of these beams is partially reflected by identical crystals. A phase shift θ between the beams is produced by a piece of aluminum. Spin flipper coils, which we can turn on or off at will, are set in the path of both beams. The beams recombine and get transmitted or diffracted by yet another identical crystal, and are finally measured in detectors. Finally, we assume that the setup is such that the experiment can be carried out neutron-by-neutron. That is, there is never more than one single neutron within the interferometer.

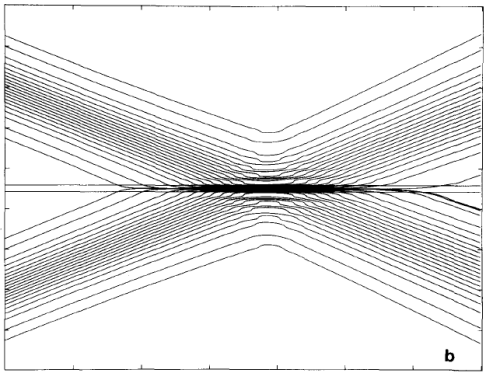


[130] The locations at which ψ is zero or undefined are 'forbidden', of course.

[131] The picture, based on the exact calculations, originally appeared in (Philippidis, Dewdney & Hiley, 1979). This graphic representation of the account of the double-slit experiment in Bohm's proposal helped to revive the interest in the theory. However, Bohm himself had already provided an equivalent verbal explanation in the original paper (1952, 173-4).

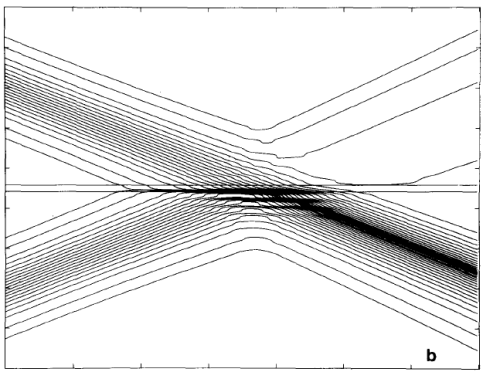
Suppose first that the spin flippers are switched off. This means that there is a phase difference given by an angle θ between the partial beams $\varphi_I = e^{i\theta}|+_z\rangle$ and $\varphi_{II} = |+_z\rangle$. Therefore, when they recombine, an interference pattern is produced and it can be measured in the neutron detectors. The situation is quite analogous to the double-slit experiment. In Bohr's view, we cannot have a continuous spatio-temporal account. If the neutron actually goes either along path I or along path II, no interference pattern would arise. We may simply live with the wave-like behavior manifested in the neutron detectors. There is no such thing as the trajectory of the neutron within the interferometer.

In Bohm's theory we do have a determinate kinematic explanation—analogous to the case of the double-slit experiment, of course. The neutron travels along one single path, *either* I or II, but its guiding wave travels along both. When the two parts of the wave recombine, the interference effect produced by the phase shift reflects in the form of the quantum potential and the resulting possible trajectories that it determines. Which of these trajectories the neutron actually follows is determined by its initial position. When the phase shift is given by $\theta = 0, \pi$, that is, when the partial beams are in phase, the following picture describes how the possible trajectories are determined in the recombination-interference region:



Notice that a quantum potential barrier takes the form of an axis of symmetry, for the trajectories cannot cross in position space.

When $\theta = \pi/2$ the result is different. Since the partial wave φ_I 'arrives first', we may say, almost all of its potential trajectories are 'transmitted', whereas all the trajectories of φ_{II} get 'reflected'. Though a quantum potential barrier divides the φ_I and the φ_{II} so that they do not cross, given the phase shift the barrier does not take the form of an axis of symmetry.



Two variations of this experiments illustrate further the difference between SQM and BQT regarding the account of interference phenomena they provide. First, suppose that only the radio-frequency spin flipper coil in II is switched on. That is, there is a one-photon energy transfer between φ_{II} and the spin flipper. The partial waves get described thus by $\varphi_I = e^{i\theta}|+_z\rangle$ and $\varphi_{II} = e^{i\frac{\Delta E}{\hbar}}|-_z\rangle$. Let the initial state of

the spin flipper to be ψ_i and its final state to be ψ_f. The final total state is then $\Psi_f = \psi_i a\varphi_I + \psi_f b\varphi_{II}$. Given the small value of ΔE, $\psi_i \approx \psi_f$, so that we can write $\Psi_f = \psi_i(a\varphi_I + b\varphi_{II})$. That is, the interaction between φ_{II} and the spin flipper does not count as a measurement in the usual sense and the interference pattern obtains in the experiment result. Since Ψ_{II} is given by a superposition of spin-up and spin-down in the z-direction, the spin of the final beam is polarized in the xy-plane—the precise polarization depends on the coefficients a and b, of course.

Since interference effects are involved, the Bohrian interpretation tells us that we cannot use the particle picture to describe what happens within the interferometer. Again, we have to live with the fact that the neutrons collectively exhibit a wave behavior when they reach the detectors, and we must accept that there is no such thing as the trajectory of a neutron in the interferometer. The interesting point of this variation—that puts some extra pressure on Bohr's interpretation—is that it is difficult to see how we may make sense of the energy interaction with the spin flipper if we cannot conceive the neutron as localized during its trip in the interferometer. Vigier et al. state that this experimental setup jeopardizes Bohr's view because, according to this interpretation

> a particle cannot exist in one beam [...] and take part in interference. However in order to describe the functioning of the coil we must use the complementary localized particle aspect. The energy transfer that takes place giving rise to the change of [φ_{II}] is described [...] in terms of photon exchange between the neutron and the field in the coil. Thus the neutron is conceived as a particle in one beam to explain energy transfer and simultaneously as a wave existing in both to explain interference. The complementarity of wave and particle descriptions is broken; both aspects must be used simultaneously in one and the same experimental arrangement. Complementary description is thus incomplete, or can energy be exchanged with a probability wave?' (Vigier et al. 1987, 186-7)

As I explained above, I think it is a mistake to attribute a wave description to what happens between the emission source and the detector in interference experimental setups. However, if, unlike Vigier et al., we hold on to the view that in Bohr's interpretation the wave picture holds only for the experimental result observed in the interference pattern exhibited in the detection screen, the problem they point out comes up anyway. To make sense of the interference pattern observed in the detection screen we use the wave picture, but in this case it is very difficult—if possible at all—to make sense of the energy transfer without referring to the *complementary* particle picture. But, again, if we use such a picture and assume that the neutron traveled along path II and interacted with the spin flipper, we get prevented from understanding how the interference pattern is generated. In this context, the mutual exclusion between the particle and the wave pictures impedes us to grasp a fully intelligible account of the process.

Unlike Vigier et al., I think that complementarity is not broken, though, for the simultaneous use of the wave and particle picture is not consistent with the observed results. The point is that the wave picture that accounts for the interference pattern cannot make sense of the energy interaction featured in the experiment. That is, the difficulty for Bohr's interpretation that this experiment exhibits is connected to the *completeness* of the description. This is not exactly shocking news for the Bohrian, of course. We have already seen that the doctrine of complementarity consists exactly in that a classically complete description cannot be obtained. Experimental setups in which interference effects are shown are not suitable to be described in continuous, determinate kinematical terms. The interesting bonus of this example is that the renunciation to the kinematical description implies an extra price to be paid: it is not possible to make sense of a dynamical aspect that features in the process: the energy interaction with the spin flipper.

In the case of BQT, on the other hand, the explanation follows again from the wave *and* particle ontological description. The particle travels either along path I or path II, but the wave goes along both. The interaction of the partial wave φ_{II} with the spin flipper is such that the resulting quantum potential determines that the spin of the particle will be polarized along the xy-plane:

> A spin-half particle is conceived as a localized entity surrounded by a real spinor wave [...]. While the particle really travels one way (path I or II), the spinor wave propagates in both paths. In path II the interaction with the [radio-frequency] spin flipper inverts the spinor symmetry of the wave while in path I the initial state is maintained. What happens in the interference region can now be represented by the action of a spin dependent quantum potential Q and a quantum torque τ which can be shown to produce a time-dependent spinor symmetry in the xy plane. The particle travelling, for example, in path I is constrained by the spinor symmetry in the interference region and its $+z$ spin is twisted into the xy plane by the quantum torque. If it travels along path II it suffers an additional inversion due to the rf coil, yielding this energy to the coil while in the intersection area its spin $-z$ is twisted again to the xy plane. Consequently, a coherent picture is established which accounts for both particle and wave aspects. (ibid., 187)[132]

Finally, we can consider what happens when both spin flippers are switched on. Both φ_I and φ_{II} are affected by the magnetic field associated to each spin flipper, so that their states get described by $\varphi_I = e^{i\theta}|-_z\rangle$ and $\varphi_{II} = |-_z\rangle$. Interference effects result from the recombination of both beams, but this time the spin polarization points down in the z-axis. The interesting feature of this setup is once again that the outgoing neutrons have lost an amount ΔE of energy due to the interaction with the radio-frequency spin flippers. From what has been said so far it is easy to see how a comparison between the accounts of SQM (in the Bohr-von Neumann interpretation) and BQT, so I leave the exercise to the reader.

James Cushing (1991; 1994, chapters 2 and 5) states that the conclusion to be drawn from this analysis is that BQT scores much better than SQM with respect to an essential and basic goal of science, the achievement of explanations that provide *understanding*. Cushing describes three distinct goals of science associated to three epistemic levels: *empirical adequacy*, *explanation* and *understanding*. The first level goal is accomplished simply when we have formal algorithms that allow us to get the numbers right, that save the phenomena. The second goal is accomplished when we have a consistent formalism from which empirically adequate formal algorithms can be deduced—Cushing explicitly assumes the nomological-deductive conception of explanation. Finally, '*understanding* is possible once we have an interpretation of the formalism that allows us to comprehend and to know the character of the phenomena and of the explanation offered' (Cushing 1991, 338).

In the case of SQM *vs*. BQT we have that the corresponding formalisms that provide the n-d explanations of the equations that correctly predict experimental outcomes are mathematically equivalent. We can present the formalism in terms of Hilbert spaces or in terms of Bohm's S and R functions—in both cases the Schrödinger equation plays the main role. The interpretations of the formalism are essentially different, though—an in this context the choice between Hilbert spaces and S and R functions becomes relevant. The interpretation(s) of the Hilbert space formalism is (are) given by the postulates we revised in section 3.2, and by the heuristic variations explained in 3.3; whereas the interpretation of Bohm's functions is given by the ontology of particles and guiding waves explained in 3.5—ontology that is grounded on the similarity between the S and R functions and the Hamilton-Jacobi presentation of classical mechanics.

In Cushing's view, the analysis of interference experimental setups illustrates that with respect to the goal of understanding, BQT is a much better theory than SQM. He assumes that 'the paradigm of an explanation that *can* (or may) produce understanding [...] for physical processes is a causal explanation, consisting either of direct cause-effect between phenomena and events or of a common cause located in the past collection of phenomena under consideration' (ibid.). It is clear that the explanations that BQT

[132] Notice that this explanation presupposes the 'objective' interpretation of spin described above. It can be adapted to the alternative interpretation, of course. In this case the particle does not intrinsically possess a spin, but the quantum potential Q is such that the particle's interaction with a magnetic field results in a trajectory behavior that can be described as spin-up or down in a certain direction. Regardless of which path the particle actually took, the interaction of the partial wave φ_{II} with the spin flipper determines that the quantum potential of the final state, after the interference zone, is such that the spin to be measured would result in xy-plane polarization. There would be a different result if the particle's spin were measured in the z-direction just *before* the interference zone, though. If the particle went along path I $|z_+\rangle$ would be obtained, whereas if it traveled along path II the result would be $|-_z\rangle$.

provides for the double-slit experiment and for neutron interferometry experiments are causal in the first sense that Cushing explains. The conclusion he draws is that BQT includes explanations that correspond to the paradigm of understanding-bearer explanations; whereas it is difficult to see in what sense the explanations of SQM provide any understanding at all[133]. We may now take a look at how good this argument is.

I think that this argument can be challenged in two ways. First, we may question the view that SQM, even in the Bohr-von Neumann framework, is poor in providing understanding. One of the ontological and epistemic main conclusions that the founding fathers of the theory drew (from the Bohr-von Neumann framework point of view) was that the physical world that the theory depicts is *not* causal in the classical sense, that is not fully describable by the classical vocabulary, and that indeterminism and probability constitute some of its objective and fundamental features. It is true that Bohr and others took it that this view is a logical and necessary conclusion that follows from the theory, and that the very existence of Bohm's theory denies such *necessity*. However, Bohm's theory is not enough in and by itself to deny that the indeterministic conception of the physical world is a *consistent* stance. That is, Bohrians are not plainly *wrong* when they argue that SQM *can* be taken as describing an indeterministic world—Bohm himself acknowledged this point in one of the passages quoted above—they are wrong in that this is the *only* possible position.

Thus, since the description of the world that Bohrians extract from the theory is essentially indeterministic, why should they expect a causal and ontologically determinate explanation of interference phenomena? For a physicist that holds on to the view that quantum theory in Hilbert space describes an indeterministic world, the assumption that the paradigm of an explanation that provides understanding is a causal and deterministic one does not make any sense. That is, the argument is, in a way, circular. The ontologically determinate and causal explanations that Bohm's theory allows can be taken as offering a higher degree of understanding than SQM *provided that we think that the world is causal and amenable to a semi-classical description given in terms of particle trajectories*. The view that the world is not so is consistent with quantum theory in its Hilbert space formulation, so if this stance is adopted the need for a causal-ontological explanation simply evaporates. It is true that the accounts of the interference experiments that the Bohr-von Neumann interpretations of SQM provide are rather strange, but if we take the theory as describing an indeterministic world, it is rather natural that our (classical) conceptual machinery is not capable to grasp it in a complete way. Actually, as it can be noticed from the analysis in section 3.3.1, this view underlies Bohr's doctrine. If there is no place for causal explanations in the Bohrian interpretation, then it cannot be argued that the explanations of SQM are poor in providing understanding because they do not explain causally.

The second way in which the argument can be challenged is by simply noting that the Bohr-von Neumann interpretation is not the only possible one that can be assigned to SQM. Recall the basic idea of the modal approach. If the post-measurement state of a composite system of observed subsystem plus measuring device is expressed by $\sum_i c_i|\psi_i\rangle|\varphi_i\rangle$, we can take the reduced density matrix of each subsystem, namely, $\rho_I = \sum_i|c_i|^2|\psi_i\rangle\langle\psi_i|$ and $\rho_{II} = \sum_i|c_i|^2|\varphi_i\rangle\langle\varphi_i|$, respectively. The modal approach tells us that the observed system *I* and the measuring apparatus *II* have determinate values for the properties associated

[133] Vigier et al. adopt a similar stance. Commenting on the neutron interferometry experiments, they state that 'we are confronted by a stark alternative, either:

1. We renounce any possibility of describing what happens in the neutron interferometry experiments; there exists then no possibility of explaining [understanding] quantum phenomena, not even in terms of a wave/particle duality which only leads to ambiguity; individual quantum phenomena are in principle and irreducibly indeterminist in character and there can be no form of physical determinism appropriate in the quantum domain; or
2. We adopt the quantum potential approach as the only known consistent manner in which the quantum world can be conceived and explained in terms of a physically determinist reality; then, even if the quantum potential approach is not taken as the finally satisfactory description of quantum mechanical reality, it at least shows in a clear way the features that such a description must entail' (Vigier et al. 1987, 190).

to the observables defined by the projectors $|\psi_i\rangle\langle\psi_i|$ and $|\varphi_i\rangle\langle\varphi_i|$, respectively. That is, the density operators ρ_I and ρ_{II} correspond to the *dynamical* states of the subsystems, which determine the range of *possible* values that the system can take with respect to the corresponding observables, and which also tell us that the subsystems do have a definite value for the mentioned observables. What is the *specific value state* of the system, the dynamical state does not say. As Laura Ruetsche puts it, 'on the Modal Interpretation, a system's quantum state does not fix its value state. Rather, it fixes a set of possible value states and offers a probability distribution over those possibilities' (2003, 26).

When compared to the Bohr-von Neumann approach to SQM, it is very clear that BQT offers a much more definite and determinate account of what happens at the quantum level in the absence of measurements. As we saw above, Bohm's theory gives us a causal and precise account of what happens with an electron travelling from the emission source to the detection screen in the double-slit experiment and of what happens with a neutron inside an interferometer. SQM, on the other hand, tells us, in the Bohr-von Neumann approach, that the process cannot be literally described by the wave function and cannot be understood through classical concepts, or that the kinematics of the electron or the neutron are not well defined during the process.

However, from the modal approach point of view things are, in principle, different. This interpretation tells us that quantum systems have determinate values for (some) properties at all times, and that SQM is a theory that provides us with a complete account of their dynamical states. Therefore, regardless of whether measurements are considered or not, in the modal interpretation there is always a determinate story to be told about what happens in quantum processes—by 'determinate story to be told' I mean that the wave function literally (not symbolically) describes physical states and that these states have (some) definite properties at all times. The theory does not give us a specific pictorial representation of this story, for the value state is not fixed by the dynamical state. But this does not mean that SQM, in the modal interpretation, does not describe a determined quantum world[134]. Therefore, the comparative advantage of BQT with respect to SQM that Cushing argues for is not so clear when we consider the modal approach.

Anyhow, two remarks are appropriate in order to attain a balanced evaluation of the modal interpretation as providing a determinate account of quantum processes. First, the elaborations and amendments that have been introduced in the modal approach result in some awkward features. We saw in 3.3.5 that the Kochen-Specker theorem requires that the value state of a system is perspectival, that is, that the definite properties that a system possesses are relative to what reference system we are considering. When special relativity is considered, it turns out that perspectivalism holds also for reference frames, for the state value of a system is now relative to the specific simultaneity hyper-plane the system is considered to lie on (see Dieks 2005, section 5). Therefore, unlike BQT, the determinate account that the modal approach provides of quantum processes is not absolute, but perspectival and relative to a reference system and to a reference frame.

Second, the fact that the dynamical state does not fix the value state of a system suggests that the account that the modal approach offers of quantum processes is not as complete as it could, or should, be. We may ask how does the value state of a system at a certain time depend on its value state at earlier times. This question naturally comes up if we consider that one of the basic features of the modal approach is that the properties for which a quantum system has a definite value may change over time. That is, we may demand for a *value state dynamics* in order to have a *complete* modal interpretation. Unless such a completion is given we could still say that BQT beats SQM with respect to the degree of understanding that their explanations provide. Actually, Bacciagaluppi explains the demand for a value state dynamics in the modal interpretation in comparison with the dynamics in Bohm's theory:

[134] For a detailed explanation and assessment of the idea that the modal approach interprets SQM as a theory that describes a quantum world that is objectively determined regardless of whether measurements are considered or not, see (Dieks 1989; 2005).

> The modal interpretation is not just a reformulation of standard quantum theory in terms of different concepts. If, indeed, it delivers the goods of definite pointer readings, it does so because it introduces new structure into the theory, namely the *actually possessed properties* of a system. The *complete state* of a system is no longer its quantum state ρ, which, indeed, can be exhaustively described in terms of a set of definite properties and a probability distribution. The complete state is a *pair* (ρ, P_i), consisting of a quantum state ρ and one of the eigenprojectors P_i of ρ, representing the property of the system that is actually possessed. And the evolution of the complete state has to be given in terms of the evolution of ρ *and* of the evolution of the possessed property P_i. Thus [...], unless the modal interpretation is further supplemented by a dynamics for the *possessed* properties, it is *not a complete theory*.
>
> This point can be made more compelling by an analogy with the Bohm theory. The Bohm theory postulates that at every instant, positions of all (say N) particles are distributed according to $|\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)|^2$. If one assumes that all outcomes of measurements are ultimately recorded in positions of particles (in a pointer, in the ink of a piece of paper, etc.), this is enough to ensure that all measurements have outcomes that are distributed according to the quantum mechanical statistics. However, the Bohm theory does not merely *state* that the particles are distributed according to $|\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)|^2$. It also provides a *dynamical* mechanism that explains *how* the distribution of the particles is always maintained equal to $|\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)|^2$, as the latter varies according to the Schrödinger equation. (Bacciagaluppi 1998, 178-9)

That a value state dynamics is required need not be taken as a criticism or argument against the modal approach. Actually, some attempts fulfill such a requirement have been proposed[135]. However, the inclusion of a value state dynamics somewhat departs from the original spirit of the interpretation. The modal approach intended to solve the measurement problem in such a way that no extra ontological or conceptual baggage (with respect to the five postulates of SQM) is introduced. The reason for this seems to be to try to hold on to a theoretical structure as economical as possible[136]. Besides, and more importantly in our context, since the introduction of a value state dynamics implies the addition of theoretical structure that is not considered in the standard formalism of the theory, a complete modal approach may no longer be an interpretation of SQM, but turn into a rival (EE) theory — at least if we hold on to the criterion that establishes BQT as a rival theory and not as yet another interpretation of SQM[137].

---

[135] (Bacciagaluppi 1998) contains a review of the main proposals and the challenges they must face. I will only mention two general theoretical: the transition probabilities—the probability that a system has the property defined by P_i at t_i, given that at t_1 it had property P_1, at t_2 it had the property P_2, etc.—must be Markovian and consistent with the Born rule. However, the observance of these two requirements is not enough to single out a value state dynamics—infinitely many inequivalent dynamical proposals satisfy them, and there is no obvious criterion to select one specific proposal. Pragmatic considerations regarding the specific processes that are to be explained may help to single out an appropriate dynamics. But it is not assured that pragmatic considerations will be available and clear in every instance, let alone that a pragmatically grounded dynamics is also physically motivated (see Ruetsche 2003, section 3).

[136] 'Because Bohm's theory preserves the structural features of quantum mechanics but adds new elements to it, while not adding to the empirical content, there are in this particular case strong *methodological arguments* that speak in favor of conventional quantum mechanics. There is no *empirical support* for the introduction of additional structure. But this of course does not amount to a proof that conventional quantum mechanics has a greater probability *of being true* than Bohm's theory' (Dieks 1989, 1417, footnote 3); 'An objection to the Bohm theory is that it achieves its aims by introducing additional theoretical structure (the preferred observable) that does not figure in the standard formalism' (Dieks 2005, 413).

[137] Ruetsche provides the following definition of an interpretation of quantum theory: 'Questions classically transparent—which value-attributing propositions have truth values?, what are those truth values?—are quantum mechanically obscure. I suggest that to give an interpretation of QM, one must answer these questions. One must say, given a system's quantum state, what its value state is or might be. That is, an interpretation requires a semantics for quantum theory, a procedure for, given an element of the theory's state space, specifying the value-attributing propositions with determinate truth values for that state' (2003, 27). Ruetsche explicitly states that according to this definition, Bohm's theory is not a rival theory to SQM, but yet another interpretation—and the same holds for a complete modal interpretation. She also acknowledges that she is being more liberal in the conception of 'interpretation' than the common view in the philosophy of physics. She defends such liberalism in the following way: 'I urge that we adopt a principle of leeway according to which the interpretation of QM needn't be a purely semantic project. This principle frees interpretations from the obligation to adjust their semantics to the state space of QM, innocently construed; they may fiddle with that state space, or unitary dynamics, or both [...]. Why be more liberal? The best, but not the only, reason is that the best, or maybe the only, way to characterize a world of which the quantum statistical algorithm is so splendidly empirically adequate, is as a world of which an adulteration of QM innocently construed is true. Corruptions of QM innocently construed deserve to be called interpretations

Summarizing, we have that Bohm's theory provides us with an ontologically determinate, continuous and causal account of quantum processes, an account which is not offered by SQM in the Bohr-von Neumann interpretation. This difference may be taken as a comparative virtue possessed by the Bohmian approach, and, along this line, Cushing argues that Bohm's theory is clearly better than SQM in terms of the understanding of quantum phenomena it provides. Thus, this difference between the theories may count as a non-empirical criterion to prefer Bohm's proposal over SQM. However, I have shown that the argument is not appealing for the supporter of Bohr's interpretation, for in this case SQM is taken as a theory that describes an intrinsically indeterministic quantum world in which causal and ontologically determinate descriptions are just not possible. Besides, if we adopt the modal interpretation we have that SQM, in principle, is capable to include an ontologically determinate description of quantum processes. However, when a full modal description is achieved, the resulting theory may count as yet another EE rival to SQM.

3.6.2 Particles are always distinguishable

The fact that in Bohm's theory particles have well-defined trajectories at all times implies another feature that may be considered a comparative virtue with respect to SQM. Let us recall that in the latter theory, if two particles have all their non-relational properties in common and their wave-functions expressed in position basis overlap, then there is nothing that can count as a differentiation criterion—the particles are indistinguishable. This feature is grasped by the conditions of symmetrization for bosons (in the case of two indistinguishable particles a symmetrized state is expressed by $\Psi = \frac{1}{\sqrt{2}}(|\alpha_1\rangle|\beta_2\rangle + |\alpha_2\rangle|\beta_1\rangle)$) and the condition of antisymmetrization for fermions ($\Psi = \frac{1}{\sqrt{2}}(|\alpha_1\rangle|\beta_2\rangle - |\alpha_2\rangle|\beta_1\rangle)$), conditions that correspond to the Bose-Einstein and the Fermi-Dirac statistics, respectively. As we saw above, this indistinguishability involves a philosophical difficulty: if the particles are indistinguishable, then it is not clear in what sense they can be called individual objects.

Let us revise what are the conditions for identical particles, regarding their intrinsic non-relational properties, in Bohm's theory. In a system of n particles, the symmetrization and antisymmetrization conditions for arbitrary particle permutations obtain provided that $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n) = \pm\Psi(\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n)$, where $i, j = 1, \dots, n$. From what has been said in the previous sections we already know that all the particles in the system have well-defined trajectories determined by the quantum potential of the total system and their initial specific positions. That is, in spite of their identity with respect to intrinsic non-relational properties and the symmetrization and antisymmetrization conditions, in Bohm's theory we can still invoke the spatio-temporal histories of the particles in order to differentiate them from one another.

The situation is thus similar to the case of identical but distinguishable particles in classical mechanics, with the important difference of the non-local correlation effects produced by the quantum potential of the total system. Consider an (anti)symmetrized total system of two particles, each associated with a partial wave, described by $\Psi_\pm = C[\psi(\mathbf{x}_1, \mathbf{x}_2) \pm \psi(\mathbf{x}_2, \mathbf{x}_1)] = C[\psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2) \pm \psi_A(\mathbf{x}_2)\psi_B(\mathbf{x}_1)]$, where C is a normalization constant. If the partial waves overlap we have that $|\Psi_\pm|^2 = C^2[|\psi(\boldsymbol{x}_1, \boldsymbol{x}_2)|^2 + |\psi(\mathbf{x}_2, \mathbf{x}_1)|^2 \pm \psi^*(\mathbf{x}_1, \mathbf{x}_2)\psi(\mathbf{x}_2, \mathbf{x}_1) \pm \psi^*(\mathbf{x}_2, \mathbf{x}_1)\psi(\mathbf{x}_1, \mathbf{x}_2)]$. That is, for an (anti)symmetrized state the quantum potential determined by the amplitude function R will include contributions from the interference terms. Now, depending on whether the total state is antisymmetrized or symmetrized, we have that the contribution of the interference terms is different, so that the non-local effects exhibited by the corresponding quantum potentials are also different, and these different non-local effects, in turn, result in

of the theory because they may turn out to be the best way of making sense of the theory's capacity to save the phenomena' (ibid., 28). The interesting point implied in Ruetsche's liberal view is that it shows that the criterion delimiting what constitutes an interpretation and what a rival theory is permeated by *pragmatic* considerations.

different statistical behavior—FD for antisymmetric states and BE for symmetric ones, of course[138]. That is, the departure from Maxwell-Boltzmann statistical classical behavior that ensembles of bosons and fermions exhibit is the result of the particular non-local effects that the respective quantum potentials determine. When a two particle system is factorizable, that is $\Psi = \psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2)$, we have that the interference terms vanish in $|\Psi|^2$, so that no non-local effects are included in the quantum potential and the particles behave independently of each other—as we saw above, effective factorization obtains when the partial waves do not overlap or when interaction with the environment wash out the interference terms. More generally, if a system of n-particles is (effectively) factorizable, it exhibits a MB statistical behavior.

Hans Reichenbach (1956) clearly explains that the assumption of distinguishable identical particles in the context of BE statistics necessarily implies non-local correlations. Consider the case of two coins 1 and 2 that we toss. If MB statistics obtained, the four equiprobable configurations that can be observationally distinguished are $H_1H_2, H_1T_2, T_1H_2, T_1T_2$ (we denote 'heads' and tails by H and T, respectively)—notice that these configurations exhaust all the possible combinations. But let us assume that BE statistical behavior is observed, that is, that the three equiprobable and observationally distinguishable configurations are HH, TT, HT—we omit the subindexes to show that permutation in the 'different state' case does not make a difference. To make sense of this statistical behavior under the supposition of particle distinguishability in terms of spatio-temporal histories we must assume that the *possible* combinations are not equiprobable. That is, whereas in the case of MB statistics we have that the probability of 'only heads', of 'only tails', of 'heads and tails', and of 'tails and head' are all $\frac{1}{4}$; in the BE case we have that—if we assume spatio-temporal distinguishability—the probability of 'only heads' and 'only tails' are both $\frac{1}{3}$ and the probability of 'heads and tails' and 'tails and heads' are both $\frac{1}{6}$. This means that there must be some correlations in the behavior of the coins, for if one of the coins shows 'heads' there is a tendency in the other one to show 'heads' as well—the conditional probabilities of H_2 and of T_2 given H_1 are $\frac{2}{3}$ and $\frac{1}{3}$, respectively. Reichenbach concludes that

> we arrive at the following result. Assume that we could assign material genidentity [spatio-temporal distinguishability] to each particle of a Bose ensemble; then we would find that the particles are mutually dependent in their motions. If one particle is in a certain state, then there exists a tendency for the others to go into the same state [...]. These causal relationships would represent action at a distance [non-locality], since the particles can be far apart; that is, the dependence relations would constitute causal anomalies. In other words: any assignment of physical identity to Bose particles leads to causal anomalies. (Reichenbach 1956, 70-1)

We can clearly see that this is precisely the way in which BE statistics arise in BQT. In a symmetrized state in which the partial waves overlap, the quantum potential determines non-local effects that correlate the behavior of the particles, resulting in BE statistics. Reichenbach does not make any reference to Bohm's theory, so his conclusion that the non-local correlations are causal *anomalies* may be somewhat exaggerated. If we accept Bohm's theory as the correct quantum theory, we should learn to live with non-local causality as a fundamental feature of the physical world.

---

[138] Antisymmetrization implies that two fermions cannot be in the same quantum state at the same time. Pauli's exclusion principle is a basic postulate in SQM. In BQT there is an explanation for it. If two fermions are in the same quantum state, we have that $\Psi = 0$. In the context of Bohm's theory this means that there is no possible trajectory defined for such a case, for trajectories cannot pass through *nodes* (regions in which the wave function is zero or undefined). It is the 'quantum forces' in the quantum potential corresponding to an antisymmetrized state what forbids a violation of Pauli's principle: 'for fermions, the antisymmetry of $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_2)$ implies that $\psi = 0$ if any two or more sets of coordinates are equal. Since the configuration space path cannot pass through nodes, it follows that two or more fermions cannot occupy the same point in 3-space [configuration space?] at the same time (Pauli's exclusion principle). The total potential will always act in accordance with this requirement' (Holland 1993, 284). For an analysis of concrete examples of MB, BE and FD statistics in BQT, including nice graphic representations of the corresponding allowed trajectories, see (Vigier et al., 1987, section 4) and (Holland 1993, section 7.4).

Peter Holland underscores an interesting point regarding identical particles in Bohm's theory. Assume that a system of two identical particles $\Psi = \psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2)$ is strictly factorizable, and that after some time the partial waves overlap. Since the Schrödinger evolution of such a state cannot turn into an (anti)symmetrized state $\Psi_\pm = C[\psi(\mathbf{x}_1, \mathbf{x}_2) \pm \psi(\mathbf{x}_2, \mathbf{x}_1)] = C[\psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2) \pm \psi_A(\mathbf{x}_2)\psi_B(\mathbf{x}_1)]$, then the assumption of strict factorizability cannot be correct. That is, we cannot really conceive that any two different identical particles are effectively independent from one another:

> From a physical point of view, where does the second term in $[\Psi_\pm = C[\psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2) \pm \psi_A(\mathbf{x}_2)\psi_B(\mathbf{x}_1)]]$ come from? One cannot pass from $[\Psi = \psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2)]$ to $[\Psi_\pm = C[\psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2) \pm \psi_A(\mathbf{x}_2)\psi_B(\mathbf{x}_1)]]$ as part of the evolution of the overlapping packets described by the Schrödinger equation. We conclude that the assumption that the wavefunction of two distantly separated identical particles is factorizable is, in fact, incorrect. The actual state is really always $[\Psi_\pm = C[\psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2) \pm \psi_A(\mathbf{x}_2)\psi_B(\mathbf{x}_1)]]$ which, when the wavefunctions ψ_A and ψ_B do not overlap, is effectively factorizable and so physically equivalent to the product state. Does this mean that we can never isolate a particle from all the other particles in the universe of the same species and ascribe to it its own wavefunction? Strictly speaking, no, if the symmetrization postulate is exact. This then implies a very strong form of state-dependence. However, in many situations in which particles are distantly separated, classical interactions may be neglected and the component parts of a symmetrized wavefunction will not overlap; ψ will then be effectively factorizable. It is legitimate in that case to treat the many-body system as a collection of independent systems obeying MB statistics. (Holland 1993, 292-3)[139]

We may now consider to what extent the fact that particles are always distinguishable in BQT can count as a virtue to determine the acceptance of the latter theory over SQM. It is true that this feature prevents us from even having to deal with the problems with the notion of objective identity that arises in the standard formulation of the theory. However, we saw above that there are some possible standpoints that, in principle, allow the supporter of SQM (in any of its interpretations) to cope with the problem. Recall that object identity can be saved by means of *haeccity* arguments, weak discernibility arguments, or by restricting the notion of objective identity to quantum fields and to consider the tag 'particle' as adequate for entities that emerge from quantum fields in the classical limit. Bohm's theory certainly provides a clear and (semi)classical account of the objective identity of quantum particles, but this feature may count as a reason to prefer it over SQM granted that we believe that the quantum world is amenable to a (semi)classical scientific representation, or granted that we have and accept arguments that the scientific enterprise *must* be such that its theoretical achievements are to be provided within a (semi)classical and ontologically determinate scaffolding. If we have no problems with an indeterministic world that is not suitable to be described in (semi)classical terms, for example, to renounce to the familiar and intuitive account of objective identity that classical physics offers should not count as a big sacrifice to make.

On the other hand, the distinguishability of fermions and bosons clearly shows that non-local causality is a feature that we must necessarily introduce in the representation of the physical world that BQT depicts. Even though I do not think that non-locality cannot be included in a rational and coherent description of the physical world, it is clear that it involves a renunciation of an important part of the intuitive and classical framework of physical science. That is, there are epistemological reasons that may advise dodging non-local causality. For example, Reichenbach explicitly asserts that the conflict with (local) causality implied in the assumption of distinguishable bosons is reason enough in order to simply assume the indistinguishability of quantum particles:

> We see now how the thesis concerning indistinguishable particles is to be qualified. In precise language we cannot simply say: the particles are indistinguishable. We must say: either the particles are indistinguishable, or their behavior displays causal anomalies. We are left the choice of selecting one or the other interpretation. Neither interpretation is 'more true' than the other; two are equivalent descriptions.

[139] The same holds, mutatis mutandis, in the case of SQM. The states of identical particles should always be (anti)symmetrical.

> However, only one of the two descriptions supplies a normal system, that is, a system free from causal anomalies; this is the description according to which the particles are indistinguishable. When we follow the usual rule of employing a normal system whenever it is possible, we may therefore say, without hesitation, that the particles are indistinguishable. (Reichenbach 1956, 71).

From the discussion of indistinguishable particles above we know that the alternative that Reichenbach presents us is a false dilemma—there are other possible standpoints on the subject. However, the quoted passage illustrates that if we assume an epistemological background in which non-locality is to be rejected from the outset, then the fact that BQT is a theory in which particles are always distinguishable will not count as significant feature to recommend its acceptance. That is, particle distinguishablity may count as a virtue that endorses a decision favoring BQT over SQM granted that we are willing to assume non-locality as a fundamental feature of the world. There are several cases in the history of physics in which cherished epistemological and ontological principles have been abandoned in the name of empirical adequacy—the best available theory may contradict those principles. However, in this case, the EE between SQM and BQT implies that we cannot have recourse to the empirical evidence that supports these theories in order to decide whether non-locality is something that we may or may not include in our conception of the physical world (furthermore, some form of non-locality seems to be postulated also by SQM, I will deal with an evaluation of non-locality in quantum theory below). More generally, considering the epistemological trade-off that *both* SQM and BQT compel us to do with respect to the classical framework of physics (rejection of definite trajectories, of local causality, etc.), there is no way to establish, in terms of empirical evidence, what is the best deal.

3.6.3 The classical limit

The next possible comparative virtue of BQT I will consider is given by the way in which in this theory the classical description emerges from the quantum level. If we assume a comparative standpoint with respect to the Copenhagen interpretation, we find that the emergence of the classical level seems to be a rather problematic issue in SQM. We may expect that a natural and appropriate account of such an emergence must be similar to the way in which Newtonian mechanics follows from relativity theory as a special case ($v \ll c$). However, given the epistemological priority that Bohr assigns to the language of classical physics, it is difficult to see how an explanation analogous to the Newtonian mechanics-relativity link can be obtained, for it would be required that the classical vocabulary were physically and mathematically grounded on the formalism of the quantum theory—the language of classical physics should arise as a special case of the quantum description. Peter Holland states the problem in the following way:

> In Bohr's interpretation the validity of classical concepts is already presupposed since, it is suggested, it is only in terms of these that one can unambiguously communicate the results of experiments in the quantum domain. Thus, classical physics must be considered as prior to quantum mechanics and the latter is a generalization of the former in that it provides a new set of laws governing the application of classical concepts. According to this view, any procedure by which classical mechanics is recovered from quantum mechanics as a mathematical limit can only be a demonstration of consistency with the already postulated epistemological relation between the two theories and not as a 'derivation' of classical mechanics from quantum mechanics [...]. Yet in spite of this admonition the discussion of the problem is usually carried on as if quantum mechanics is the more fundamental theory from which classical mechanics emerges when certain parameters naturally occurring in the theory are varied (in much the same way as classical physics is supposed to be a special case of relativity when the velocity is small compared to the speed of light). As a result a great deal of confusion still surrounds this question and there is not even agreement on what constitutes a universal mathematical criterions to characterize the classical limit. (Holland 1993, 219)

Since in BQT there is no assumption of classical language priority, this conceptual difficulty for an account of the emergence of the classical world does not arise. In principle, at least, the classical limit

seems to be a rather simple, technical issue. Let us recall the equation $\frac{\partial S}{\partial t}+\frac{(\nabla S)^2}{2m}+V-\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}=0$. We might say that a derivation of classical mechanics is obtained from the quantum theory if we are capable to deduce from this formula the Hamilton-Jacobi equation $\frac{\partial S}{\partial t}+\frac{(\nabla S)^2}{2m}+V=0$. The easy way to achieve this looks to be to take the limit $\hbar \to 0$, so that the quantum potential $Q=-\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}$ vanishes. But this strategy does not work. The quantum potential Q does not depend only on $\hbar$, but also on S and R, which are also $\hbar$-dependent. That is, the way in which Q is determined by Planck's constant is subtle, so that even when $\hbar \to 0$ its dependence on S and R is such that $Q \nrightarrow 0$. Even more obviously, if we simply plug $\hbar = 0$ in the Schrodinger equation $i\hbar\frac{\partial \Psi}{\partial t}=-\frac{\hbar^2}{2m}\nabla^2\Psi+V\Psi$ we obtain a nonsensical result.

The appropriate conditions under which the classical limit arises can be formulated by inspecting the following two equations that hold in BQT

$$\frac{d\mathbf{p}}{dt}=-\nabla(V+Q) \qquad\qquad \frac{dE}{dt}=\frac{\partial}{\partial t}(V+Q),$$

and comparing them with Newton's second law and the law of energy change as derived from the Hamilton-Jacobi equation, namely, $\frac{d\mathbf{p}}{dt}=-\nabla V$ and $\frac{dE}{dt}=\frac{\partial V}{\partial t}$ [140], respectively. For the equations in BQT just mentioned to reduce to their classical counterparts it is required that the quantum potential is constant over time and the involved trajectories. That is, $(-\nabla Q)$ and $(\partial Q/\partial t)$ must be negligible[141]. Besides, we must consider that in classical mechanics the zero energy level is given by $V = 0$. Therefore, we also require that in regions where $V \to 0$ it also holds that $Q \to 0$, otherwise in a case in which V vanishes but Q does not, the quantum potential would contribute with an amount of energy resulting in quantum behavior. Thus, $Q \to 0$ stands as the necessary and sufficient condition for the classical limit to arise from BQT.

The resulting general explanation that BQT gives us for the emergence of classical behavior is simply that under certain physical conditions the participation of the quantum potential in the behavior of physical systems becomes idle[142]. Notice that this natural explanation does not require any special reference to measurements or to observers, and consequently it does not rely on the microscopic-macroscopic distinction—no arbitrary 'cut' between observed system and measuring apparatus is needed, 'emergence' of classicality is just an objective physical process that the theory, at least in principle, can naturally describe. Actually, the theory tells us that in some macroscopic cases quantum behavior obtains anyway,

---

[140] These equations are obtained, respectively, by applying the gradient operator ∇ and the operator $\frac{\partial}{\partial t}$ to the Hamilton-Jacobi equation, and identifying $-\frac{\partial S}{\partial t}=E$. The analogous equation of Bohm's theory follow applying the same procedure to the equation $\frac{\partial S}{\partial t}+\frac{(\nabla S)^2}{2m}+V+Q=0$ (see Holland 1993, sections 3.9.3 and 6.2).

[141] As Holland explains, 'an important point is that it is not sufficient in obtaining the classical limit that the absolute value of the quantum potential, $|Q|$, be small in comparison with the other energies appearing in the Hamilton-Jacobi equation. [...] $|Q|$ is practically negligible compared with the kinetic energy of an electron in the two-slit experiment, and yet of course the channeling of the trajectories which yields the characteristic interference pattern is entirely due to the action of Q. The particle responds to the spatio-temporal *variation* of Q (i.e. force) rather than its numerical value' (1993, 226).

[142] This is just an exposition of the conceptual framework that clarifies what are the general conditions for the classical limit in BQT. How the classical description is achieved in specific contexts is a more complicated matter. For example, we already know that since the position of a quantum particle depends only on its previous position and on the quantum potential then trajectories cannot cross in configuration space (or in 3-space in the case of a two-particle system), whereas in classical mechanics the trajectories of particles can cross—actually, this is so in the case of particles that exhibit MB statistical behavior, see (Vigier et al., 1987, section 4; Holland 1993, section 7.4). A solution for this difficulty can, in principle, be given by the decoherence effect, for 'the non-interfering components produced by decoherence can indeed cross, and so will the trajectories of particles trapped inside them' (Bacciagaluppi 2012, 23). For an overview of how classical behavior arises from SQM in specific cases see (Bohm & Hiley 1993, chapter 8; Bacciagaluppi 2012, section 3.2 and the references therein).

for the physical conditions are such that the action of the quantum potential is not washed out. An example is given by a thought experiment due to Einstein (1953). Consider a box with impenetrable and perfectly reflecting walls separated by a length L and a particle freely moving in one dimension within the box. The wavefunction of the particle is given by $\psi = \frac{1}{\sqrt{2L}} e^{-iE_n t/\hbar} \sin\left(\frac{n\pi x}{L}\right)$, where n is an integer and $E_n = \frac{(n\pi\hbar/L)^2}{2m}$. Since the quantum wave is a standing wave such that $\nabla S = 0$, Bohm's theory tells us that the velocity of the particle is $v = \nabla S/m = 0$. That is, the particle is at rest within the box and its probability to be in a certain position $\mathbf{x}$ between the walls is given by $|\psi|^2$, and, although the kinetic energy is zero, the total energy is fully given by the quantum potential[143].

Einstein stated that the description of the system would be given by ψ even in the macroscopical level— it would still hold for a particle with diameter 1 mm in a box whose walls are 1 m apart[144]. That is, even at the macroscopic level this example displays an obvious quantum behavior. Einstein took this feature of Bohm's theory as a flaw. The reason he invoked was the violation of a requirement he considered as a principle that gets violated in Bohm's account of the particle in the box. In his reply, Bohm paraphrased Einstein's principle in the following way: 'all microscopic theories must always become identical with previously accepted macroscopic theories, when one considers sufficiently large dimensions' (1953, 16). That is, Einstein demanded a correspondence between the macroscopical level and classical behavior.

Bohm replied in a twofold way. First, he stated that the predictions of his theory neatly correspond to the predictions of the standard formalism. In SQM the location of the undisturbed particle within the box is not determinate, but a momentum measurement would yield the equiprobable results $\pm\hbar n\pi/L$—the particle would be found moving back or forth with equal probability. This prediction also obtains in BQT, for a momentum measurement would imply a disturbance of the quantum potential (by removing one of the walls, for example) such that the particle would be set in motion with a momentum $\pm\hbar n\pi/L$— with the direction depending on the specific initial position of the particle, that we do not know. Second, Bohm compellingly pointed out that, although Einstein's principle may be useful as a heuristic principle in the quest for adequate theories—actually, the 'correspondence principle' seems to be a version of the principle at issue—there are no good reasons to hold it as an *a priori* criterion to *reject* theories. The absolute demand for correspondence between macroscopy and classical behavior is nothing but a metaphysical commitment that may not be sanctioned by nature:

> In conclusion, the author would like to state that he would admit only two valid reasons for discarding a theory that explains a wide range of phenomena. One is that the theory is not internally consistent, and the second is that it disagrees with experiment. Principles and requirements such as those suggested by Einstein frequently serve as valuable heuristic aids, when they are used in the search of a *positive* theory. But when they are used to *discard* a logically consistent theory that agrees with all known experimental facts they may lead to another form of precisely that subjectivism which Einstein finds unsatisfactory in Born's interpretation of the quantum theory. As Einstein himself often emphasized, one must be very careful in judging the admissibility of a theory on the basis of general arguments. (Bohm 1953, 18-9).

---

[143] A similar description accounts for stationary atomic states. Recall that de Broglie's the condition for a stationary orbit is that the wave function of the orbiting electron around the nucleus is given by a standing wave. In BQT this means that the phase function of the electron is given by $\nabla S = 0$, which means that the electron is at rest in a fixed position with respect to the nucleus, it does not really orbit! This naturally accounts for the fact that no energy radiation occurs: the electron is not accelerating so it does not collapse into the nucleus. That the electron possesses an 'angular momentum' in spite of being at rest is accounted by the contribution of the quantum potential, not by the electron's motion—in stationary states, the quantum force and the classical force exactly balance each other according to $V + Q = E_n$ (see Holland 1993, section 4.5). Moreover, in SQM there are no quantum jumps. Excited electrons do go through a continuous path in regions where Q is not 'stationary'. A nice illustration of the continuity of transition between stationary states is given by Bohm's account of the Franck-Hertz experiment (Bohm 1952, 175-7).

[144] Einstein's conditions for macroscopicality are $n \gg 2$ and $p_n L \gg h$. This example shows that taking the limit $h \to 0$ and/or $n \to \infty$ does not work as a sufficient condition to obtain the classical limit from Schrödinger's equation.

From what has been said one may get the impression that the goal of deriving classical physics from BQT in a way analogous to the derivation of Newtonian mechanics from relativity theory can be easily fulfilled. Holland stresses out one loose end in this issue that impedes that we jump to this conclusion. Just as there are some quantum states for which the limit $Q \to 0$ never obtains—quantum systems with no classical analogue like the case of the particle in the box—the possibility cannot be excluded *a priori* that there are solutions to the classical equations of motion that do not correspond to the limit of corresponding quantum solutions. That is, classical systems with no quantum analogue may exist. Although in most of the physically interesting cases, the values of the classical potentials are consistent with classical limit solutions of the quantum equations, there is no general proof that assure that *all* classical solutions can be obtained in this way. From this situation Holland concludes that

> it appears reasonable to conceive of classical mechanics as a special case of quantum mechanics in the sense that the latter exhibits new elements ($\hbar$ and Q) not anticipated in the former. But the possibility that the classical theory admits more general types of ensembles than can ever be reached from the limit of quantum ensembles [...] suggests that the two statistical theories are disparate while having a common domain of application. The intersection is characterized by $Q \to 0$ in the quantum theory. There is a well-defined conceptual and formal connection between the classical and quantum domains but as regards ensemble theory they merely intersect rather than one being contained in the other. (Holland 1993, 228-9)

Anyhow, when compared to Bohr's interpretation of SQM, it is clear that BQT offers a much better approach to the question of the classical limit. As mentioned above, the main problem is connected to the epistemological primacy that Bohr attributes to the language of classical physics. But we know that this view is not the only possible interpretation of SQM, and it may be the case that the question of the classical limit is not as problematic from the perspective of the alternative interpretative approaches. Moreover, we saw above that environmental decoherence—especially in the context of the consistent histories interpretation—may in principle provide an adequate account of how the classical description of the physical world emerges from the standard formalism. Even though there still much to be clarified, the hope that this goal will be achieved is not unfounded. Thus, one may conclude that if Bohr's argument concerning the priority of classical language is rejected and decoherence is taken on board, the question of the classical limit in SQM is not that stringent anymore.

But this counterargument can be challenged. If we consider the meaning that the wavefunction is assigned in the different interpretations of SQM we find that the conceptual link between the classical and quantum description is rather dubious. In Bohr's theory the wavefunction does not really have a descriptive meaning in the absence of measurement interactions, but a symbolic one. In a more literal interpretation *a là* von Neumann, the wavefunction describes a physical system for which properties like position are not well-defined. Neither the consistent histories interpretation nor the Everettian ones really deal with the meaning of the wavefunction insofar as measurements are not performed. Actually, in the many-worlds interpretation the question of the classical limit is closely associated to the problem of the preferred basis. In the modal approaches the dynamical state refers to the possible values that physical systems can take for properties, and there is always an underlying value state, but the property for which the system has a determinate value may not be its position. Therefore, the concept 'trajectory' is not one that is (always) well-defined in either of the alternative interpretative approaches. This fact, in turn, may be taken as a sort of semantic incommensurability between quantum theory and classical physics that threatens the possibility of a satisfactory account of the question of the classical limit:

> The '$\mathbf{x}$' appearing in the argument of the Schrödinger equation is usually supposed to represent only the potential locations of a particle, one of which is realized with a probability $|\psi|^2$ if a measurement is performed to determine the position. It does not refer to a current material location. The classical analogue of this system is an ensemble of point particles of mass m each deterministically pursuing a trajectory $\mathbf{x}(t)$, with a certain probability of occupying a particular region of space at each instant. How can one pass from

> a theory in which ψ merely represents statistical knowledge of the state of a system to one in which matter has substance and form independently of our knowledge of it? That is, even if one contrives to obtain $[\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V = 0]$ in some limit, why are we justified in identifying this with the Hamilton-Jacobi equation describing the propagation of the S-function associated with an ensemble of precisely defined trajectories whose law of motion is given by $\dot{\mathbf{x}} = \nabla S/m$? Or in treating the probability of being as a limit of the probability of finding? The answer is that one is not justified in doing these things: one cannot logically deduce a model of substantial matter and its motion from an algorithm which has not theory of matter and motion in it at all (i.e., which makes no statements as to what matter *is*). In order to connect quantum and classical dynamics smoothly in the usual approach the law $\dot{\boldsymbol{x}} = \nabla S/m$ and the notion of a precise initial position $\mathbf{x}_0$ of a material object are slipped in as additional postulates. Indeed these postulates, which cannot be derived from the wavefunction, are made in a domain where *quantum mechanics is valid* (for $[\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V = 0]$ is here an instance of the Schrödinger equation). (Holland 1993, 221-2)

That is, one may not expect only a formal account of how the classical limit emerges from the standard formalism, but also an explanation of why the classical world can be described in terms of precisely localized objects if at the quantum fundamental level the notion of precisely localized physical systems does not hold. More generally, we have that BQT offers an account of the question of the classical limit that does not suffer of basic conceptual problems. The main challenge is to show in each case why and how the quantum potential vanishes—decoherence may play an important role in the clarification of this issues, see (Schloshhauer 2004, section F). On the other hand, SQM need not only to formally explain the emergence of classicality, but must also clarify how the conceptual description of the classical world in terms of classical concepts can be grounded on a theoretical framework in which the notion of definitely localized objects, for examples, is not defined. To be fair, though, this task is one of the goals that the work on decoherence tries to reach. Environmental decoherence, especially through scattering processes, may be the answer for how localization of systems described by spatially spread wavefunctions occurs (see Schlosshauer 2007, chapter 3), and successful results of research on this field would be available for the alternative interpretations of SQM in order to provide an account of the classical limit question.

3.6.4 No measurement problem

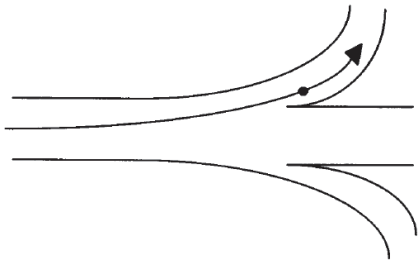
Maybe the most notorious feature of BQT when it comes to an evaluative comparison with respect to SQM is that its account of measurement interactions is completely transparent and such that no problems regarding definite outcomes arise, that is, in BQT there is no measurement problem whatsoever. To understand in what way measurements are described in Bohm's theory we need just to generalize the analysis offered in section 3.5.2 where we discussed contextuality. Let us consider a composite system of observed subsystem and measuring device at t_o given by $\psi_o(\mathbf{x})$ and $\phi_0(y)$, respectively (we assume the relevant coordinates of the measuring system to be defined in one dimension, like in a 'pointer position' observable). Before the measurement interaction the total wave function is factorizable and given by $\Psi_0 = \psi_o(\mathbf{x})\phi_0(y)$—the subsystems are independent.

We want to find out the value of subsystem ψ for a property corresponding to the operator A, so we make the subsystems ψ and ϕ to interact according to a suitable Hamiltonian H determined by a coupling constant g. At the time T, corresponding to the termination of the interaction, the total wavefunction is thus $\Psi(\mathbf{x}, y, T) = \sum_a c_a\psi_a(\mathbf{x})\phi(y - gaT)$—we have expanded Ψ on a basis given by the eigenvectors $\psi_a(\mathbf{x})$ of the spectral decomposition of operator A (we assume there is no degeneracy), where the a are the eigenvalues of A. Given the wave overlap due to the interaction, now the total wavefunction is not factorizable and the point $(\mathbf{x}(t), y(t))$ defining the trajectory of the full system in configuration space undergoes a complicated motion in configuration space due to the fluctuating values of Q.

Just as the measurement requires, we see that the y-coordinate gets correlated to the eigenvalues of A. The centers of the wave packets ϕ move along the 'channels' $y_{at} = gat$, with $0 < t < T$, so that when the

interaction ends the separation between the channels associated to different eigenvalues is given by $\delta y_a = gT\delta a$. With a suitable choice of g, at the time T the value of gT will be considerably larger than the width of the packets Δy, so that no significant overlap between the channels is present. After the end of the interaction at time T the system evolves according to the free Hamiltonian, so the packets ϕ acquire a classically describable separation—that is, the condition for observationally discernible measurement outcomes is $\Delta y \ll \delta y_a$, whose observance depends on the value of g.

Due to the measurement interaction as described, the total wave-function then splits up into a collection of non-overlapping configuration space partial wave-functions. The system enters in only one of these different channels, so the measurement outcome is given by the eigenvalue a associated to that channel. The non-occupied channels spread out, and since no overlap is present, they do not influence the behavior of the system—after T it becomes effectively factorizable—and since the channels are separated by regions where $\Psi = 0$ or where the amplitude is overwhelmingly small, the probability of the system to cross from one channel to another is zero or negligible—once the system enters one of the channels it remains there. Consequently, we can simply replace the total wave-function $\Psi(\mathbf{x}, y, T) = \sum_a c_a\psi_a(\mathbf{x})\phi(y - gaT)$ by one of its summands $c_a\psi_a(\mathbf{x})\phi(y - gaT)$—and since the quantum potential does not depend on the amplitude of the wave-function, we can renormalize the new wave-function in order to calculate consistent probabilities in later experiments. The following figure clearly illustrates the process:



It is rather clear that in this account of measurements no problems regarding definite outcomes arise. The specific summand that corresponds to the actual state of the system after the measurement interaction is determined by the (unknown) initial conditions $\mathbf{x}_0, y_0$ and by the form of ψ_0 and ϕ_o, and it can be found out simply by checking the y-coordinate in the measuring device—but notice that the process takes place regardless of whether we look at the apparatus or not. Thus, the replacement $\Psi \to \psi_a(\boldsymbol{x})\phi(y - gaT)$ neither yields nor requires any kind of wave-function collapse:

> The definite outcome is obtained from the usual Schrödinger evolution and does not require the sudden intervention of an unexplained 'collapse'. Our knowledge of the state of the system of interest changes because of the objective transformation [$\Psi \to \psi_a(\boldsymbol{x})\phi(y - gaT)$] of the physical wave. Wave packet collapse puts it the other way around: the change in the wave (collapse) occurs when our knowledge changes. In our approach no collapse, in the sense that the unrealized summands in [$\Psi(\mathbf{x}, y, T) = \sum_a c_a\psi_a(\mathbf{x})\phi(y - gaT)$] cease to 'exist', actually occurs. When a detector clicks the wavefunction does not 'collapse' from all over space to a point, it is simply that only part of it is now relevant [...]. And because by assumption the apparatus coordinates are always well defined there is no need to invoke further systems that can put the observer into a definite state. We are therefore able to avoid an infinite and inexplicable regress. (Holland 1993, 344)

As to the probabilities of the possible outcomes, we have that at $t > T$, when the wave-packets cease to overlap, the probability of the system to lie in a certain point $(\mathbf{x}, y)$ in configuration space is given by $P(\mathbf{x}, y) = |c_a|^2|\psi_a(\mathbf{x})|^2|\phi(y - gAt)|^2$. Therefore, in order to determine the probability of a measurement outcome a we need to integrate over the probability of all the points $(\mathbf{x}, y)$ contained in the ath channel

or wave- packet, that is, $P(a) = \int |c_a|^2 |\psi_a(\boldsymbol{x})|^2 |\phi(y - gAt)|^2 \mathrm{d}^3x\mathrm{d}y$. Since we assume that both ψ_a and ϕ are normalized, we have that $\int |\psi_a(\boldsymbol{x})|^2 |\phi(y - gAt)|^2 \mathrm{d}^3x\mathrm{d}y = 1$, so that $P(a) = |c_a|^2$. In other words, just as in SQM, in BQT the probability of obtaining an outcome a in an A-measurement is given by the Born rule.

We are entitled to perform the replacement $\Psi \to \psi_a(\boldsymbol{x})\phi(y - gaT)$ granted that the after measurement interaction wave-function is effectively factorizable. We may then comply that the 'empty waves' could somehow be brought together again, so that their overlap would forbid us from inferring the A value from y—and the theory of measurements in BQT would get problematic. Though this is theoretically possible, its practical possibility is negligible. The reason is that the y-coordinate is coupled to an enormous amount of degrees of freedom making up the measuring apparatus. To reestablish the overlap, all these degrees of freedom must also be coupled, an operation that is practically impossible. More precisely, we can introduce the mentioned degrees of freedom by the variables $z_1, \ldots, z_n$ with $n \sim 10^{23}$ and assign them a wave-function $\xi_0(z_1, \ldots, z_n)$. The total wave-function is then $\Psi_0 = \psi_0(\mathbf{x})\phi_0(y)\xi_0(z_1, \ldots, z_n)$. After the measurement interaction, the wave-function is given by $\Psi(\mathbf{x}, y, z_1, \ldots, z_n, T) = \sum_a c_a \psi_a(\mathbf{x})\phi(y - gaT)\xi_a(z_1, \ldots, z_n)$. For the mentioned problem to come up, it would be necessary that all the ξ_a are brought back together and overlap, but the probability of such an event is virtually null.

Another important feature in the account of measurements of BQT is that, in general, experiments do not *reveal* the property value the observed system at the instant previous to the measurement. Actually, the measurement 'produces' the obtained value. As we have seen, the measurement interaction *changes* the wave-function $\psi_0(\mathbf{x})$ into a different one that is finally correlated to a pointer position. The value a obtained in the measurement outcome corresponds to this transformed wave-function, though it is causally determined by $\mathbf{x}_0$ and ψ_0, of course. More precisely, we have that the evolution of the property A as 'possessed'[145] by system $\psi(\mathbf{x})$ evolves along a measurement interaction in the following way:

$$\langle A\rangle(\mathbf{x}_0, t_0) = \mathrm{Re}\psi_0^*(\boldsymbol{x}_0, t_0)A\psi_0(\boldsymbol{x}_0, t_0) \to \langle A\rangle(\mathbf{x}, y, t) = \mathrm{Re}\,\Psi^*(\mathbf{x}, y, t)A\Psi(\mathbf{x}, y, t) \to a = \mathrm{Re}\,\psi_a^*(\mathbf{x}, t')A\psi_a(\mathbf{x}, t')$$

where $t_o < t \leq T < t'$, and assuming that ψ_0, Ψ and ψ_a are normalized. In the context of Bohm's theory, the only case in which a measurement reveals the relevant property value possessed by the system prior to the interaction occurs when at t_0 the wave-function $\psi(\mathbf{x})$ is an eigenfunction of A. As Holland puts it,

> in order that we may extract from the apparatus variable unambiguous information on the system of interest, the wavefunction of the latter must, in general, undergo an irreducible and unpredictable transformation so that the values of the relevant physical properties of the particles that it determines are no longer those obtaining prior to the interaction. As a result, measurements do not and cannot, in general, reveal the current values of physical quantities. (1993, 337)

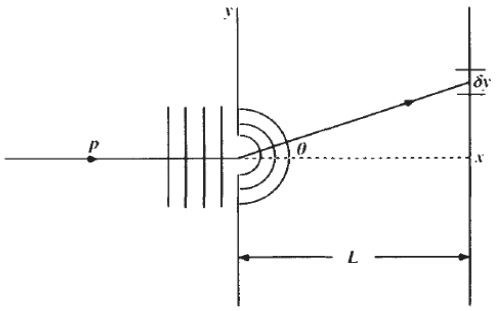
> Whatever the initial true value $[A_0(\mathbf{x}_0, t_0)]$ may have been, it has been deterministically and continuously transformed into an eigenvalue. Which result will be found is unpredictable and depends critically on the initial values $\mathbf{x}_0, y_0, \psi_0$ and ϕ_0. Only if $\psi_0 = \psi_a$ do we find $A_0 = A = a$ whatever the initial positions $\mathbf{x}_0, y_0$. It is then obvious that is we immediately repeat the same impulsive operation on a system that has been mapped from an arbitrary initial state ψ_0 to ψ_a, the same result $A = a$ will be obtained for all elements of the ensemble associated with ψ_a. (ibid., 343)[146]

[145] I write 'possess' to indicate figurative language in the case of contextual properties, given the restrictions imposed by the KS theorem on property possession in BQT.

[146] Bohm himself was explicit on this issue in the 1952 paper: 'the measurement of an "observable" is not really a measurement of any physical property belonging to the observed system alone. Instead, the value of an "observable" measures only an incompletely predictable and controllable potentiality belonging just as much to the measuring apparatus as to the observed system itself' (1952, 183). The reader may confront this analysis of measurement in BQT with Bub's view concerning the real lesson we should learn from von Neumann's 'impossibility proof'. I quote once again the relevant passage: 'So in a hidden variable theory in which dispersion free (deterministic) states are the extremal states for quantum probability distributions, the quantum probabilities could not reflect the distribution of pre-measurement values of beables, but would have to be derived in some other way, e.g., as in Bohm's theory, where the probabilities are an artefact of

In BQT a measurement involves a transformation of the wave-function of the observed system into an eigenvalue of the property being measured in such a way that the obtained eigenfunction gets correlated with an eigenstate of the 'pointer position' observable in the measuring device. For this reason, we cannot measure in one single experimental setup two properties corresponding to non-commuting observables. In general, if a system is described by an eigenfunction of an operator P, it is not an eigenstate of an operator Q that does not commute with P (with the exception of non-commuting observables defined in Hilbert spaces of dimension 3 and greater whose spectra have one eigenvector in common). This is the meaning of the expression $[\hat{x}, \hat{p}] = i\hbar$ in the context of Bohm's theory, it displays the limitations concerning the *accuracy* of measurements that the uncertainty principle imposes: 'if P and Q do not commute, then, by definition, no ψ-function can be simultaneously an eigenfunction of both. In this way, we understand in our interpretation why measurements, of complementary quantities, must (as in the usual interpretation) necessarily be limited in their precision by the uncertainty principle' (Bohm 1952, 183).

This operational limitation implies that it cannot be empirically verified, in a direct way, that quantum particles possess a well-defined position and momentum simultaneously and at all times. But this does not mean that the uncertainty principle proves the stance that quantum systems do not have well-defined position and momentum simultaneously—with the result that 'trajectory', and even 'particle', are meaningless concepts. Actually, from the point of view of BQT, the momentum of a particle can be *inferred* from position measurements. For example, if the wave-function is known so that the momentum $\mathbf{p} = \nabla S$ at each point is given, a precise position measurement allows us to calculate the corresponding momentum. A more concrete illustration is the following. Consider a one-slit experiment in which a particle corresponding to a plane wave is incident on a diaphragm containing the slit. After going through the slit, the wave propagates and hits a detection screen that registers the arrival of the particle. On passing through the slit the energy of the particle remains $E = p^2/2m$, and the momentum p in the x-direction can be calculated from the wave-function. When the particle hits the screen at the coordinate y, the y-component of the momentum of the particle can be calculated from $p_y = p \sin\theta$, with $\theta = \tan^{-1}(y/L)$:



Thus, we conclude that the uncertainty principle does have a fundamental meaning in Bohm's theory. it indicates that a joint direct measurement of position and momentum is impossible--the momentum in the example has been calculated from position, not measured. This impossibility is not due to practical limitations like the accuracy of apparatuses, but to the fact that the involved operators do not commute. However, the restriction imposed by the uncertainty principle is not fundamental in the sense that quantum systems cannot be said to have a well-defined position and measurement simultaneously at the *ontological* level. Though we cannot measure them both with an accuracy greater than $\hbar/2$, this does not mean that the particles do not have determinate values for both properties at the same time. Moreover,

a dynamical process that is not in fact a measurement of any beable of the system. What von Neumann's proof excludes, then, is the class of hidden variable theories in which (i) dispersion free (deterministic) states are extremal states, and (ii) *the beables of the hidden variable theory correspond to the physical quantities represented by the Hermitian operators of quantum mechanics'* (Bub 2010, 1339-40, my emphasis).

the limitation is not even *epistemic*, for we can infer the momentum of a particle from a position measurement.

Another important feature of Bohm's theory that this thought experiment stresses out is that position measurements can be unlimitedly accurate, and, more importantly, that such measurements are an important exception to the 'property change' implied by the observed-system disturbance explained above. That is, though a measurement implies a disturbance of the wave-function of the observed system, the wave-function 'condenses' around the current position of the particle and so we are able, in principle, to infer the premeasurement position as defined by the casual interpretation' (Holland 1993, 351).

This possibility, in turn, gives us to think about the sense in which particle positions constitute the 'hidden variables' in BQT. As it is clear, this variable can be measured and observed in a way such that the premeasurement value is faithfully revealed, so one wonders why is it called 'hidden'. The natural answer is that value of the position variable is not specified by the wave-function—so that we could say that it is *hidden from the formalism* but not from observation. The relevance of being hidden from the formalism is big, though. We may think that the fact that position can be measured without disturbance and with unlimited precision may allow us to go beyond the statistics of the theory and predict experimental outcomes deterministically. But this is not so. Consider the thought experiment just explained. We can say that the position of the particle has been precisely measured without disturbing the premeasurement trajectory. However, we have that once the particle hits the screen, its wave-function has been changed and we cannot calculate the momentum as implied by the newly created wave-function, so that we are not able to predict the subsequent trajectory. This is a general result: though position measurements can be carried without trajectory disturbance, the wave-function is nevertheless altered, so that deterministic predictions cannot be done. That is, there is no way to go beyond the statistical predictive power of the theory. As mentioned, the wave-function does not specify the position of the particle, but if we experimentally determine such a position, we lose the original wave-function.

The way that BQT deals with measurement interaction is quite an attractive feature. As we have seen, no problem regarding definite outcomes arises, so that no mysterious collapse of the wave function nor a corresponding projection postulate is required. Considering how fundamental the measurement problem is in the standard formulation of quantum theory, it is reasonable to argue that with respect to this issue BQT scores better than SQM—and a criterion to make a choice seems to come up. However, as we saw in section 3.3, there are other interpretations of SQM that reject the projection postulate and that intend to account for the determinate outcomes of experiments we observe, while retaining the view that the wave-function gives us a complete description of physical systems and that the dynamics is always governed by the time evolution of the Schrödinger equation.

Both the modal and the Everettian interpretations are heuristic approaches that intend to avoid the measurement problem without assuming wave-function collapse. In the case of the former, we saw that this goal has not been completely achieved. Non-ideal measurements of observables with continuous spectra, even considering the environmental decoherence effect, cannot be accounted for in a fully satisfactory way—the value state defined by the spectral decomposition of the density matrix does not approximate enough to the observed 'pointer position'.

In the case of the Everettian interpretations, the conceptual approach to solve the measurement problem is quite clear: all of the branches of the wave-function are somehow real. However, some technical problems, especially the preferred basis, have not been completely solved—for example, it is not completely clear whether decoherence is enough in order to select the preferred basis that is consistent with the world we observe. Notwithstanding these remaining difficulties, supporters of these interpretations are not in a hopeless situation. Though the problems are there, there is nothing unreasonable in the hopeful attitude that they will be solved through further research—especially concerning the details of decohering processes. Thus, even though BQT does offer an account of measurements that is problem-free, it

could still be argued that in Everettian and modal understanding of the formalism of SQM the measurement problem, in principle, can be coped with. And if for other epistemological reasons we got convinced that the inclusion of hidden variables—in the sense of variables not specified by the wave-function—should be avoided, we may conclude that the Everettian or modal approaches are more attractive. Once again, the strength of the absence of the measurement problem counts as a reason to prefer BQT over SQM granted that some epistemological principles are assumed.

An interesting example of this can be found in an argument that H. Brown and D. Wallace have introduced. These authors offer an analysis of the Bohmian theory of measurements in which the avoidance of the measurement problem relies only on the non-overlapping partial wave-functions that result from a measuring interaction—the 'channels'—and in such a way that the role of the particle is idle. They start with a formulation of what they take to be the Bohmian view, namely, that 'for Bohm it is the entered *wave packet* that determines the outcome; the role of the hidden variable, or apparatus corpuscle, is merely to pick or select that packet from amongst the other non-overlapping packets in the configuration space associated with the final state of the joint object-apparatus system' (Brown & Wallace 2005, 523). Their main point is that an analysis of the account of measurements of BQT of the predictable case—the case in which the initial wave-function of the observed system is an eigenfunction of the relevant observable—shows that the particle does not play any substantial role. They take it that in the predictable case the single wave-packet that results from the measurement interaction does all the work in determining the definite outcome obtained. If this is so, then the same holds in the more interesting case in which several channels results:

> if analysis of the predictable case is successful without appeal to hidden variables, then Bohm's Result Assumption [that it is the entered channel what determines the observed outcome] in the general case is problematic. In the general case, *each of the non-overlapping packets in the final joint-system configuration space wave-function has the same credentials for representing a definite outcome as the single packet does in the predictable case.* The problem, if it is one, is that there is more than one of them. But the fact that only one of them carries the de Broglie-Bohm corpuscles does nothing to remove these credentials from the others. Adding the corpuscles to the picture does not interfere destructively with the empty packets. (ibid., 524, their italics)

Brown & Wallace claim that the wave-packets determined in measurements interactions are the only relevant aspect in the account for definite outcomes of experiments because they interpret the ontology of the quantum wave in BQT as defining fully determinate *worlds*:

> The corpuscle's role is minimal indeed: it is in danger of being relegated to the role of a mere epistemological "pointer", irrelevantly picking out one of the many branches [...] while the real story—dynamically and ontologically—is being told by the unfolding evolution of those branches. The "empty wavepackets" in the configuration space which the corpuscles do not point at are none the worse for its absence: they still contain cells, dust motes, cats, people, wars and the like. (ibid., 527)

Thus, from this point of view the explanation of measurements processes and the definite outcomes obtained that BQT provides boils down to a many-worlds account. If BQT provides a solution to the measurement problem at all it does insofar as it reduces to a many-worlds theory, with the decisive drawback that an ontologically otiose entity is purported: the quantum corpuscle. At first sight BQT may look more economic from an ontological point of view (quantum wave and particles) with respect to the many-worlds approach. However, Brown & Wallace say, if we compare one of the recent developments of the many-worlds view in which the purported ontology is monistic (for example Saunders 1995; Wallace 2002, 2003) with BQT, but considering that the quantum corpuscle does not play any relevant role in the definite outcomes of measurements, then it is the Everettian view which is ontologically more economic. Thus, one it comes to a treatment of the measurement problem, we should prefer the many-worlds approach.

The main premise in Brown & Wallace's argument is that the empty waves of BQT constitute worlds—in the sense that in and by themselves they account for definite outcomes. But this presupposition constitutes a misinterpretation of the ontological meaning of the quantum wave in Bohm's theory. Brown & Wallace refer to three earlier works by D. Deutsch (1996), H. D. Zeh (1999) and D. Wallace (2003) that they take as proving that BQT is a many-world theory in disguise. The arguments offered by Deutsch and Zeh are rather similar. Since the wave-function must be considered as part of the furniture of the world—for it plays dynamical and explanatory roles in physical processes—and since the laws of physics are the same for the empty and the occupied channels, they conclude that empty waves contain worlds[147].

I think that the reasons to include the quantum wave as a part of the fundamental ontology in BQT are quite compelling—though this stance has indeed been challenged in some elaborations on Bohm's proposal, as we will see below. However, that such an inclusion implies that the empty waves represent *worlds* is a blatant *non sequitur*. That the quantum waves determine the trajectories of the particles associated to them and that they produce non-local effects and interference patterns, etc., does not mean that in and by themselves they constitute worlds, or, more precisely, that they determine definite *actual* outcomes in measurement interactions. I think it is rather clear from the account of measurements interactions above that a definite outcome is obtained granted that the quantum corpuscle enters one of the wave-packets. More precisely, that a value a of the measured property is obtained means that the quantum system is described by the corpuscles coordinates $(\mathbf{x}, y)$ lying in the channel $\psi_a(\boldsymbol{x})\phi(y - gaT)$. If in a measurement we make the exercise of suppressing the corpuscles and retain the quantum wave, we have that although the empty wave, under suitable conditions, is still an entity capable of producing non-local effects and interference patterns, this is not sufficient for us to obtain a definite outcome. With respect to the measurement results, the quantum wave plays a sort of *modal* role (similar to the role that the dynamic state plays in the modal approach): if the particle would enter the ath channel, the outcome would be a, but if there is no particle there in the ath channel there is no a outcome[148]. This holds also in the predictable case that Brown & Wallace use in their argument.

Wallace (2003) proposes a criterion, that he takes from Daniel Dennett, for the reality of physical objects according to which it does follow that the empty waves in BQT constitute worlds: 'a macro-object is a

---

[147] 'The question that pilot-wave theorists must therefore address, and over which they invariably equivocate, is what are the *unoccupied* grooves? It is no good saying that they are a merely theoretical construct and do not exist physically, for they continually jostle both each other and the 'occupied' groove, affecting its trajectory [...]. so the 'unoccupied grooves' must be physically real. Moreover, they obey the same laws of physics as the 'occupied groove' that is supposed to be 'the' universe. But that is just another way of saying that they are universes too' (Deutsch 1996, 225).
'It is usually overlooked that Bohm's theory contains the s*ame* "many worlds" of dynamically separate branches as the Everett interpretation (now regarded as "empty" wave components), since it is based precisely on the same ("*absolutely* real") global wave function. Its robust components *branch* by means of decoherence (rather than combining by means of recoherence) because of a fundamental initial condition to the global wave function that is also responsible for the Second Law—not because of "increasing knowledge" about a classical state (the reduction of Bohm's ensemble). Only the "occupied" wave packet itself is thus meaningful, while the assumed classical trajectory would merely point at it: "This is where *we* are in the quantum world". However, this can be done without using a trajectory' (Zeh 1999, 200).

[148] Peter Lewis offers a similar counterargument: 'in order for the empty branch to contain a dead cat, it must not only obey the same physical laws as the occupied branch, but it must also have the relevant physical state—the state that the occupied branch would have if it contained a dead cat. But the physical state of the empty branch is *not* the same as the state the occupied would have if it contained a dead cat. The wavefunction states of the two branches are the same, but according to Bohm's theory, the physical state of a system consists of its wavefunction *and* its particle state. An occupied branch and an empty branch plainly do not have the same particle state, and hence Deutsch fails to establish that empty branches contain measurement outcomes. [...]
It is true that the wavefunction structure is the same in Bohm's theory as in the many-worlds theory, but it does not follow that the empty branches in Bohm's theory are worlds. In the context of Bohm's theory, objects are (prima facie, ate least) made out of Bohmian particles, so the fact that two branches have the same wavefunction structure does not establish that they contain the same objects. But if an empty branch does not contain the same objects as an occupied branch, it is hard to see in what sense it is the same world. Since Zeh provides no reason to think that the wavefunction alone determines the contents of a branch, his argument begs the question against Bohm's theory' (Lewis 2007, 792-3).

pattern, and the existence of a pattern as a real thing depends on the usefulness—in particular, the explanatory power and predictive reliability—of theories which admit that pattern in their ontology' (Wallace 2003, 93). He applies this criterion to standard quantum theory and concludes that each branch of the wave function represents a real cat: 'in each of the branches there is a 'cat' pattern, whose salience as a real thing is secured by its crucial explanatory and predictive role. Therefore, by Dennett's criterion there is a cat present in *both* branches after measurement' (ibid., 97). Then, as an aside, Wallace affirms, reference to Dennett's criterion provides support to the view that the empty branches denote worlds in the same sense as the occupied ones:

> To predict the behaviour of the corpuscles we have to predict the behaviour of the wavefunction, and to predict the behaviour of the wave-function we have to study the emergent patterns within it. Thus, cats and all other macro-objects can be identified in the structure of the wave-function just as in the structure of the corpuscles. But the patterns which define them are present even in those parts of the wave-function which are very remote from the corpuscles. So if we accept a structural characterisation of macroscopic reality, we must accept the multiplicity of that reality in the de Broglie–Bohm pilot wave as much as in the Everettian universal state. (ibid., 99)

Though there are ambiguities in the criterion—especially regarding the meaning of 'predictive pattern'—we can take it as certainly entailing that empty waves constitute worlds and import definite outcomes in measurements. However, why should we take Dennett's criterion for granted as the last world regarding physical ontology? To my mind, at least, it sounds as the outcome of epistemological-metaphysical speculation (and even as especially tailored for the many-worlds interpretation!), and in these matters there is hardly ever general agreement. Actually, as P. Lewis clearly points out, BQT implies a *rejection* of this criterion. In Bohm's theory a Dennettian pattern is insufficient in order to assure the reality of a physical object. A pattern, as defined by the wave-function, must be 'instantiated' by the quantum corpuscle if it is to refer to a real object. Moreover, given the EE between SQM (in its many-worlds version) and BQT, it cannot be said that the empirical evidence favoring one of the theories sanctions a corresponding criterion of reality. As Lewis concludes, 'given that the Bohmian solution to the measurement problem presupposes the falsity of Dennett's criterion, Dennett's criterion cannot be taken for granted in arguing that Bohm's theory fails to solve the measurement problem' (Lewis 2007, 794)[149]. In other words, Brown & Wallace's evaluative comparison between BQT and many-worlds SQM is, in a way, circular. It only works provided that we assume a criterion of reality that is amenable to the many-worlds interpretation, and that the corresponding criterion in BQT is wrong. Only under an assumption like this we could say that the empty waves in BQT correspond to worlds and are enough to determine definite measurement outcomes.

3.6.5 Pauli and Heisenberg's objection: the momentum-position asymmetry

Now we can turn to a consideration of the features of Bohm's theory that may be taken as flaws determining a choice favoring SQM. I will begin with a criticism that played an important historical role. We

[149] Brown & Wallace do not really state that Bohm fails to solve the measurement problem. As mentioned, they claim that BQT solves the problem only insofar as it reduces to a many-worlds theory in disguise. Lewis' criticism can be adapted and leveled against such a view as well. Interestingly, this author elaborates on his criticism in a way that converges with something I have mentioned a number of times: how we evaluate some features of BQT and SQM—as flaws or virtues that can be invoked to make a decision—depends on previous epistemological or ontological commitments which, given the EE at issue, cannot be sanctioned by the empirical evidence: 'suppose the tables were turned; a Bohmian who took for granted all our intuitive views about probability and identity and argued on that basis that the many-worlds theory fails to solve the measurement problem would rightly be accused of begging the question against the many-worlds theory. This is because the many-worlds solution is predicated on the rejection of some or other of our intuitive views. But then, by the same token, the fact that Bohm's theory fails as a solution to the measurement problem given Dennett's criterion should not be held against it, since Bohm's theory is predicated of the rejection of Dennett's criterion' (ibid., 794-5).

saw above that von Neumann's 'impossibility proof' of 1932 had already set a hostile context towards HVT, which may explain why Bohm's 1952 proposal did not receive much attention. However, some of the great names of the times, namely, Einstein, Pauli and Heisenberg, did dedicate some words to Bohm's theory. Anyhow, the fact that these appraisals consisted in harsh criticisms contributed even more to the fact that BQT was not taken as a viable program. We have already dealt with Einstein's criticism in the context of the classical limit in quantum theory, so we now turn to Pauli and Heisenberg[150].

In a 1952 collection of essays dedicated to de Broglie, Pauli criticized Bohm's theory on the basis that 'the artificial asymmetry introduced in the treatment of the two variables of a canonically conjugate pair characterizes this form of the theory as artificial metaphysics' (from Pauli 1952, translated and quoted in Myrvold 2010, 11). Pauli's point can be clarified by an analogy between classical mechanics and SQM, which does not hold in the case of Bohm's theory. In classical mechanics, a canonical transformation is a change of canonical coordinates $(\mathbf{q}, \mathbf{p})$ in phase space that preserves the form of Hamilton's equations—although the Hamiltonian itself may not be preserved. That is, the Hamilton equations $\frac{dq}{dt} = \frac{\partial H}{\partial p}$ and $\frac{dp}{dt} = -\frac{\partial H}{\partial q}$ transform into the equations $\frac{dQ}{dt} = \frac{\partial H}{\partial P}$ and $\frac{\partial P}{\partial t} = -\frac{\partial H}{\partial Q}$, with $Q = p$ and $P = -q$. Something analogous holds in SQM:

> any unitary transformation
>
> $$\hat{q}_i \to \hat{U}^\dagger \hat{q}_i \hat{U} \qquad \hat{p}_i \to \hat{U}^\dagger \hat{p}_i \hat{U}$$
>
> leaves the canonical commutation relations, and hence the equations of motion, invariant. There is, in particular, a unitary transformation that (up to change of sign and multiplication by an arbitrary constant) interchanges position and momentum:
>
> $$\hat{q}_i \to \hat{p}_i/\alpha \qquad \hat{p}_i \to -\alpha \hat{q}_i$$
>
> (here α is an arbitrary constant of dimension mass/time). (Myrvold 2003, 18)

In SQM the canonically conjugate quantities q and p are treated on a par—in the sense that the basic formulas of the theory are invariant under substitutions $x \leftrightarrow p$ and $i \leftrightarrow -i$. According to Pauli, it follows that neither quantity can be considered as more fundamental than the other. In Bohm's theory, though, momentum and position are treated in an asymmetric way. As we saw above, only position measurements can draw outcomes that correspond to the pre-measurement value of the property, whereas in the case of momentum, the result obtained in a measurement indicates the value of the property once the system has been transformed by the measurement interaction. This means, according to Pauli, that an unjustified treatment of position as a more fundamental physical quantity than momentum:

> As Bohm points out, the result of this interaction [between system S and experimental apparatus A] naturally depends also on the values of the parameters of A. What has been said here about position and momentum is valid generally for any pair of canonically conjugate variables: at most the values of *one* of these variables can be interpreted as a "property of S". This strips of its physical sense the simple passage via Fourier analysis from a wave function to a function of the conjugate variable (which leaves us the choice of considering either one or the other as the "primary" function), or introduces an asymmetry with regard to the interpretation of canonically conjugate magnitudes for which one finds reason neither in the system of our experiences nor in the mathematical formalism of wave mechanics. (From Pauli 1952, quoted in Myrvold 2010, 11-2)

[150] This overview of Pauli and Heisenberg's criticism is based on (Myrvold 2003). This article also contains a brief treatment of Einstein's objection.

Heisenberg's objection (1955; 1958) is similar. In a rather positivistic vein, he claims that Bohm's theory is nothing but a repetition of (the Copenhagen version of) SQM in a different language, and that, therefore, the Bohm's proposal cannot be refuted by experiment—meaning that no different or further empirical predictions can be obtained from it. From this point of view, there is no real choice to be done, of course. However, Heisenberg states that the language of SQM (in its Copenhagen semantics) is preferable to the Bohmian alternative. The reason for this is twofold. First, Heisenberg complains about the trajectories purported by BQT, for they are not observable and do not contribute to derive predictions foreign to SQM—they constitute unnecessary 'ideological superstructure'. Secondly, this the use of this 'language' results in the asymmetry between momentum and position that Pauli pointed out:

> Besides the objections already made that in speaking of particle orbits we are concerned with a superfluous "ideological superstructure", it must be particularly mentioned here that Bohm's language destroys the symmetry between position and velocity which is implicit in quantum theory [...]. Since the symmetry properties always constitute the most essential features of a theory, it is difficult to see what would be gained by omitting them in the corresponding language. Therefore, one cannot consider Bohm's counterproposal to the Copenhagen interpretation as an improvement. (Heisenberg 1958, 133)

The main reason why Pauli considers that the disparity between position and momentum constitutes a flaw in Bohm't theory is analogous to the reason why Einstein criticized the postulation of the ether: the resulting asymmetries are artificial and have a metaphysical flavor. That is, the criticism is not merely a matter of aesthetical taste, for, according to Pauli, the momentum-position asymmetry in Bohm's theory reflects its dubious foundations.

However, Myrvold (2010, section 5) compellingly argues that the momentum-position symmetry, both in classical and (standard) quantum mechanics, does not reflect a fundamental aspect of their descriptions of the physical world. Therefore, the symmetry violation in Bohm's theory does not have to do with a foundations issue—according to Myrvold, the situation is not analogous to the asymmetries that do not belong to the phenomena. In the case of classical mechanics, we have that the Hamilton equations retain their form under *arbitrary* canonical transformations, independently of the form that the Hamiltonian takes under the transformations. Consequently, the fact that the particular canonical transformation $Q = p,\ P = -q$ leaves the Hamilton equations invariant does not really express a symmetry pertaining to the physical systems involved. Actually, this kind of invariance is simply a mathematical property of phase space, but the structure of phase space does not determine the physical behavior of systems in a fundamental way—certainly not as Minkowski space-time does. Special relativity, in its Minkowski space-time formulation, is clearly a chrono-geometric theory, so its symmetries do determine (or do reflect, depending on the ontology of space-time we assume) the kinematical features of spatio-temporal objects. The momentum-position symmetry connected to the canonical transformation in phase space is not a symmetry of this kind:

> There is, therefore, a profound disanalogy between the invariance of, say, the equations of electromagnetism under Lorentz transformations, and the invariance of the Hamiltonian equations of motion under the canonical transformation $[Q = p, P = -q]$. One could say: the former reflects a symmetry of the physical system under consideration, the latter a symmetry of the phase space. The fact that structure of phase space does not distinguish between position and momentum should not be taken as an indication that there is no important physical distinction between position and momentum—and herein lies the disanalogy, because the fact that the spacetime, Einstein-Minkowski spacetime, invoked in special relativity is invariant under Lorentz transformations should be taken as an indication that, if the theory is correct, distinctions not preserved under Lorentz transformations do not belong to the structure of physical spacetime. (Myrvold 2003, 18)

Now, since the symmetry between momentum and position present in SQM is of the same nature of the corresponding symmetry in classical mechanics, it follows that the conclusion that a fundamental flaw of the theory is expressed in the symmetry violation in BQT is not justified:

> The canonical commutation relations are preserved under *arbitrary* unitary transformations (the analogue, in quantum mechanics, of classical canonical transformations), no matter what form the Hamiltonian takes. The fact that the fundamental equations of the theory are invariant under the transformation $[\hat{q}_i \rightarrow \hat{p}_i/\alpha\,, \hat{p}_i \rightarrow -\alpha\hat{q}_i]$, therefore, yields no information about the symmetries of a system. (ibid., 19)

Thus, Pauli and Heisenberg's criticism cannot be taken as an objection leveled on the basis of foundational problems in Bohm's theory. Their attack can only be grounded on aesthetic criteria according to which formal symmetries like the one between momentum-position are desirable features in a theory. It is clear, though, that in this case the argument against Bohm's theory becomes rather weak and infected with subjective considerations—as it is always the case in aesthetic matters[151].

To conclude this section we may consider Heisenberg's remaining objection that the quantum particle trajectory is unobservable and therefore constitutes superfluous ideological superstructure. As we saw above, it is true that precise joint measurements of momentum and position cannot be performed for fundamental reasons. However, we also saw that the momentum of a particle can be deduced from a precise position measurement in suitable contexts. I think that this is a clear indication that the concept of quantum trajectory, though directly unobservable, does not reduce to metaphysical baggage in the theory. Moreover, in the discussion of measurements and empty waves in BQT we saw that particles entering a specific channel is what determines definite outcomes in measurements: it is the particle following one of the possible trajectories given by the quantum potential what determines experimental results. Thus, from the Bohmian point of view it is natural to take the mark a particle produces when it hits the detection screen as an empirical trace of the trajectory the particle followed. Heisenberg's commitment to positivistic principles are well known, and from such a stance a strong disregard of particle trajectories makes full sense. However, from a less prejudiced standpoint, we have that direct unobservability is not reason enough to immediately qualify a term as not referring to anything physical—there are many examples in the history of science. As Michel Janssen comments with respect to unobservability of the ether in the Einstein vs. Lorentz case, 'the problem with this type of argument is that it derives its force from a blanket rejection of unobservables in scientific theories, whereas it is widely accepted that such elements should not be banned automatically. Rather than condemning unobservables in general, I think it is wiser to demand arguments to put forward on a case-by case basis to show why a particular unobservable is otiose' (Janssen 2002, p. 438). In this particular case, there are not reasons enough to consider the concept of a particle trajectory as empirically otiose.

3.6.6 The foundations of probability

In section 3.5.4 we saw that in Bohm's theory $P = |\Psi|^2$ represents the location distribution of an ensemble of particles associated to Ψ, and hence it gives us the probability for a particle associated to Ψ to

[151] Moreover, the asymmetry can be provided physical justification within the Bohmian framework. As Callender & Weingard state: 'the simple reason why position is special in Bohm's theory is that the measurement problem requires it to be. Arguably, all measurements are the measurements of the positions of things. That is partly why Bohm's theory works. But the other reason [...] concerns an asymmetry between position measurements and momentum measurements: "effective collapses" of position superpositions also collapse momentum superpositions, but not *vice versa*. An effective collapse of a momentum superpositions will sometimes leave a system in a *superposition of macroscopically distinct positions*. Since solving the measurement *just is* ensuring that this doesn´t happen, a momentum-space version of Bohm's theory—while mathematically possible—cannot solve the measurement problem in the real world' (1997, 26-7).

be at a specific location. The formula $\frac{\partial P}{\partial t} + \nabla \cdot \left(P \frac{\nabla S}{m}\right) = 0$ assures that this probability distribution is conserved over time. That is, the way that probability enters in Bohm's theory is analogous to the way in which probability enters in classical statistical mechanics. As it was also mentioned, this distribution postulate determines that the observable predictions of SQM and BQT coincide.

So far, so good. But we may of course ask *why* the distribution postulates holds. There is no *a priori* reason forbidding that the particles associated to a wave-function Ψ are spatially distributed in a way such that $P \neq |\Psi|^2$. The easy way out would be just stating that there is one initial condition such that it leads to $P = |\Psi|^2$—once quantum equilibrium is reached, the formula $\frac{\partial R^2}{\partial t} + \nabla \cdot \left(R^2 \frac{\nabla S}{m}\right) = 0$ guarantees that it will be conserved . There is nothing intrinsically wrong or incoherent with this statement, but if a full explanation of the distribution postulate is required, this would not be accepted as a satisfactory answer. This is especially clear if we recall that in BQT the probability distribution is interlocked with the dynamics—both P and Q are determined by the wave-function. Given the fundamental importance of this postulate in the theory, to leave its foundations to the "chance" or contingency of the initial conditions of the universe is not really an explanatory view.

The 'initial conditions' answer could be given more explanatory content by showing that *every* or *most* initial states lead to the distribution postulate. The first option is not possible, for there exist well-defined wave-functions that are not in quantum equilibrium and that cannot evolve towards it. The second approach was the option that Bohm took originally. The basic idea, as we saw above, was that the chaotic interactions between the particles in an ensemble are such that quantum equilibrium is quickly approached. That is, the dynamics are supposed to force non-equilibrium states into equilibrium, or at least states which are observationally indistinguishable with respect to $P = |\Psi|^2$. Actually, for empirical adequacy all that is needed is that the dynamics assure that distributions such that $P \neq |\Psi|^2$ quickly approach $P = |\Psi|^2 \pm \varepsilon$, where ε is within the range of observational error. Bohm (1953) showed that in a specific system of molecules out of equilibrium subject to random perturbations, a distribution according given by $|\Psi|^2$ is reached. A *supplementation* of the dynamics has also been attempted. Bohm & Hiley (1993), for example, introduce stochastic terms that operating in a sub-quantum level are responsible for equilibrium approach.

However, though both the dynamical attempts mentioned contribute in showing that given certain initial states quantum equilibrium is approached, a general justification for most initial conditions has not been achieved. Another challenge that the justification of the distribution postulate must face is given by the distinction between the universal and the effective wave-function. From our discussion of non-locality in BQT it follows that if a many-particles system is not factorizable, then an 'independent' wave-function for the subsystem constituted by a single particle does not exist, and in this sense we cannot say the such a subsystem is governed by Bohmian mechanics, and 'therefore, in a universe governed by Bohmian mechanics there is a priori only one wave function, namely that of the universe, as there is a priori only one system governed by Bohmian mechanics, namely the universe itself' (Dürr, Goldstein & Zanghi 1995, 5). However, when it comes to predictions for measurement results—whose statistics are given by the quantum equilibrium hypothesis—it is not the universal wave-function that is relevant, but the wave-function of the particular subsystem of the universe under scrutiny. Thus, a condition for measurements is that such a subsystem can be 'factorized out' from the rest of the universe. The condition for this has been introduced by Dürr, Goldstein & Zanghi:

> A subsystem has effective wave function ψ (at a given time) if the universal wave function $\Psi = \Psi(x, y)$ and the actual coordinates $Q = (X, Y)$ (at that time) satisfy
>
> $$\Psi(\mathrm{x}, \mathrm{y}) = \psi(x)\Phi(y) + \Psi^{\perp}(x, y)$$
>
> with Φ and $\Psi^{\perp}$ having macroscopically disjoint y-supports, and

$Y \in \text{supp}\ \Phi$

> Here, by the macroscopic disjointness of the y-supports of Φ and $\Psi^{\perp}$ we mean not only that their supports are disjoint, but that there is a macroscopic function of y whose values for y in the support of Φ differ by a macroscopic amount from its values for y in the support of $\Psi^{\perp}$. (1992, 863)

The *generic* configuration of the spatial variable is here given by $q = (x, y)$, whereas the *actual* configuration of particles is given by $Q = (X, Y)$. That is, $\Psi(x, y)$ gives us only the universal wave-function with the possible trajectories it defines, whereas the *state* of the universe is given by (Q, Ψ). The universe-system has been split into the x-subsystem (which in turn can be subdivided in observed system and apparatus) and the environment y-subsystem. The condition $\Psi(x, y) = \psi(x)\Phi(y) + \Psi^{\perp}(x, y)$, with $Y \in$ supp Φ, means that the support of $\Psi^{\perp}$ is given by empty waves, whereas the actual positions (X, Y) correspond to the support of ψ and Φ. Thus, the 'observed system + apparatus' subsystem can indeed be factorized out from the rest of the universe and described by its own effective wave-function $\psi(x)$ which is suitable for measurement predictions.

Now, the distinction between effective and universal wave-function is relevant in the context of the discussion of the distribution postulate because the equation $\frac{\partial P}{\partial t} + \nabla \cdot \left(P \frac{\nabla S}{m}\right) = 0$ holds for the latter. Though this equation assures that the dynamics preserve $P = |\Psi|^2$, it does not imply that the dynamics preserve $\rho = |\psi|^2$—where Ψ stands for the universal wave-function and ρ is the location distribution of the particles in the ensemble associated to the effective wave-function ψ. The two outlined approaches to the justification of the quantum equilibrium hypothesis do not deal with this distinction, but it is clear that the real goal is to justify $\rho = |\psi|^2$ rather than $P = |\Psi|^2$.

Dürr, Goldstein & Zanghi directly address the task of justifying the distribution postulate as holding for effective wave-functions. Their approach is based on the concept of *typicality*. In mathematics, given a number $x = y.x_1x_2x_3$, we say that x is *normal* if and only if for all bases and all subscripts i, $1/10^{\text{th}}$ of the x_i's are zero, $1/10^{\text{th}}$ of the x_i's are one, $1/10^{\text{th}}$ of the x_i's are two, and so on. It is a theorem (Borel's) that the complement of the collection of normal numbers in the interval [0,1] is of zero length given a Lebesgue measure. That is, "almost all" numbers x are normal, and in this sense, normal numbers are *typical*. The details of the argument by Dürr, Goldstein & Zanghi are rather intricate, but the basic idea is that

> the counterpart of our numbers x are histories of the Bohm particles $Q(t)$. As we found patterns in the x's, we try to find patterns in the configuration variable histories. In particular, the pattern we are interested in is the one wherein the particles for *subsystems* of the universe are distributed according to $|\psi|^2$ when the subsystem merits an effective wave-function, just as we might look for the relative frequency of fives in a real number x. The *a priori* measure used by DGZ, the counterpart to Lebesgue, is the natural volume measure on configuration space, modified by $|\Psi(q_1 \ldots q_N; 0)|^2$:
>
> $$|\Psi(q_1 \ldots q_N : 0)|^2 d^{3N} q.$$
>
> Note that [this equation] uses the wavefunction of the universe. (Callender 2007, 362-3)

The conclusion that DGZ draw is that ensembles described by effective wave-functions with a distribution $\rho = |\psi|^2$ are typical. This result seems to be a full attainment of the goal set. However, one can still pick on the choice of the *a priori* measure: $|\Psi|^2$ has been chosen as a measure to prove the typicality of $|\psi|^2$. Though the argument is not circular, other measures could have been used instead. DGZ argue that even though $|\Psi|^2$ is not a choice dictated by the logical features of the argument, this value for the measure is suggested by its special dynamical properties: it is preserved by the dynamics, trajectories as

measured by it do not run into nodes, escape to infinity or run into singularities of the potential. Anyways, this argument is more a piece of mathematics than a substantial *physical* justification of the distribution postulate. To obtain such a justification an interpretation of what it means that effective wavefunctions in quantum equilibrium are typical is needed. That is, in order to this mathematical result to count as a fully satisfactory explanation of the distribution postulate we may first venture into the realm of the interpretation of probability—which is an open philosophical issue.

In his critical article on Bohm's theory (1953), Pauli also mentions that simply stipulating the distribution postulate is not an acceptable move in BQT. He demanded for its derivation in terms of more fundamental features of the theory. That is, he took the lack of explicit justification of the quantum equilibrium hypothesis as yet another reason to reject Bohm's proposal. The sketch of the dynamical and typicality attempts to provide such a justification show that, although the foundations of probability in BQT is an issue that still requires clarification, the situation is not as problematic as to imply a fundamental conflict in the foundation of the theory such that its plain rejection is suggested. Many open questions remain, but this does not mean that the distribution postulate cannot be, in principle, justified—especially if we consider that the approaches mentioned are work in progress. The DGZ group continues working on the typicality approach, and Valentini & Westmann (2005) have proposed a new dynamical method. Besides, the formal result achieved through typicality is highly valuable. Though it is true that questions regarding what typicality means still stand, such questions have more to do with the interpretation of probability in the context of physical theories as a general issue rather than with BQT in particular. In this sense, the situation in Bohm's theory concerning the foundations of probability is similar to the situation in classical statistical mechanics, and the remaining questions in the latter case certainly do not suggest the rejection of the theory. As Callender states,

> But what is the justification of the microcanonical probability measure in classical statistical mechanics? Few questions in all of the foundations of physics are more vexed and yield such a diversity of answers. Some believe the thermodynamic limit solves the foundational puzzles, others environmental perturbations, others symmetries and ignorance, others mixing dynamics, and others the H-theorem. The answers these theories give are sometimes no more similar than chalk and cheese. Faced with this foundational bedlam, we see that criticisms Like Pauli's [...] are problematic: one certainly cannot do uncontroversially for the microcanonical probability distribution what they want done for $[\rho = |\psi|^2]$, yet presumably neither recommend the rejection of classical statistical mechanics. Today, most Bohmians agree that the distribution postulate has roughly the same justification as the microcanonical probability posit—so long as no one asks what that justification is! If one asks this question, foundational bedlam arises again, this time at the subquantum level. (2007, 355)

On the other hand, there is nothing intrinsically wrong or inconsistent with the 'default position', that is, the view that the quantum equilibrium hypothesis is an axiom of the theory that does not necessarily require further explanation—we may simply adopt the stance that there is a class of initial conditions that result in quantum equilibrium and assume that the universe developed from an initial state belonging to such a class. That the default position may be regarded as incomplete does not mean that it is an unacceptable view—and this reflects that the foundations of probability do not constitute a fundamental problem in BQT[152]. Anyways, it is true that when it comes to a comparative evaluation with respect to SQM, we have that in the latter, at least in the Copenhagen interpretation, there is no need for a justification of the Born rule, for indeterminism is assumed as a fundamental aspect of physical reality. The rule

[152] Consider for instance the view endorsed by Callender & Weingard: 'Our opinion is that we should abandon the search for a justification of Born's rule. The world began with initial conditions in the Good Set. That is just the way it is. Nothing more can be said, for the Good Set is not probable (for what would this mean?), and if it is "natural", this is not explanatory. Explaining Born's rule in a deterministic theory like Bohm's means explaining the boundary conditions of the universe, and it is just here that we believe explanation must come to a halt' (1997, 40).
Callender expresses a softened position in his (2007): 'The Default Position seems defensible. The question is then only whether we can do better' (356).

is not merely a description of the distribution of particles in an ensemble, its essential meaning is simply to determine the probability of finding certain experimental results. However, though no *justification* of the Born rule is needed, we may still ask about the meaning of the probabilities that the rule *expresses*. Heisenberg, for example, took them as manifesting an objective tendency or *potential* in the Aristotelian sense. In contrast, the answer for *this* question that BQT gives is much clearer: the probability expressed by Born's rule is simply the distribution of the particles in an ensemble. That is, though in SQM there is no distribution postulate to be explained, the question about the meaning of probability is also relevant and open in this case.

3.6.7 The ontological status of the wave function

As it has already been mentioned a number of times, the basic ontology that is naturally associated to Bohm's theory is given by the quantum corpuscles and the guiding or pilot-wave. We also saw that the latter is a rather *sui generis* entity. Unlike the usual fields described in other physical theories, the quantum field of Bohm's theory does not have any source. The pilot-wave determines the possible trajectories that a particle can follow, but the particle's motion along one of such paths does not affect back the wave—it does not respect Newton's third law. The action of the wave-function depends on its form, not on its amplitude, so, in general, the action of the quantum field does not diminish as the distance increases and its amplitude diminishes. Finally, the wave-function is defined and described in $3N$-configuration space (this property is closely related to the non-local effects that the quantum field determines), not in 3-space, as the typical waves and fields of other theories do[153]. Thus, it is clear that this ontology is committed to a very strange entity that does not have any counterpart in other physical theories. This commitment opens a flank for criticism in the Bohmian approach. It has been argued that the postulation of such a weird entity is not justified[154], especially when it is considered that it cannot be directly observed or measured[155]. Thus, the strange ontological status of the quantum field may be taken as advising the rejection of Bohm's theory.

The response that some Bohmians have given to this challenge is simply to avoid any ontological commitment towards the wave-function and interpret it in a nomological way (Dürr, Goldstein & Zanghi 1995; Goldstein & Zanghi 2012). The basic idea is simple. The wave-function is taken to be a component of physical law, not a term describing an entity in the physical world. The rationale for this answer is given by means of an analogy with respect to the role that the Hamiltonian plays in the phase-space formulation of Newtonian mechanics:

> We propose that the reason, on the universal level, that there is no action of configurations upon wave functions, as there seems to be between all other elements of physical reality, is that the wave function of the universe is not an element of physical reality. We propose that the wave function belongs to an alto-

[153] If we consider the distinction between the universal and the effective wave-function explained above, we notice that it is the former that constitutes the basic ontological element of the theory (along with the corpuscles, of course).

[154] Anandan & Brown (1995) point out the troubles with the reality of the wave-function on the basis of its violation of the Action-Reaction principle of reality—a sort of ontological generalization of Newton's third law. Monton (2002, 2006) picks on the fact that the wave-function is defined in $3N$-space.

[155] That the wave-function is not directly observable does not mean that there are no empirical traces that the Bohmian could take as manifestations of its existence. Just recall the crucial importance of the quantum potential Q for the derivation of predictions. That is, the observation of typical quantum effects like interference patterns in particle trajectories can be taken as indirectly witnessing the existence of the wave-function. The situation is analogous to the case of the ether: the v-dependence of mass can in principle be taken as a trace of the ether. I showed in chapter 2 that when Poincaré established that the Lorentz transformations are symmetric this stance was no longer possible—for the ether became idle for the formal derivation of predictions. However, it is obvious that in BQT the quantum potential is essential in the derivation of predictions—though directly unobservable, it is not *superfluous*.

> gether different category of existence than that of substantive physical entities, and that its existence is nomological rather than material. We propose, in other words, that the wave function is a component of physical law rather than of the reality described by the law.
> We note in this regard that nobody objects to classical mechanics because it involves a Hamiltonian $H_{class}(\mathbf{q}_1, \dots, \mathbf{q}_n, \mathbf{p}_1, \dots, \mathbf{p}_n) \equiv H_{class}(\xi)$ that is a function on a space, the phase space, that is of greater dimension and even more abstract than configuration space. This is because we think of the state in classical mechanics as given by the q's and p's, and we regard the Hamiltonian as the generator of the evolution of the state—i.e., as part of the law—and not as an object in whose behavior we are directly interested. (Dürr, Goldstein & Zanghi 1995, 10)

It is important to underscore that the wave-function that is interpreted in a nomological way is the universal one, not the effective one. This remark blocks a possible objection: we can arbitrarily prepare effective wave-functions in the laboratory, but this does not mean that we are manipulating the laws of nature at our will—the universal wave-function cannot be manipulated. But then two technical questions naturally arise. First, the effective wave-functions that obey the Schrödinger equation and govern the parcels of the universe we treat as subsystems are time-dependent, and we may ask how these time-dependent wave-functions emerge from a nomological universal wave-function that is assumed to be fixed and timeless. Second, the universal wave-function as specified by the Schrödinger equation is not unique, under different initial conditions different wave-functions result. The problem with this is that there are no other clear elements in the theory that indicate how the form of the universal wave-function can be uniquely specified—if this uniqueness condition cannot be grounded it is difficult to conceive Ψ as nomological, for a law of nature whose form can vary according to different initial conditions is a rather unappealing idea.

A proposed answer to the first question is that the equation that determines the wave function of the universe is the Wheeler-de Witt equation, in which time plays no role. Though the Wheeler-de Witt equation *determines* the universal wave-function, the so determined wave-function is a solution of the Schrödinger equation—but no fundamental role is assigned to the latter. Now, that the universal wave-function is a solution of the Schrödinger equation assures that the effective wave-functions are also so. Therefore, the time-dependence of the effective wave-functions is inherited from a timeless universal wave-function determined by the Wheeler-de Witt equation. As to the second question, a tentative answer has been suggested by Goldstein & Teufel (2001). If it is demanded that the universal dynamics of the theory is given by a first-order and covariant equation, these demands result in that strict constraints on the form of the universal wave-function are posed—besides, that the solution of the Wheeler-de Witt equation may be unique would provide further assistance, of course. As it can be seen (in the essential reference to the Wheeler-de Witt equation), these answers are highly speculative. But this could be a nice thing in the long run, for the Bohmian approach may turn out to be more fruitful, adequate and less problematic than SQM when it comes to the quest for a quantum gravity or quantum cosmology theory.

An important remark concerning the nomological approach to the wave-function is that, as Esfeld et al. state (2013), it can take two different forms. The difference between them is given by the two different usual epistemological views that can be taken regarding the status of the laws of nature. Under a Humean standpoint, laws of nature simply express certain contingent regularities and have a merely inductive support. If this view is adopted in the case of the nomological wave-function in BQT, we have that the universal wave-function simply denotes the collective pattern of behavior of all the quantum corpuscles in the universe[156]. An important feature of this approach is the meaning of entanglement and the connected non-locality present in the theory. Since this property resides in the wave-function, and since the

[156] 'Humeanism about laws is applicable to Bohmian mechanics. Assume that one knows the positions of all the particles in the universe throughout the whole history of the universe. The wave-function of the universe, then, is that description of the universe that achieves, at the end of the universe, the best balance between logical simplicity and empirical content. In other words, the wave-function of the universe supervenes on the distribution of the particles' positions throughout the whole of space-time; the same goes for the law of motion. This supervenience relationship implies that the wave-function

wave-function is a Humean law, then entanglement and non-locality are mere contingent regularities expressed in the spatio-temporal behavior of the particles as well:

> The mosaic of the particle positions in the actual world happens to be such that an entangled wave-function figuring in a non-local law of motion supervenes on it. But there is no real physical relation of entanglement that exists as a non-supervenient relation in four-dimensional space-time in addition to the relations of spatio-temporal distance among the particle positions. By the same token, there is not any sort of holistic physical property instantiated in space and time over and above the local particle positions' (Esfeld et al. 2013, 10-1).

The alternative standpoint is given by a dispositional conception of the nomological wave-function. According to this view, laws of nature express the behavior of physical objects as determined by certain dispositional properties possessed by such objects. The support of laws of nature thus understood is not merely inductive, there is an underlying basis in reality, the dispositions, that determines the regularities described by the laws. In the case of BQT, we have that since it is the universal wave-function that plays the role of a law, then the disposition that determines the trajectories of the quantum particles is possessed by the collection of all the particles in the universe as a whole. Now, since the disposition described by the nomological wave-function is holistic in this sense, then this approach can naturally accommodate entanglement and non-locality:

> On this view, the universal wave-function, Ψ_t, of the system of particles at a given time is a mathematical object that represents the disposition to move in a certain manner at that time. This disposition is a holistic property of all the particles in the universe together—that is, a relational property that takes all the particles as relata. It induces a certain temporal development of the particle configuration, that development being its manifestation. In other words, given a spatial configuration of the particles (actual or counterfactual) and the disposition of motion at a time as represented by the wave-function as input, the Bohmian law of motion yields the velocities of the particles at that time as output. (Ibid, 13)

There are several stringent problems with the nomological conception of the wave-function in BQT. First, we have that a good deal of the explanatory power that the dualistic ontology (particle and pilot wave) offers gets lost. For example, consider the simple explanation that we reviewed above in the case of neutron interferometry experiments—the quantum wave played a central role in that account, of course. Another example is given by the explanation that BQT, in its dual ontology, offers for the EPR correlations. Assume that an entangled system in the singlet state is constituted by two space-like separated particles, and that the spin *z*-direction of only one of the particles is measured with a Stern-Gerlach device. The interaction of this particle with the device draws a spin value, up or down, and we know that the disturbance of the measurement device also determines the spin-value of the non-measured particle—according to the correlations in the singlet state. The natural explanation for this in BQT is that since the wave-function, ontologically understood, is 'non-locally extended' (it is defined in the 6-dimensional configuration space of the two particles), the disturbance of the wave-function by the S-G device is the cause of both spin values[157]. That is, though the particles are space-like separated, the 'non-locally extended' wave acts as the substratum supporting the non-local effect that the disturbance on the measured particle infringes on the non-measured one.

---

of the universe applies not only to the actual distribution of particle positions throughout space-time, but also to other possible distributions. Given that we are ignorant about the exact positions of the particles in the universe, we then get, through the quantum equilibrium hypothesis, effective wavefunctions and quantum statistics as the best description we can achieve for subsystems of the universe. But note that on this view, only the universal wave-function that supervenes on the particles' positions throughout the whole history of the universe has a nomological status. No effective wave-function describing subsystems can claim such a status' (Esfeld et al. 2013, 9-10).

[157] Whether the non-measured particle has an 'actual' or 'potential' spin value depends on what interpretation of spin one assumes. For a detailed and technical account of the EPR correlations in the context of BQT, see (Dewdney, Holland & Kyprianidis 1987).

If the wave-function is excised from the ontology of the theory and assigned only a nomological meaning, the explanation provided for these examples is not that natural anymore. Actually, in the Humean approach, no explanation whatsoever is available, for the non-local correlations of the EPR experiment and the peculiar results of neutron interferometry experiments are nothing but contingent regularities—this question is an essential feature of the Humean standpoint of laws nature in general, of course, but it becomes particularly clear (ad stringent) in a quantum context. The dispositional attitude can still provide an account, but it is impoverished and almost a *de facto* one: non-local and interference effects occur because the dispositions of the collection of particles in the universe say so. Besides, the ontology that the dispositional approach presupposes carries some metaphysical baggage, the dispositions, which is rather unattractive from the framework of modern science.

The supporter of the nomological stance could reply that the dispositional explanation is enough, or that the Humean approach is such that no explanation is needed. However, there are two other stringent problems that, to my knowledge, have not been considered in the discussion of this subject. First, as it can be inferred from an analysis by Brown, Dewdney & Horton (1995), there are physical reasons that suggest that the wave-function should be regarded as real in the context of BQT. Consider again the neutron interferometry experimental setup described in 3.6.1. We saw that, assuming the dualistic ontology, all the peculiar results get explained by the fact that the quantum corpuscle travels along one single path, while the quantum wave travels along both. Brown, Dewdney & Horton consider an interesting variation of the setup. Assume that path *II* is the one the particle takes and path *I* is the empty. As it is clear, with path *I* blocked the exiting neutrons would be deflected in two paths *0* and *H* by the last crystal in a half-half proportion (due to the trajectories no-crossing principle in BQT), whereas if path *I* were left open all the identically prepared neutrons would be deflected along the same outgoing path. The setup variation consists in that the interferometer, that was originally lying horizontally, is rotated in a way such that paths *I* and *II* (both open) stand in different heights in the gravitational field of the Earth. The result is that the outward neutrons get deflected along the two possible paths, but on a proportion which is not half-half—this is due to a phase-shift induced by the rotation angle and the gravitational mass of the neutron, gravitation acts as a phase-shifter in this case. The point is that since gravitation determines this outcome by affecting the partial wave along path *I* (if path *I* were blocked then all the neutrons would be deflected along the same outgoing path) the conclusion that Brown, Dewdney & Horton draw is that what they call the *localized particle properties hypothesis* (LPP) is violated: the mass of the neutron cannot be thought as fully localized in the corpuscle:

> Notice now that a significant number of the Bohm path *II* particles must enter the *0*-beam, which would have not done so were path *I* blocked, because of the no-crossing principle. This number depends on *inter alia* the angle α and the mass of the neutron. How can this effect on path *II* particles in the region of the third crystal plate be understood intuitively if the empty path *I* carries no gravitational mass? How is the difference in the gravitational potential integrated over the two paths felt by the particle after traversing path *II* if all its gravitational mass is concentrated in that path? It is difficult to avoid the conclusion that gravitational mass provides a counter example to the LPP thesis. (1995, 338-9)

It is clear that we can use this analysis to draw a more basic conclusion, namely, that the wave-function has to be taken as denoting something real. The nomological approach surely faces a difficulty in explaining how the gravitational effect involved in this setup determines the outcome if there is just no wave at all. In that case, the LPP thesis should hold—if there is no wave the gravitational mass of the neutron must be localized in the corpuscle. But then it is difficult to conceive a suitable interpretation or description of how gravitation produces the phase-shift that results in deflection along the two outgoing beams. I am not saying that the nomological approach is in principle incapable of providing an explanation, but I think it is rather clear that any possible explanation would be rather contrived, especially compared to the dualistic ontology account. Though Brown, Dewdney and Horton's analysis indicate yet another

strange property of the quantum wave in BQT, we already knew that it is an entity that has no counterpart in other physical theories[158].

The second problem I would like to point out is that the nomological approach crashes with an essential feature of space-time theories. Recall the guidance equation $\frac{\partial S}{\partial t} = -\frac{(\nabla S)^2}{2m} - V + \frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}$, where $Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}$. We can rearrange this equation and apply the gradient operator to obtain $\nabla\left[\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m}\right] = -\nabla(V + Q)$. Now, the left-hand side of this equation is equal to $\frac{d}{dt}\nabla S + \frac{\nabla S}{m} \cdot \nabla \cdot \nabla S$, and this expression can be interpreted as the result of the convective derivative $\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$ applied on $\nabla S = \mathbf{p}$. That is, we thus obtain the total derivative of momentum for a stream of particles associated to the field S and distributed according to $P = R^2 = |\Psi|^2$. Finally, we can wrap up all this to set the equation

$$\frac{d}{dt}\nabla S = \frac{d\mathbf{p}}{dt} = -\nabla(V + Q),$$

which is the Bohmian quantum version of Newton's second law $\frac{d\mathbf{p}}{dt} = -\nabla V = \mathbf{F}$.

Now, let us see what happens, according to this law, to a free quantum particle and assuming the nomological interpretation of the wave-function. According to this view, the term Q cannot be a quantum *potential* associated to a quantum wave, it has to be a nomological term. Therefore, it cannot express a force, for there is no entity to apply it. Since we are considering a free particle, the classical potential V is zero. It is then clear that the *inertial* motion of the particle when V is zero and Q is non-zero is not rectilinear and uniform. That is, the nomological term Q implies that the motion of a free quantum particle is, in general, not rectilinear and uniform. This is already strange, but the Bohmians that support the nomological stance regarding Ψ could simply say 'so what? At the quantum level inertial motion is not uniform and rectilinear'. However, the problem is that this view violates an essential and fundamental principle in the foundations of space-time theories, namely, that free particles follow space-time geodesics in their trajectories. The modern formulation in four-dimensional space-time of Newtonian mechanics, special relativity and general relativity are all theories in which this principle holds. That is, by assuming the nomological stance, Bohmians seem to be in conflict with our best space-time theory (at least in its generally accepted interpretation). Though the conflict with space-time theories is not empirical, the nomological interpretation of the wave-function in BQT seems to tell us that we are wrong in the way we usually understand the meaning of space-time theories.

For all these reasons I think that with the nomological stance we lose more than what we gain, so we better assign a real status to the wave-function in BQT and try to learn to live with this odd entity. The unease with the ontological status of the quantum wave that has been most discussed in the literature is the fact that it is defined in $3N$-configuration space rather than in 3-space. This is usually taken as implying that the wave somehow inhabits in configuration space, and we would be thus obliged to assign a physical status to configuration space. Actually, given the fundamental role of the wave-function in the theory, $3N$-space should be taken as more fundamental than 3-space.

One way to deal with this issue is simply state that 3-space somehow supervenes on $3N$-space. That is, that the three-dimensional space we experience is the result of a certain configuration of the N particles in the universe—the 'universal particle'—in configuration space[159]. An objection that Monton (2002) levels against this view is that the correspondence between configurations in $3N$-space and 3-space is not

[158] Brown, Dewdney & Horton (1995) consider two other variations of the interferometry experiment in order to show also charge and magnetic moment do not obey the LPP.

[159] As Monton (2002) puts is, according to Bohmians that follow this path 'on Bohm's interpretation "the cat is in the box" is true if and only if the universal particle is in the region of $3N$-dimensional space which corresponds to the N particles

one-to-one—to one single arrangement in 3-space, many different configurations in $3N$-space correspond. The reason is that no intrinsic specification of which dimensions in $3N$-space correspond to which particles in 3-space is included in the formalism[160]. A possible reply to this problem is that this shortcoming is just a matter of mathematics. To draw a more substantial ontological implication one may demand for a general proof that a one-to-one correspondence cannot be introduced—actually, Lewis (2004) argues that some physical considerations can ground such a correspondence. Even if the formalism of the theory does not point it out, in an ontological level the one-to-one link may hold.

However, it is clear that there is a more basic difficulty with this approach: a highly dubious space dualism is involved—one may immediately ask how the spaces are connected, how is it that physical 3-space supervenes on a (physical!) configuration $3N$-space. One way to avoid this problem is to reject the space-dualism and simply assume that physical space is given by $3N$-space and that 3-space is nothing but a kind of illusion we are led to. D. Albert (1996) endorses this view. According to this author, the wave-function must be understood as denoting a real object,

> and of course the space those sorts of objects *live* in, and (therefore) the space *we* live in, the space in which any realistic understanding of quantum mechanics is necessarily going to depict the history of the world as *playing itself out* [...] is *configuration*-space. And whatever impression we have to the contrary (whatever impression we have, say, of living in a three-dimensional space, or in a four-dimensional space-time) is somehow flatly illusory. (Albert 1996, 277)

Though Albert proposes an explanation for why the 3-dimensional delusion occurs—he argues that the form of the Hamiltonian operator of QM is such that the natural (but false) hypothesis it suggests is that we inhabit a 3-dimensional space—I think that this standpoint falls into the realm of mere metaphysical speculation. We should, and can, do better than this.

I think that the whole discussion about the ontological status of the wave-function in connection with its description in $3N$-configuration space is misplaced. Consider first the position representation of a quantum state, and assume that the corresponding vector in Hilbert space is superposed in this representation. In any of the interpretations of SQM, we have that this conception of the spatial localization of the system is rather strange. However, that we represent this 'property' by means of a vector in Hilbert space defined on the basis of position operator does not mean that we need to conceive the depicted state as inhabiting Hilbert space rather than physical space. Something similar can be said in the case of the wave-function in BQT. The theoretical description of this entity requires that we represent it in $3N$-configuration space rather than in 3-space. However, this does not mean that we are obliged to think that the wave inhabits $3N$-configuration space. Actually, configuration space is just a mathematical tool that allows us an adequate representation of its properties—thus the analogy with Hilbert space and the representation of quantum states.

That the wave is to be described in configuration space shows that it is not a wave or a field in the usual sense—actually, it manifests that the wave involves non-local properties and effects, for example. We dub and describe it with these terms ('wave', 'field', 'potential', 'phase field', 'amplitude field') by analogy with the corresponding terms in other physical theories, but if we understand these terms in their classical meaning, they do not neatly describe it. The wave-function of BQT represents an entity that does not correspond to any conceptual category in the catalogue of the entities that, according to other

---

arranged in three-dimensional space in such a way that the cat-particles and the box-particles are such that the cat is in the box' (267).

[160] 'As an example, consider Bohm's interpretation. On the mixed ontology, there exist both a three-dimensional space with N point particles and a $3N$-dimensional space with a wave function field and a universal particle. The evolution of the universal particle represents the configuration-space evolution of the N particles. It is not the case, however, that given just the evolution of the universal particle, one could determine the evolution of the N particles [...]. It could be that the x, y and z coordinates of particle number 3 in the three-dimensional space correspond to the seventh, eighth, and ninth coordinates of the universal particle, but it could be that they correspond to different coordinates' (Monton 2002, 268).

physical theories, defines the furniture of the world. To accept it implies to accept that the world contains a highly *sui generis* thing. That the wave-function of BQT is defined in configuration space does not mean that we need to conceive it as 'inhabiting' that space. Such a description tells us, among other things, that we need to conceive it as 'non-locally extended', but this does not imply that the wave does not inhabit 3-space and that we need to conceive configuration space as a *physical* space.

Norsen (2010) offers a historical outlook of the reasons for the unease that some of the founding fathers of the theory felt about the wave-function. When the physicists that manifested an early sympathy for the pilot-wave approach as introduced by de Broglie noticed that Ψ could not be defined in 3-space—in a way such that for a system of N particles a number N of particle-wave pairs would suffice for a dynamical account—the sympathy for such an approach, and even for Schrödinger's interpretation of the meaning of his own equation started to vanish. The important point is that when one reads the opinions of those scientists what one finds is a dissatisfaction grounded on the strange nature of the pilot-wave, but not on the fact that $3N$-space needs to be conceived as physical—actually, the description of the wave in configuration space was taken as manifesting its odd ontological state compared to the waves of classical theories. As time went by, it became clearer and clearer that the description of the wave-function in configuration space was an expression of the non-local features associated to entangled systems. Consider for example, the following passage by Bell and Norsen's comment on it. Bell remarks that thw wave must be defined not in 3-space, but

> in a much bigger space, of $3N$ dimensions. It makes no sense to ask for the amplitude or phase or whatever of the wavefunction at a point in ordinary space. It has neither amplitude nor phase nor anything else until a multitude of points in ordinary three-space are specified. (Bell 2004, 204)

> For any theory in which the wave function has beable status, then, it is necessarily a non-local beable. And this provides a convenient alternative way to state what is surprising and unfamiliar about the de Broglie-Bohm pilot-wave theory: in addition to positing local beables (the particles), the theory also posits a genuinely non-local beable (the configuration space wave function which pilots them). (Norsen 2010, 1862)

We can make full sense of these accurate statements without falling in any realism about configuration space. Consider, for example, the sketchy description of the account of BQT for EPR correlations mentioned above. The non-local effects tell us that we need to think about the wave as existing 'here' and also in a space-like separated 'there', and both 'here' and 'there' are regions in 3-space, of course. If we want to provide a dynamical explanation of how the wave determines the physics of this example, we need to use the concepts of a phase field, an amplitude field, a quantum potential, etc., and these notions can only be defined in the 6-dimensional configuration space of the entangled particles—this is what Bell tells us. Now, that the mentioned terms are described in $3N$-space indicates that the wave of BQT is a non-local beable—and this is what Norsen remarks. In neither case we need to assume that the wave literally inhabits in a physically real configuration space. Expressions like 'the Bohmian wave propagates in configuration space' do not need to be taken literally. We can simply understand this statement as saying that the description of the dynamical evolution of this strange entity is described by terms which are necessarily defined in configuration space. We may never forget that the 'wave' in BQT is not a wave in the usual sense[161].

[161] Norsen (2010) has constructed a toy theory that he calls a *theory of exclusively local beables* (TELB). The ontology of such a theory is given by the quantum particles; a pilot-wave, defined in 3-space, corresponding to each particle; and an infinite set of what he calls 'entanglement fields', also defined in 3-space—as expected, the pilot-waves as local beables are not enough to predict the non-local features connected to entanglement, thus the need for the introduction of an infinite amount of entanglement fields. As Norsen himself points out, a TELB capable of reproducing all the predictions of BQT and SQM must pay the cost of a highly implausible ontology and a very intricate mathematical structure. However, the fact that Norsen's toy TELB, in which no entities are defined in $3N$-space, constitutes an alternative presentation of BQT

Summarizing, we have that, attractive as it seems at first sight, the nomological interpretation of the wave-function in BQT is highly problematic. First, it is rather dubious from an explanatory point of view. A real quantum 'wave' naturally explains the peculiar experimental outcomes in the quantum world. Brown, Dewdney & Horton's argument that in BQT the *localized particle properties* thesis is highly questionable, if not plainly false, can be certainly taken as an indication that the wave-function must be understood as representing something real. Furthermore, a purely nomological understanding of the wave-function entails that BQT is in open conflict with the principle of space-time theories that state that inertial motion is essentially linked to space-time geodesics. Thus, the price to be paid is too high, so a realistic interpretation of the wave-function is advised. That this stance is less problematic is reinforced by the fact that the problem of $3N$-space realism that is usually thought to follow is specious. That the wave-function is theoretically described and defined in configuration space does not mean that we have to think of it as literally inhabiting a $3N$-space—it only indicates that we are dealing with a very peculiar 'non-locally extended' entity. However, the issue of the highly strange ontological status of the wave-function still stands. We can conceive it as existing in 3-space, but it is still an entity without any counterpart in other physical theories. It certainly violates Newton's third law, for example, and this is taken by some as a reason to reject its reality. I think that such a rejection follows if one assumes an ontological generalization of the third law and elevate it to a sort of *a priori* ontological principle, but maneuvers like this are rather questionable and have been refuted in the history of science—just ask Kant. Anyways, this analysis shows that if adopting Bohm's theory is the choice to be taken, it must be openly acknowledged that a puzzling and questionable ontology is assumed.

3.6.8 BQT, SQM, and SR

Now I will deal with a very important issue concerning the evaluative comparison we are considering: the question of whether quantum theory is compatible with SR or not. A common opinion is that the explicit non-locality in BQT precludes from the outset that this theory can be made compatible with SR. As we saw above, in a many-particle system, the form of the quantum potential (which is defined in configuration space) implies that the position of each of the particles at a certain instant depends on the position of all the other particles at that same instant, even if they are all space-like separated. This clearly requires the introduction of a suitable foliation of space-time that allows us to make sense of the clause 'at that same instant', and this seems to be very close to the introduction of a preferred reference frame in which such a foliation determines a hyper-plane of simultaneity. That is, BQT requires the postulation of some space-time structure that is not considered by SR, and this structure seems to go against the spirit of Einstein's theory. On the other hand, and based on the work by Jarrett (1984) and Shimony (1984), there is a widespread opinion that since in SQM the possibility of faster than light signaling is essentially precluded, this theory can 'peacefully coexist' with SR.

---

reinforces my point: there is no need to understand the wave function in BQT as inhabiting in a physically real configuration space. Though in his very interesting paper Norsen assumes that the representation of the wave-function in configuration space implies the problematic $3N$-space realism—thus the motivation to build his TELB—there is a passage in which he makes a point that comes close to what I have just argued: even in the $3N$-representation of the wave-fucntion there is no reason to assume a $3N$-space realism. He ponders on what would have happened if the pilot-wave theory had been originally formulated in 1927 in his TELB version: 'From this perspective—and to whatever extent one finds this a plausible thing to imagine—the proposed theory sheds an interesting new light on ordinary quantum theory and in particular the status of the wave function therein. For one could imagine, after the present theory had been proposed and tested, mathematical explorations of its structure and predictions revealing the possibility of a mathematically-equivalent formulation in terms of a single abstract pilot-wave on configuration space. *From this perspective, the (in our world, familiar) configuration space wave function would be merely a convenient mathematical device, analogous to Hamilton's principal function in classical mechanics—an abstract mathematical quantity which perhaps in some situations makes calculations simpler or more elegant, but which one needn't take as indicating the real existence of anything like a physically real field on configuration space*' (Norsen 2010, 1873, my italics).

Let us begin with an analysis of this second standpoint. In order to appraise Jarret's analysis and its subsequent interpretation, we may first take a look once again at Bell's theorem in a different and more general version. In section 3.4.4 I introduced the theorem as showing that non-locality is a feature that empirically adequate HVT must exhibit. Now we will look at it as showing that the assumption of locality, in and by itself, leads to observance of the Bell inequalities, so that non-locality is a feature that *any* empirically viable theory must exhibit.

Consider the setup in which the spin direction of a two-particle system in the singlet state is measured in two space-like separated regions. Let us assume that Alice and Bob perform the corresponding measurements to obtain the outcomes A and B, respectively. Let us also assume that $\hat{a}$ and $\hat{b}$ are the specific setups (orientations) of the S-G devices, when the measurement is performed, that Alice and Bob operate, respectively, and that the state of the system is described by λ—notice that no assumption about hidden variables is made. Since we are dealing with correlated measurement results, the correlations imply that the probabilities of finding a certain value for A and for finding a certain value for B are not independent, that is, $P(A,B) \neq P(A) \times P(B)$. However, when we consider that the correlations are stipulated by the state as described by λ, we obtain $P(A,B|\lambda) = P(A|\lambda) \times P(B|\lambda)$. In other words, when we consider the ground of the correlations, the probabilities get screened-off from each other.

Now, the locality condition requires that the causes of any event must lie within the past light-cone of the event, which in turn implies that the outcome A must be independent of the outcome B and of the setup of the S-G device that measures B, so that the value of the outcome A depends only on λ and on the setup $\hat{a}$ of Alice's S-G device. That is, locality implies that it must hold that $P(A|\hat{a},\hat{b},B,\lambda) = P(A|\hat{a},\lambda)$— and by the same token, $P(B|\hat{a},\hat{b},A,\lambda) = P(B|\hat{b},\lambda)$. Finally, and given the screening-off condition, we have that

$$P(A,B|\hat{a},\hat{b},\lambda) = P(A|\hat{a},\lambda) \times P(B|\hat{b},\lambda)$$

This last expression is commonly known as the *factorizability condition*, which results from the assumption of locality. For any theory that predicts correlated measurements, if we want such a theory to offer an account of those correlations such that local causality is respected, then the theory must observe the factorizability condition

Now, we have that the expectation values for the outcomes for Alice and Bob are given by $E(\hat{a},\lambda) = \sum P(A|\hat{a},\lambda)A$ and by $E(\hat{b},\lambda) = \sum P(B|\hat{b},\lambda)B$ (just as the probabilities for a specific outcome, the expectation values for Alice and Bob are screened-off each other), respectively, so that $E(\hat{a},\hat{b},\lambda) = \sum P(A,B|\hat{a},\hat{b},\lambda)AB$. Now, assuming that we assign the value $+1$ to 'spin-up' and -1 to 'spin-down', it follows that $\left|E(\hat{a},\hat{b}|\lambda) - E(\hat{a},\hat{b}'|\lambda)\right| + \left|E(\hat{a}',\hat{b}|\lambda) + E(\hat{a}',\hat{b}'|\lambda)\right| \leq 2$, where $\hat{a}',\hat{b}'$ simply represent a different orientation for the S-G devices that Alice and Bob operate[162]. However, quantum theory tells us that for the singlet state the expectation value is given by $E(\hat{a},\hat{b}) = -\cos\theta$, where θ stands for the angle subtended by the directions $\hat{a}$ and $\hat{b}$, and it turns out that for certain angles θ subtended by suitable choices of $\hat{a},\hat{a}',\hat{b},\hat{b}'$, the inequality that follows from the factorizability assumption is violated—and as I mentioned

[162] Actually, this inequality does not require us to assume a particular λ. We assume a normalization condition for the the probability density ρ: $\int_\lambda \rho(\lambda)d\lambda = 1$. In this case, the expectation value is given by $E(A,B|\hat{a},\hat{b}) = \int_\lambda d\lambda P(A,B|\hat{a},\hat{b})\rho(\lambda)$, so that the inequality becomes independent of a particular λ-choice and given by $\left|E_\rho(\hat{a},\hat{b}) - E_\rho(\hat{a},\hat{b}')\right| + \left|E_\rho(\hat{a}',\hat{b}) - E_\rho(\hat{a}',\hat{b}')\right| \leq 2$. This independence allows that the inequality violation tests do not depend on the production of perfect (anti)correlations (singlet state) in the laboratory. This generalized version of Bell's theorem is due to Clauser, Horne, Shimony & Holt (1969).

above, Aspect's experiments are widely taken as an experimental confirmation of the inequality violation[163].

It is important to underscore that in this formulation of Bell's famous argument, no essential reference to hidden variables is involved. In section 3.4.4 we explicitly referred to a hidden variable that was responsible for the correlations which operated as a local common cause. The intention there was to clearly state that Bell's theorem implies a constraint for any empirically viable HVT: it has to be nonlocal. Although there is a widely spread opinion (sometimes implicitly assumed) that this is *the* meaning of the theorem, this alternative formulation does not assume in any extent that we are dealing with a HVT. Therefore, the general conclusion that can be extracted from the theorem is that any empirically viable quantum theory must be nonlocal. It is true that the most natural way to instantiate local causality in quantum theory is by introducing some hidden variable (that can act as a local common cause to explain the singlet state correlations, for example), but the main point of the theorem is that it shows that no theory that respects the factorizability condition and thus the local causality principle can be empirically adequate. Actually, the nonlocality in SQM is partially expressed in that the theory does not postulate any hidden variable to locally account for the singlet state correlations—this is why the theory manages to violate facorizability. Bell's theorem does not only tell us something important about HVT, but about quantum theory.

We are now in position to assess Jarrett's and Shimony's arguments. The key is given by a logical decomposition of the factorizability condition $P(A,B|\hat{a},\hat{b},\lambda) = P(A|\hat{a},\lambda) \times P(B|\hat{b},\lambda)$. First, Jarrett introduces his 'locality' sub-condition $P(A|\hat{a},\hat{b},\lambda) = P(A|\hat{a},\lambda)$, which means that the probability to obtain a certain value for the outcome A depends only on $\hat{a}$ and λ, and not on the space-like distant experimental setup $\hat{b}$—the same holds, mutatis mutandis, for the probability of a certain value for outcome B. Jarrett's second sub-condition, which he calls 'completeness', consists in that $P(A|\hat{a},\hat{b},B,\lambda) = P(A|\hat{a},\hat{b},\lambda)$, meaning that the specific value for outcome A is stochastically independent from the specific value of outcome B—the same holds, mutatis mutandis, for the value of outcome B with respect to the specific value of outcome A. It is easy to see that, taken together, these two sub-conditions entail the factorizability condition, so that the violation of either of them assures that the factorizability condition is violated. In SQM, only the second of Jarrett's sub-conditions is violated, for recall that in an entangled system constituted by two subsystems, all of the statistical information regarding local measurements performed on one subsystem are contained in the reduced density matrix of the measured subsystem, which traces-off the statistics of the other subsystem—thus, the first subcondition, Jarrett's 'locality', is respected. On the other hand, the correlations expressed by the singlet state tell us that given a 'spin-up' *outcome* in Bob's S-G device, we know that the value of the outcome that Alice obtains is 'spin-down', so that the second of Jarrett's subconditions, 'completeness', is violated—and this violation alone is responsible for SQM's violation of the factorizability condition and of the Bell inequality.

Jarrett takes it that compliance to the 'locality' condition suffices for SQM to be compatible with SR. He states that since the probabilities of a specific value for outcome $A(B)$ is independent of the experimental setup $\hat{b}(\hat{a})$, even though Bob(Alice) can control the orientation setup of his(her) S-G device, he (she) cannot take advantage of this possibility to send a faster than light signal to Alice(Bob), for no matter how the manipulation is done, the statistics in the space-like separated regions where Alice(Bob) performs her(his) measurement remain unaltered. That is, Jarrett assumes that the main restriction of SR

[163] This reasoning assumes that λ is independent of $\hat{a}$ and $\hat{b}$, which is of course a very plausible assumption. If this requirement is not observed, Bell's theorem can be circumvented—the singlet correlations could be accounted for without violating factorization. The two ways to ground that λ-independence from $\hat{a}$ and $\hat{b}$ is broken are (i) to assume a sort of natural conspiracy: to postulate a common cause that λ and $\hat{a}$ and $\hat{b}$ share, such that it correlates some types of λs with some types of $\hat{a}$s and $\hat{b}$s; and (ii) to postulate a kind of *backwards* causation going from $\hat{a}$ and $\hat{b}$ to λ (this is assumed in the so-called *transactional* interpretation). Though conceptually interesting, both alternatives are rather bizarre in terms of physical plausibility. Anyhow, they are not relevant here.

poses for a quantum theory is that his conception of locality, as expressed by the first sub-condition, is respected—and this means in turn that superluminal signaling must be forbidden. He states that what locality excludes,

> is the possibility that the preparation of either measuring device in some particular state can exert a causal influence on the other subsystem so as to affect the probabilities for the possible outcomes of measurements performed on that other subsystem. Since these two events, the preparation of one measuring device in a given state and the measurement executed by the other measuring device, can be space-like related, locality is a requirement of relativity theory.
> In order to establish the relativistic basis for locality, it suffices to show that if enough control over the state preparations is assumed, violations of the locality condition provide (at least in principle) the means for superluminal signal transmission. (Jarrett 1984, 573)

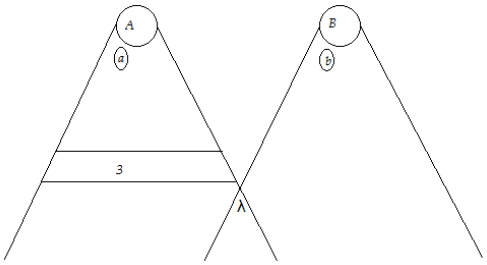
On the other hand, Jarrett argues that the violation of the second sub-condition does not amount to an incompatibility with SR, for it only indicates that there is something missing in the state description λ, a missing element that, were it present, would explain the correlations observed in EPR-type experiments in such a way that (assuming that the first sub-condition is respected) no non-local features are involved:

> Completeness asserts the stochastic independence of the two outcomes in each pair of spin measurements. This can be interpreted as a natural requirement for theories which represent observable phenomena as the effects of interacting (but otherwise independently existing) entities whose physical state may be exhaustively characterized by the specification of some (not necessarily unique) set of definite, well-defined properties.
> For such theories, given the state of the two-particle system and the states of both measuring devices, the probability for a given outcome of the spin measurement at L(R) is independent of the outcome of the other spin measurement at R(L); that is to say, the probability for the given outcome at L(R) is invariant under conditionalization on the outcome of the measurement at R(L). However, unlike local theories, the probability for that outcome of the L(R) measurement may very well exhibit a dependence on the state of the distant measuring device R(L). *By including in the state descriptions of the two-particle system and the measuring devices all those properties of the systems (i.e. precise numerical values of whatever physical quantities are appropriate for the types of entities posited by the theory) in virtue of which measurement interactions yield each possible outcome with its assigned probability, such state descriptions automatically "screen off" any correlation of the pairs of measurement outcomes which might arise from the omission from the state descriptions of predictively relevant information*. (ibid, 578-9, my emphasis)

Summarizing, according to Jarrett, a violation of his 'locality' condition would amount to a violation of the basic constraint imposed by SR, which is expressed in the prohibition of faster-than-light signaling;, but a violation of his 'completeness' condition (assuming that locality is respected) only shows that the state description does not contain the information that accounts for the correlations observed in spin space-like distant measurements of the singlet state. Now, since SQM violates the factorizability condition only in terms of 'completeness' violation, there is no conflict between this theory and SR.

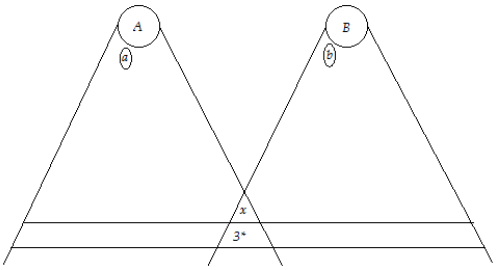
Travis Norsen (2009) argues that Jarrett's line of thought is misleading. He claims that Jarrett overlooks a basic element in Bell's formulation of the factorizability condition. A careful analysis of Bell's own formulation of $P(A, B|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda) \times P(B|\hat{b}, \lambda)$ indicates that it *directly* follows from the assumption of locality (it is not the consequence of locality *and* something else), and the fact that the factorizability condition is a direct expression of the locality assumption implies that Jarrett's second condition cannot be interpreted as a 'completeness' condition whose violation is enough to break factorizability, even if Jarrett's 'locality' is respected. In other words, even though the formal decomposition of the factorizability condition that Jarrett introduces is logically correct, his interpretation of the physical meaning of the two sub-conditions is inadequate. To see why, Norsen argues, we only need to take a look at Bell's own formulation of the 'principle of local-causality' that underlies his theorem:

> The direct causes (and effects) of events are near by, and even the indirect causes (and effects) are no further away than permitted by the velocity of light [...]. Thus, for events in a space-time region 1 [Alice's] [...] we would look for causes in the backward light cone, and for effects in the future light cone. In a region like 2 [Bob's], space-like separated from 1, we would seek neither causes nor effects of events in 1. Of course this does not mean that events in 1 and 2 might not be correlated. [...]
> A theory will be said to be locally causal if the probabilities attached to values of local beables in a space-time region 1 are unaltered by specification of values in a space-like separated region, *when what happens in the backward light cone of 1 is already sufficiently specified, for example by a full specification of local beables in a space-time region 3*. [...]
> It is important that region 3 completely shields off from 1 the overlap of the backward light cones of 1 and 2. And it is important that events in 3 be specified completely. Otherwise the traces in region 2 of causes of events in 1 could well supplement whatever else was being used for calculating probabilities about 1. The hypothesis is that any such information about 2 becomes redundant when 3 is specified completely. (Bell 1990, 239-40, my emphasis)



The figure illustrates Bell's important remark. Bell's region 1 is our region A, whereas his region 2 is our B. The point is that in the formulation of the locality that leads to the factorizability condition, Bell says, it is required that the value obtained in A is sufficiently specified by the beables in region 3, which is space-like isolated from the past light-cone of region B. More formally, Bells is requiring that $P(\mathscr{b}_1|\mathcal{B}_3, \mathscr{b}_2) = P(\mathscr{b}_1|\mathcal{B}_3)$, where $\mathscr{b}_i$ means 'some beable in region i' and $\mathcal{B}_i$ means 'a full specification of beables in region i'[164]. Bell's rationale for this requirement is that without it, the breakdown of the factorization condition would not guarantee that some way of nonlocal causation is present. Assume that we are dealing with a non-deterministic theory and that some event, a beable that we can call x, comes into existence in the future of region 3*, a region that includes a sub-region in which the past light-cones of A and B overlap. If x lies in the future of 3* and also in the overlapping region of the past light-cones of A and B, and if x somehow influences events both on A and B, 'there is therefore the possibility that specification of events from $[B]$ could allow one to infer something about $[x]$ from which one could in turn infer more about goings-on in $[A]$ than one could have inferred originally from just the full specification of beables in 3*' (Norsen 2009, 278). On the other hand, if $\mathcal{B}_3$ did not provide a completely specified description of the beables in region 3, that is, if we had an incomplete $\overline{\mathcal{B}}_3$, then it could be the case that the beables that 'carry' the causal influence from x to A are omitted by $\overline{\mathcal{B}}_3$, and the violation a corresponding weaker condition $P(\mathscr{b}_1|\overline{\mathcal{B}}_3, \mathscr{b}_2) = P(\mathscr{b}_1|\overline{\mathcal{B}}_3)$ would not signify the presence of a some sort of nonlocal causation.

[164] The talk of 'the beables of the theory' does not imply that assumptions about hidden variables are involved, and do not restrict the meaning of the theorem commented above. As Norsen points out, 'note that everything in the above discussion refers to some particular candidate physical theory. For example, there is a tendency for misplaced skepticism to arise from Bell's use of the concept of "beables" in the formulation of local causality. this term strikes the ears of those influenced by orthodox quantum philosophy as having a metaphysical character and/or possibly committing one (already, in the very definition of what it means for a theory to respect relativistic local causality) to something unorthodox like "realism" or "hidden variables". Such concerns, however, are based on the failure to appreciate that the concept "beable" is theory-relative' (2009, 279). Once again, in Bell's derivation of factorizability, no specific assumption about the role of hidden variables is made. No matter how a theory instantiates locality, be it through hidden variables or in some other way, factorizability and Bell's inequality will follow.



Thus, According to Norsen, the disagreement between Bell's and Jarrett's views on the factorizability condition relies in that whereas Bell puts forward a requirement $P(\mathscr{b}_1|\mathcal{B}_3, \mathscr{b}_2) = P(\mathscr{b}_1|\mathcal{B}_3)$, Jarrett only assumes in his decomposition of the factorizability condition that $P(\mathscr{b}_1|\overline{\mathcal{B}}_3, \mathscr{b}_2) = P(\mathscr{b}_1|\overline{\mathcal{B}}_3)$, which allows him to interpret $P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \hat{b}, \lambda)$ as a 'completeness' condition.

Now, Jarrett's argument could be taken as a correction of Bell's analysis in the sense that it shows that it is not locality in and by itself, but the conjunction of 'locality', $P(A|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda)$, and 'completeness', $P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \hat{b}, \lambda)$, what leads to factorizability. It is then clear that Jarrett's physical interpretation of his two sub-conditions implies that any quantum theory that violates $P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \hat{b}, \lambda)$ but respects $P(A|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda)$ is capable to predict the correct correlations for a singlet state without any kind of nonlocality being involved (for it is its incompleteness what allows it to violate the Bell inequality), and any such theory would be compatible with SR (for 'locality' is the constraint posed by SR).

However, Norsen also shows that a simple toy-model proposed by Tim Maudlin (2011) respects $P(A|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda)$, violates $P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \hat{b}, \lambda)$, but nevertheless features a form of nonlocality that puts it in conflict with SR. Consider two entangled particles in the singlet state. Our toy-model states that the spin of the particles in indeterminate unless a measurement is performed. Assume that Bob's particle arrives first to the S-G device. At this point the particle 'flips a coin' in order to decide whether it gets deflected 'up' or 'down'—the model is in this sense essentially probabilistic. The particle then sends a massless tachyon to Alice's particle that somehow tells it how to behave: in such a way that the correlations predicted by the expectation value $E(\hat{a}, \hat{b}) = -\cos\theta$ get respected. In this model, Jarrett's 'locality' sub-condition, $P(A|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda)$, is respected, but the 'completeness' sub-condition, $P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \hat{b}, \lambda)$, is violated. However, it is obvious that there is a form of causal nonlocality involved—the action of the tachyon—and that the theory is not necessarily incomplete, for it may include a full description of the tachyonic process. This analysis indicates that Jarrett's physical interpretation of the two sub-conditions of factorizability is not adequate. Factorizability does not follow from the conjunction of 'locality' and 'completeness'. Rather, as Bell points out, it is a direct consequence of locality.

Shimony (1984) also draws on the logical decomposition of factorizability that Jarrett introduced, but he interprets differently the physical meaning of the two sub-conditions. Concerning the first sub-condition, Shimony does not diverge from Jarrett, but instead of dubbing it 'locality', Shimony uses the expression 'parameter-independence' to refer to $P(A|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda)$. Both authors agree in that this condition means that the probability of obtaining a certain value for outcome A does not depend on the particular setting $\hat{b}$ of a space-like distant measuring device. The interpretative difference is given with respect to the second sub-condition $P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \hat{b}, \lambda)$. Shimony calls it 'outcome-independence' and it simply asserts that the probability to obtain a certain value for A does not depend on the specific value of outcome B that is obtained in a space-like distant region. The point of divergence consist in that Shimony does not interpret this condition as a requirement for 'completeness'. Actually, he agrees with Bell in that factorizability is a direct consequence of the locality condition, so that the violation of any of the two sub-conditions yields a violation of locality—which in turn implies that Shimony understands Bell's theorem in its full significance, namely, that any empirically viable theory must be nonlocal.

However, the idea is now that the *type* of nonlocality that results from the violation of each sub-condition is different. Shimony agrees with Jarrett in that the essential constraint that SR poses on quantum theory is that superluminal signaling is forbidden. Now, the fact that in SQM the statistical information about local measurements in A is exhausted by the reduced density matrix that traces-out the statistical information about B entails that in SQM 'outcome-independence' violation is responsible for the nonlocality of the theory, whilst 'parameter-independence' is respected, which in turn guarantees that the possibility of superluminal signaling is forbidden. As mentioned above, Bob cannot manipulate the setup $\hat{b}$ of his S-G device in order to send a faster-than-light signal to Alice, for such a manipulation would not yield any disturbance on the expected value of A. It is true that given a result $+1$ in B, the probability of -1 in A becomes one, but since we cannot go further than Born's probabilistic rule in the predictions of SQM, that 'outcome-independence' is violated does not mean that we can send superluminal signals either. Therefore, according to Shimony, notwithstanding the locality violation in terms of 'outcome independence' breaking, the fact that in SQM faster-than-light signals are forbidden allows a peaceful coexistence between this theory and SR.

Though Shimony's interpretation of the physical meaning of the two sub-conditions is certainly better than Jarrett's, the conclusions he draws from such an analysis are rather dubious. Let us first take a look at the connection between 'parameter-independence' and superluminal signaling in the context of Bohm's theory. We can easily envisage a method for Bob to send faster than-light-signals to Alice. We know that Bob's measurement implies a disturbance in the partial quantum wave associated to his particle, and that this disturbance implies an instantaneous non-local disturbance on the quantum potential affecting Alice's particle. Besides, the theory is deterministic at the fundamental level, thus, if Bob knows the quantum state $\Psi(x, y)$ corresponding to the ensemble of entangled particles *and* the initial conditions (x, y) for each of the systems in the ensemble corresponding to Ψ, then Bob could control the measurement outcomes that Alice obtains by affecting the partial wave corresponding to his own particle in a suitable way. For example, Bob and Alice could determine that she will measure her particle always in the x spin direction, so that Bob can alter at will her measurement outcomes by measuring his own particle along different spin directions. The implementation thus requires a decision regarding what kind of measurement Alice will perform, a code based on the measurement outcomes that Alice will obtain, and that Bob knows both Ψ and the initial conditions (x, y) of each system in the ensemble associated to Ψ. But, as we know from the dynamics of measurement interactions, our knowledge of the initial positions of the systems in the ensemble is merely statistical and given by the distribution postulate $P = |\Psi|^2$, and if Bob tried to experimentally find such initial conditions, he would disturb the wave function Ψ. That is, it is impossible to know, at the same time and for an individual system, in the ensemble, both Ψ and (x, y), so that the method of superluminal signaling just described cannot be instantiated. Though the results of Alice's measurements certainly depend on the experimental setup that Bob applies, Bob cannot *control* her Alice's measurement outcomes.

Now, does BQT violate 'parameter independence'? In an ontological sense, it certainly does. Each outcome that Alice obtains is deterministically determined by the initial positions of the particles in the system, and by *how* Bob's particle is disturbed if it is measured before Alice performs her measurement[165]. Thus, the *relative frequency* of the outcome values that Alice obtains is indeed dependent on Bob's spacelike separated experimental setup. However, if we understand that the probabilities for Alice measurements express nothing but our ignorance with respect to the initial positions of each specific system in the ensemble, then, since there is no way to gain further knowledge that allows Bob to control Alice's

[165] Notice that the relevant nonlocal feature that determines Alice's result is how the global *quantum potential* associated to a specific entangled system is altered by Bob's measurement, so that the particular *outcome* of Bob's measurement is redundant concerning the value that Alice measurements will yield. In other words, Alice's outcome depends on the initial position of the system in configuration space, and on how Bob's measurement disturbs Alice's branch of the global quantum potential, but not on Bob's measurement outcome. In this sense BQT respects 'outcome independence'.

outcomes, we could say that 'parameter independence' is observed in an epistemic sense. In a word, we may say that at an ontological level, BQT violates 'parameter independence', but at an epistemological level it is respected. Anyhow, the sense in which BQT contravenes Shimony's first sub-condition for factorizability is substantial enough to state that although Bohm's theory violates 'parameter independence', faster-than-light signaling is not allowed by the theory[166].

Thus, considering that there is a friction between BQT and SR, it follows that the impossibility of superluminal signaling is not a sufficient condition for compatibility with SR. Moreover, as Maudlin has shown (2010, chapter 4), SR does admit some kinds of superluminal signaling. Troubles with faster-than-light signals come up when they imply casual paradoxes (this happens when the signals are such that it is possible to send a message to an event that lies in the past light-cone of the emission source), and when the signals yield a special or privileged foliation of space-time. That is, the real sense in which the impossibility of (some type of) faster-than-light signals is a constraint imposed by SR is that such signals presuppose that some space-time structure is postulated, structure that is at odds with the space-time structure that SR, in Minkowski's formulation, posits. This clearly explains why BQT is in conflict with SR even though no superluminal signaling is possible according to the former, and also why some forms of faster-than-light signals are tolerated by SR—the ones that do not require to postulate the problematic structure. Thus, *pace* Shimony, we have that *(i)* 'parameter independence' observance is not enough to ensure that superluminal signaling is forbidden, and *(ii)* prohibition of superluminal signaling is neither a sufficient nor a necessary condition for compatibility with SR.

We can now turn to an examination of another view that is assumed by Shimony: that SQM violates 'outcome independence' is harmless for its compatibility with SR. Once again, since he takes it that observance of 'parameter independence' is an expression that superluminal signaling is forbidden, and that this is the essential constraint that SR imposes, Shimony states that the fact that the probabilities of the possible outcomes that Alice can obtain certainly depend on which specific values Bob obtains in his measurements does not constitute a source of conflict between SQM and SR—violation of 'outcome independence' allows SQM to violate Bell's inequalities without implying an open conflict with SR. This view has been also highly influential, but Maudlin (2010) compellingly argues against it. The main reason why such a view is wrong is that outcome dependence certainly implies a causal correlation between space-like separated events, correlation which in turn yields a foliation of space-time that is not purported by the special-relativistic description of the chrono-geometry of the physical world.

Maudlin states that despite the fact that a sound and adequate definition of causality is a complicated issue, a minimum and non-controversial criterion for causal correlation between events can be formulated. We can safely assume the following sufficient condition for a causal connection: 'given two events A and B, if B would not have occurred had A not occurred (or if B would have been different had A been different) then A and B are *causally implicated with each other*' (2010, 116). This is a minimum criterion for causal *connection*, but it must not be taken as a method to determine what is the specific causal connection between the events involved: 'we do not suppose that it follows from the fact that A is causally implicated with B that A caused B or B caused A' (ibid, 117)—causal implication or correlation is a weaker notion than causation. The next step is thus to apply this criterion to causal implication between space-like separated events:

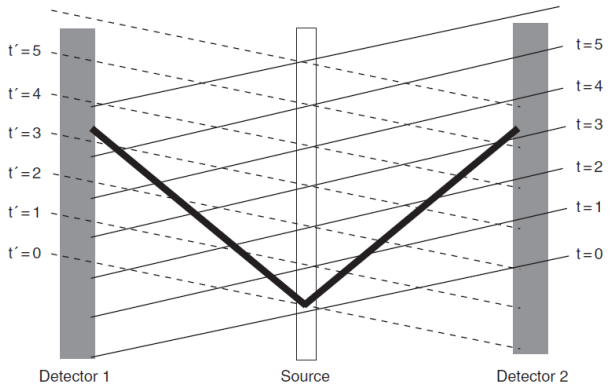
> if no causal influences are superluminal then it *cannot* be the case for space-like separated A and B that A would not have occurred had B not occurred *and everything in A's past light cone been the same* [...]. Taking the contrapositive of [this] conditional yields our sufficient condition: given a pair of space-like separated events A and B, if A would have not occurred had B not occurred even though everything in A's past light cone was the same then there must be superluminal influences. (ibid, 118)

[166] Furthermore, since the theory respects 'outcome independence', then it *must* violate 'parameter independence'.

Bell's theorem tells us that a violation of the inequality—which is demanded by empirical adequacy—requires a violation of the factorizability condition, and, *a fortiori*, a violation of the principle of local causality. Furthermore, it is easy to see that 'outcome independence' violation is an instantiation of Maudlin's criterion for causal correlation between space-like separated events. We know we cannot blame a common cause in the intersection of the past light cones of *A* and *B* for the correlations between the spin measurement outcomes that Alice and Bob obtain, so a violation of 'outcome independence' clearly means that if *B* had been different then *A* would be different even if everything in *A*'s past light cone is kept the same: 'satisfying this condition does not guarantee that *B* is a cause of *A*. As in all cases of causal implication, *A* might cause *B*, or *B* might cause *A*, or there might be a common cause. But since the common cause, if any, must lie outside of *A*'s past light cone, all three possibilities involve superluminal influences' (Maudlin 2010, 119).

Thus, the violation of 'outcome independence' implies that non-local causal connections are featured in SQM. We have already seen that the prohibition of superluminal signaling does not adequately represent the constraint that Einstein's theory imposes (unless one assumes a dubious instrumentalist or operationalist interpretation). The prohibition of non-local causality does better in this sense, though not in and by itself. A violation of such a prohibition is a manifestation that space-time structure not contemplated by SR, and which seems to be at odds with the relativity principle, must be assumed. If non-local causal influences are allowed by a quantum theory, then a preferred foliation of space-time—a hyperplane that joins the space-like separated events which are casually connected—needs to be assumed; and this preferred foliation may pick a preferred reference frame: the frame in which the mentioned hyperplane is a simultaneity slice, so that the ghost of an absolute time comes to spook us. We can thus conclude that Jarrett's and Shimony's analyses of the logical decomposition of Bell's factorizability condition do not allow to conclude that the non-locality in SQM is such that this theory can peacefully coexist with SR. Prohibition of superluminal signaling does not express an essential constraint of Einstein's theory, observance of 'parameter independence' is neither a sufficient nor a necessary condition for compatibility with SR, and violation of 'outcome independence' contemplates non-local causal connections which indicate that space-time structure which goes against the spirit of SR must be postulated.

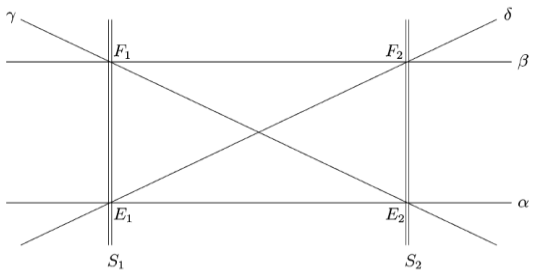
This last point is particularly clear if we assume a projection postulate. If Bob's measurement collapses the wave function of the total system in the singlet state into a state in which his particle has a definite spin, then we naturally conclude that such a collapse instantaneously and simultaneously produces that Alice's particle spin becomes definite as well. The problem is that if we take relativity into the picture, we should then ask simultaneously *in what frame?* That is, we may inquire what is the simultaneity hyperplane that joins the space-like separated events involved in the collapse. We could try to answer this question trying to respect the democratic spirit of SR (the relativity principle) and say that every inertial frame is free to use its own set of simultaneity slices. To illustrate what would happen if we take this path we can consider the following figure (taken from Maudlin 2010, 188):



Assume that Bob's and Alice's detectors are detectors 1 and 2, respectively. If we consider the simultaneity hyperplanes of the unprimed frame to describe the collapse, we have that Bob's particle is detected first, at $t = 3.2$, and before that t neither of the particles has a definite spin, but after it, both particles are spin-definite (in the direction measured by Bob). When Alice performs her measurement, at $t = 5.5$, the particle's spin is already determined, of course, and we know what the outcome will be with probability one. But if we consider the primed reference frame, the story is symmetric, but completely different: Alice's particle is measured first at $t' = 3.2$ and Bob's is detected at $t' = 5.5$. But this situation is rather perplexing. Consider the event in Alice's particle world-line that occurs at $t = 4$. If we take the unprimed reference frame, we have that to such an event a definite spin state corresponds, for in this frame the collapse already happened. However, if we describe the situation according to the primed frame, we have that Alice's particle does not have a definite spin state, for the collapse has not happened yet—the event we are considered occurs at $t' = 2.3$. In SR we are used to deal with relative times and lengths, but that a property like being (or not) in a definite spin state seems to be rather difficult to conceive as frame-relative. We can understand that kinematical properties and processes (and dynamical effects associated to such properties and processes) are frame-relative, for we can make sense of such a relativity by invoking the relativity of simultaneity. However, that a purely dynamical property like a particle having or not having a definite spin cannot be explained in this fashion—actually, here the relativity of simultaneity is the source of perplexity rather than the explanation. I am not affirming that there is some *a priori* reason to reject this form of frame-relativity, but there is certainly no available firm and plausible background (either physical or philosophical) to support it.

The alternative is thus to say that there is one hyperplane that defines the simultaneity between the events involved in the collapse. This would be highly problematic. First, if we assume that the preferred family of hyperplanes that determine the wave-collapse hyperplanes are written in the chrono-geometric structure of space-time, we would be blatantly at odds with SR: it would amount to a rejection of the relativity principle. Second, if we want to avoid this way of determining the hyperplane, we find that there is nothing in SQM that allows us to do it according to dynamical considerations—this is just a consequence of the fact that there is no physical underpinning for the projection postulate. Besides, as Maudlin states (2010, 185-7), neither the physical state of the source of the particles, nor the physical state of the particles and/or the detectors, nor the overall disposition of all the matter in the universe can be effectively invoked to physically define the collapse-hyperplane. Thus, the non-local causal connections involve either very strange form of frame-relativity regarding dynamical properties, or they imply an open conflict with SR.

In the case of the modal interpretation of SQM, we have that even though no wave collapse is assumed, the situation is rather complex as well (see Myrvold 2002; and Dieks 2005). Consider two localized subsystems S_1 and S_2 (say, two entangled electrons in the singlet state), and let α and β be two simultaneity hyperplanes of a reference frame Σ. The systems S_i are localized in α in the events E_i, and in β in the events F_i (we assume that the region where the systems are localized are small enough so they can be treated as points). Finally, γ is a simultaneity hyperplane of a frame Σ', and δ is a simultaneity hyperplane of a frame Σ''—so that E_1 and F_2, and E_2 and F_1 are space-like separated (the figure is taken from Dieks 2005, 410):



If the systems are isolated during their evolution between α and β there are corresponding operators U_i such that the total system $S_1 \oplus S_2$ on β is related to its state on α by $\rho(\beta) = U_1 \otimes U_2 \rho(\beta) U_1^\dagger \otimes U_2^\dagger$. Now, let $\rho(\gamma)$ and $\rho(\delta)$ be the states of the system on γ and δ, respectively. We have that on the assumption of the unitary evolution of the system, we can relate these states to $\rho(\alpha)$ by means of $\rho(\gamma) = U_1 \otimes I_2 \rho(\alpha) U_1^\dagger \otimes I_2$ and $\rho(\delta) = I_1 \otimes U_2 \rho(\alpha) I_1 \otimes U_2^\dagger$. Notice that since we are transforming a state in a hyperplane of simultaneity in frame Σ to states on different hyperplanes (that correspond to simultaneity slices in frames Σ' and Σ''), operators corresponding to the Lorentz transformations are implicitly involved in the last two expressions[167]. That is, we assume Lorentz-invariance as an expression of the chrono-geometric structure of space-time.

Let us now assume that A_1 and A_2 are definite properties possessed by S_1 and S_2, respectively and on α, and that B_1 and B_2 are definite properties of the systems on β[168]. Given the unitary evolution we assumed and the space-like separation between E_1 and F_2 and between E_2 and F_1, we would naturally expect that the value of A_1 possessed by the system S_1 at E_1 is possessed regardless of which hyperplane we are considering in the description of the state, and the same holds for the other values of the properties possessed by the corresponding system at E_2, F_1 and F_2, of course. That is, unlike what is the case with respect to kinematically grounded properties, we may naturally expect that what occurs at these four space-time points regarding the values of the properties A_i and B_i (they could be spin values in a certain direction) does not depend on the way that space-time is foliated in simultaneity hyperplanes.

However, this natural expectation would crash with Bell's theorem. If it were satisfied, then there would be a joint probability distribution over the values of the four observables, such that—assuming, in accordance with SR, that the Born rule for probabilities applies equally in all reference frames, and thus applies equally on α, β, γ and δ—it would yield as marginal probabilities the Born probabilities on all four hyperplanes. But given the relation between $\rho(\gamma)$ and $\rho(\alpha)$, and between $\rho(\delta)$ and $\rho(\alpha)$ stated above, we have that the existence of the joint probability distribution is equivalent to the existence of a joint probability distribution in state $\rho(\alpha)$ that yields as marginal probabilities the statistics for the observables $A_1 \otimes A_2$, $A_1 \otimes C_2$, $C_1 \otimes A_2$, $C_1 \otimes C_2$, where $C_i = U_i^\dagger B_i U_i$. Now, Bell inequalities violation entails that such a joint probability distribution cannot exist[169]. Therefore, 'if $\rho(\alpha)$ is a state such that a Bell inequality is violated for the observables A_1, C_1, A_2, C_2, then it cannot be the case that A_1 at E_1, A_2 at E_2, B_1 at

[167] More precisely, the states of the system according to Σ' and Σ'' are $\rho'(\gamma) = \Lambda(U_1 \otimes I_2)\rho(\alpha)(U_1^\dagger \otimes I_2)\Lambda^\dagger$ and $\rho''(\delta) = \Lambda'(I_1 \otimes U_2)\rho(\alpha)(I_1 \otimes U_2^\dagger)\Lambda'^\dagger$, respectively, where Λ and Λ' are the Lorentz transformations from Σ to Σ', and from Σ to Σ'', respectively. $\rho(\gamma)$ and $\rho(\delta)$ are the states of the system on γ and δ, but as represented in the coordinate basis of frame Σ, and they are related to $\rho'(\gamma)$ and to $\rho''(\delta)$ by $\rho(\gamma) = \Lambda^\dagger \rho'(\gamma)\Lambda = U_1 \otimes I_2 \rho(\alpha) U_1^\dagger \otimes I_2$ and $\rho(\delta) = \Lambda'^\dagger \rho''(\delta)\Lambda' = I_1 \otimes U_2 \rho(\alpha) I_1 \otimes U_2^\dagger$, respectively (see Myrvold 2002, 1775-6).

[168] At this point the ontology assumed by the modal interpretations enters the picture. Recall the these proposals state that quantum systems can have definite values for properties associated to a certain operator even if the system is not an eigenstate of that operator. We simply apply this ontological assumption to the properties A_i and B_i, they represent properties with a well-defined value although the quantum state is not a eigenstate of operator A and B . Myrvold sets the context of the argument we are revising in the following way: 'modal interpretations of quantum mechanics posit that the state vector obeys linear, unitary evolution at all times, and supplement the state vector with a set of possessed properties sufficiently rich to account for the occurrence of definite events at the macroscopic level, including definite outcomes for experiments, but sufficiently restricted to avoid a Kochen-Specker contradiction. The question arises whether this can be done within the restrictions imposed by special relativity. In a relativistic context, the notion of an instantaneous state of a spatially extended system must be replaced by the notion of a state on a spacelike hyperplane, or, more generally, a spacelike hypersurface. Since hyperplanes belonging to distinct foliations will intersect, we must ask whether the definite properties assigned to systems on these intersecting hyperplanes can be made to mesh in a coherent way' (2002), 1773.

[169] As Dieks explains, in the context of Bohm's version of the EPR experiment 'it is a mathematical fact that the Bell inequalities remain satisfied as long as the spin values found in the measurements on the individual particles can be regarded as coming from one joint probability distribution. The latter would be the case if the measurement results were determined by spin values already jointly possessed by the electrons, independently of which—or whether—measurements are made. If that were true, there would be well defined, definite spin values in the four directions under discussion in each run of the experiment; in repetitions these values would vary and form an ensemble that defines a joint distribution of the four

F_1 and B_2 ate F_2 constitute space-time events that exist independently of the context, i.e., the spacelike separated events with which they are correlated' (Dieks 2005, 412).

Thus, if the modal interpretation is to be viable, the values of the properties A_i and B_i must be conceived as frame-relative—that is why recent developments of the modal approach are dubbed *perspectivalist* (see Bene and Dieks 2002, for example). Two problems result from this maneuver. First, we already mentioned that there a frame-relativity of the kind we are discussing is quite hard to swallow. If perspectivalism is going to be considered as a viable and fruitful stance, further argumentation is required. To my mind, at least, that it is required for the viability of the modal interpretation is not enough to claim that perspectivalism is a philosophically and physically cogent view.

Second, to avoid the crash against Bell's theorem the modal interpretation must avoid to generate the problematic joint probability distribution. Now, what has been said above indicates that such a maneuver seems to be associated to a rejection of what Myrvold calls 'the relativistic Born rule'. The rule states that 'for any spacelike hypersurface σ, if the quantum state of the combined system $S_1 \oplus S_2$ on σ is $\rho(\sigma)$, and if X_1 and Y_2 are local definite properties of S_1 and S_2 on σ, then the probability that $X_1 = x$ and $Y_2 = y$ on σ is equal to $\mathrm{Tr}\left[P_{X_1}(x)P_{Y_2}(y)\rho(\sigma)\right]$, where $P_{X_1}(x)$ and $P_{Y_2}(y)$ are the projections onto the eigenspaces $X_1 = x$ and $Y_2 = y$, respectively' (2002, 1777). That is, if a perspectivalist formulation of the modal interpretation is Lorentz-invariant, this does not guarantee a full compatibility with SR, for if such an invariance is achieved by rejecting the relativistic Born rule, then there would be a clear violation of the spirit of the relativity principle anyway—Myrvold states that this is exactly what happens in an earlier proposal introduced by Dieks (1998): the theory is Lorentz-invariant, but it does not respect the relativistic Born rule. Actually, that Lorentz-invariance is not a sufficient condition for compatibility with SR can be adopted as a general principle. As we saw in chapter 2, the LPT is a Lorentz invariant theory, but it is incompatible with SR. Lorentz-invariance is to be taken as a *sign* of compatibility with SR, granted that such an invariance is an expression that the theory considered respects the chrono-geometric structure of space-time that Einstein's theory purports. The relativity principle is of course one of the essential features of that structure, so that if a modal perspectival interpretation of SQM is capable to achieve Lorentz-invariance on the pain of the rejection of the relativistic Born rule, then Lorentz-invariance could not be taken as a sign of full compatibility with SR.

In the case of the many worlds interpretation, the common view among its supporters is that no problems regarding non-locality and compatibility with SR arise. Wallace states that 'the overall story about locality in Everettian quantum physics, then, is this: the dynamics of the theory are local: there is no action at a distance, and no clash with relativistic covariance' (2012, 304-5). He supports that 'the dynamics of the theory are local' by invoking a principle of quantum field theory, namely, that 'spacelike separated field operators commute', so that the statistics of the expectation values to be recorded in a region A cannot be manipulated by making measurements in a space-like separated region B. This, of course, amounts to providing a physical foundation for the observance of 'parameter independence' invoking quantum field theory, but we saw that observance of this condition is neither sufficient nor necessary for compatibility with SR. furthermore, reference to this principle has been criticized for being beside the point (Maudlin 2010, 178), or even question-begging (Seevinck 2010).

In the particular case of EPR-like situations, Wallace claims that there are no non-local effects associated to the branching involved in measurements of the space-like distant particles. He states that

spin quantities. Only two of them could actually be measured in any single experimental run (one direction for each particle); but the measured values would evidently be samples from a joint distribution. The violation of Bell's inequalities by the predictions of quantum mechanics, and by the experimental results, therefore shows that we cannot think of the EPR situation in classical terms—the measurements do not reveal pre-existing jointly defined quantities' (2005, 409). Actually, the table of joint probabilities constructed in the derivation of the Bell-Wigner inequality we revised in section 3.4.4 could be regarded as coming from one joint probability distribution.

> In Everettian quantum mechanics, violations of Bell's inequality are relatively uninteresting. For Bell's theorem, though its conclusion entails not only non-separability but action at a distance [...], simply does not apply to the Everett interpretation. It assumes, tacitly, among its premises that experiments have unique, definite outcomes.
> From the perspective of a given experimenter, of course, her experiment does have a unique, definite outcome, even in the Everett interpretation. *But Bell's theorem requires more: it requires that from her perspective, her distant colleague's experiment also has a definite outcome. This is not the case in Everettian quantum mechanics – not, at any rate, until that distant experiment enters her past light cone*. And from the third-person perspective from which Bell's theorem is normally discussed, no experiment has any unique definite outcome at all. (2012, 310, my emphasis)

Wallace's analysis is, to my mind, too brief and rather obscure. However, considering his discussion of measurements on particles in superposed spin states (308-10), and his reference to Timpson & Brown (2005)[170], I think that the idea behind this passage is the following. Timpson & Brown state that the distinction between proper and improper mixtures, in the context of a relative-state view, must be understood in the following way. Consider the mixed state described by the density operator $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. Assume that the states $|\psi_j\rangle$ are generated by a preparation device, and which specific state of the $j = 1, \dots, n$ is prepared is randomly determined by a quantum die that operates a knob in the device. The orthogonal states of the die are given by $|d_j\rangle$, with $j = 1, \dots, n$ so that each eigenstate is correlated with one of the possible object system states $|\psi_j\rangle$. If the die begins in a superposed state $|D\rangle = \sum_j \sqrt{p_j}\,|d_j\rangle$, then the joint state of preparation device and object system is given by $|\Psi\rangle = \sum_j \sqrt{p_j}\,|d_j\rangle|\psi_j\rangle$, so that the reduced density matrix of the object system is $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, as we assumed originally. Now, and from the stance of the relative-state interpretation, we have that whether this mixed state corresponds to a proper or improper mixture is a relational issue: if there is an interaction of the systems with an observer, then, for such an observer, the state represents a proper mixture—more precisely, in each branch of the total object system + observer, the object system is a proper mixture relative to the observer. If the observer does not interact with the system, then for such an observer it is an improper mixture:

> the characteristic feature of a proper mixture is that there is some fact about the state our object system is in, that goes beyond the density operator ascribed to it. In the no-collapse context, such a fact must be understood as relational, that is, as a matter of correlations between the object system, or systems, and states of the environment and observers. Thus we may understand the preparation procedure just outlined as giving rise to a proper mixture if it turns out that following the preparation, the relative state of the object system with respect to the state of some particular observer is one of the states $|\psi_j\rangle$. For this to happen, the interaction between the systems involved in the preparation procedure and the environment must be such that the observer will become correlated to the die and object system states, or, must be such that an effectively classical record (that is, a record robust against decoherence) of the state of the die and object system is left in the environment.
> Following the preparation of state $|\Psi\rangle$ in $[|\Psi\rangle = \sum_j \sqrt{p_j}\,|d_j\rangle|\psi_j\rangle]$, then, we can imagine two distinct scenarios. In the first, an observer, Alice, indeed becomes correlated to the states produced in the preparation procedure; with respect to her, the object system is in a proper mixture. In the second scenario, another observer, Bob, remains uncorrelated to the states $|d_j\rangle|\psi_j\rangle$; with respect to him, the object system is in an improper mixture. It is in this sense that the proper/improper mixture distinction becomes relative to the experimental context in no-collapse quantum mechanics. (Timpson & Brown 2005, 3)

[170] Wallace refers to Timpson & Brown (2005) '*for a more detailed analysis of Bell's theorem in an Everettian context*' (310). Oddly, this paper does not directly deal with Bell's theorem in connection with the many worlds interpretation. It contains interesting insights about what is the precise sense of the distinction between improper and proper mixtures in the context of no-collapse interpretations (especially relative-state ones), but, excepting the passage I quote, there is no discussion of Bell's theorem in it. Actually, Timpson & Brown do *not* explicitly assert that Everettian views of quantum theory are non-locality free. Wallace's assertion that the Everett interpretation does not purport non-locality or friction with SR may be right, but he concludes it in too quick a manner.

Now, consider the singlet state $\frac{1}{\sqrt{2}}(|+\rangle_1|-\rangle_2 - |-\rangle_1|+\rangle_2)$, where particle 1 is Alice's and particle 2 is Bob's. For each particle, the corresponding reduced density operator traces-off the statistics of the other particle. Take it that Bob performs a measurement on his particle. I think that what Wallace has in mind is that since the observer, Bob, interacts only with his own particle, but not with particle 1, we have that the reduced density operator corresponding to particle 2 represents, relative to Bob, a proper mixture for which 'there is some fact about the state our object system is in, that goes beyond the density operator ascribed to it', whereas the reduced density operator corresponding to particle 1 represents an improper mixture for which there is no fact of the matter regarding what is its definite spin state.

According to the many worlds interpretation, when Bob performs the measurement there is a branching in two different worlds corresponding to each possible spin value. In both worlds, Bob's particle has a definite spin state, but Alice's doesn't—for its state represents an improper mixture. Only when an observer (Alice or anyone else) interacts with particle 1 it acquires a definite spin state (in that same world-branch), and since the whole process evolves unitarily, then the value that particle 1 exhibits yields the correlations predicted by quantum theory. This is, I think, why Wallace affirms that from the experimenter's point of view (Bob's), the experiment performed by his distant colleague does not have a definite result, at any rate, until that experiment enters his past lightcone. More generally, when Bob's measurement splits the world in two, in both resulting branches particle 2 has a definite state, but not particle 1—this will happen only when an observer interacts with it. Thus, the measurement that Bob performs does not yield any non-local effect, so that no menace for the compatibility between quantum theory and SR arises.

If this view is correct (assuming I reconstructed it well), and if we forget about the rather bizarre (untenable?) ontology of the many worlds interpretation, we would have not only a solution for the measurement problem, but we would also avoid all the conflicts with SR that non-locality brings on. However, some questionings can be posed. First, we have that even though in the world that results from Bob's measurement we have that particle 1 does not have a definite spin state until it gets related to an observer, we (also Bob) know with probability one what value it will yield if it is measured in the same spin direction. It is very strange to say that the particle has a probability one of exhibiting a certain value for a property in a measurement, but at the same time to say that it does not have a definite state for that property. Furthermore, given the space-like separation between Alice's and Bob's measurements, there are reference frames in which Alice's happened first, so that the problematic frame-relativity of properties comes back with a vengeance: in such frames it is Alice's particle that has a definite spin value, but not Bob's; and if in the branching world that results from Bob's measurements there are frames in which Alice measured particle 1 before Bob measured particle 2, can we still maintain that Bob's measurement determined the branching? Finally, why should we assume that the branching is given by the interaction between Bob and particle 2 as described by is reduced density operator? Why not simply assume that the branching occurs according to $\frac{1}{\sqrt{2}}(|+\rangle_1|-\rangle_2|O\rangle - |-\rangle_1|+\rangle_2|O\rangle)$, where $|O\rangle$ stands for the observer state? In such a case is clear that the branching would yield a non-local effect. I introduce these queries in order to show that the common view among supporters of the many worlds interpretation has still many gaps to fill. It may be right, but, at least to my mind, the way it is usually defended is yet rather obscure.

From this analysis we can then conclude that the situation between SQM and SR is far from being given by a peaceful coexistence. Jarrett's and Shimony's analyses are not correct. That SQM respects $p(A|\hat{a},\hat{b},\lambda) = p(A|\hat{a},\lambda)$, but violates $p(A|\hat{a},\hat{b},B,\lambda) = p(A|\hat{a},\hat{b},\lambda)$, cannot be interpreted as a sign of its compatibility with SR. The violation of outcome independence certainly implies non-local causal correlations which, assuming a wave-function collapse, imply either the introduction of a preferred foliation of space-time or a strange form of frame-relativity regarding physical properties. Modal interpretations of SQM, which do not include a projection postulate, crash with Bell's theorem if both Lorentz-invariance and the relativistic Born rule are assumed. This empirical unviability can be avoided by introducing the

same strange form of frame-relativity of physical properties just mentioned. In addition, if Lorentz-invariance, in a perspectivalist context, is achieved, it cannot count as a sign of true compatibility with SR unless the relativistic Born rule is respected. The many worlds interpretation looks like a more promising approach in this sense. However, although the idea that the branching of worlds in terms of measurements is not committed to non-local effects seems to be a possible way out, there are still many questions and worries that supporters of this approach must answer clearly and explicitly. The upshot of this conclusion is that the fact that BQT is not compatible with SR cannot be taken as a definitive reason in order to dismiss it in favor of SQM—supporters of Bohm's theory could simply reply that *tu quoque*. Anyway, it is in order to take a deeper look to the specific way in which BQT conflicts with SR, and also to evaluate how serious are the consequences of such a conflict.

The friction between SR and BQT is normally expressed by saying that Bohm's proposal cannot be made a Lorentz-invariant theory. Berndl et al. (1996) have offered a general proof of this. The reasoning is actually analogue to the theorem by Myrvold (2002) above. Violations of Bell's inequalities imply that a single joint probability distribution for spin values in all directions cannot exist. Put differently, th4is means that there are (entangled) states that draw inconsistent probabilities for spin values for measurements in different combinations of spin-directions[171]. Now, if Bohm's theory were Lorentz-invariant in the sense that quantum equilibrium (the distribution postulate $P = |\Psi|^2$) holds in all reference frames, then consistency of the probabilities predicted for measurement in all possible directions would follow—so one single joint probability distribution for all spin measurements would exist. But BQT certainly violates Bell's inequalities, which implies that quantum equilibrium cannot hold in all reference frames:

> We consider an arbitrary theory for $N(\geq 2)$ particles, i.e., a (possibly statistical) specification of all possible N-tuples of space-time paths for the N particles [...]. We shall call each such possible ''history'' an *N-path*. We assume that each spacelike hypersurface is crossed exactly once by each trajectory, and consider an arbitrary probability measure P on the N-paths. This determines the distribution of crossings ρ^{Σ}: $\Sigma^N \rightarrow \mathbb{R}$ for any spacelike hypersurface Σ.
> We now want the probabilistic predictions of the theory to agree as far as possible with those of quantum theory. Complete agreement would be straightforward if for any quantum state ψ there were a P such that for all spacelike hypersurfaces Σ the distribution of crossings ρ^{Σ} agrees with the quantum mechanical joint distribution of the (measured) positions on Σ. For Σ a spacelike hyperplane, i.e., a simultaneity plane or constant-time slice of a Lorentz frame Λ, this is given by $|\psi^{\Sigma}|^2$ where $\psi^{\Sigma} = \psi$, the wave function in frame Λ. However, this is not in general possible.
> *Assertion: There does not in general exist a probability measure P on N-paths for which the distribution of crossings ρ^{Σ} agrees with the corresponding quantum mechanical distribution on all spacelike hyperplanes* Σ. [...]
> The assertion above is more or less an immediate consequence of any of the no-hidden-variables nonlocality theorems [...] for the spin components of a multiparticle system: By means of a suitable placement of appropriate Stern-Gerlach magnets the inconsistent joint spin correlations can be transformed to (the same) inconsistent joint spatial correlations for particles at different times. Since the existence of a probability measure P on N-paths implies the existence and hence the consistency of all crossing distributions, the assertion follows. (Berndl et al. 1996, 2064-5)

It is important to remark once again that the impossibility of Lorentz-invariance in Bohm's theory is not in and by itself a manifestation of BQT incompatibility with SR. Lorentz-invariance can be regarded this way inasmuch as it illustrates that the theory at issue does not postulate a metrical structure of space-time that contradicts or goes further than the structure of Minkowski space-time. Following Maudlin, we can assume a criterion for compatibility based on this observance of Minkowski metrical structure:

---

[171] The 'Hardy state', $\psi_{Hardy} = \frac{1}{\sqrt{3}}(|+\rangle_z^a|-\rangle_z^b - \sqrt{2}|-\rangle_z^a|+\rangle_z^b) = \frac{1}{\sqrt{3}}(|-\rangle_z^a|+\rangle_z^b - \sqrt{2}|-\rangle_z^a|+\rangle_x^b) = \frac{1}{\sqrt{3}}(|+\rangle_z^a|-\rangle_z^b - |+\rangle_z^a|+\rangle_z^b + |-\rangle_z^a|+\rangle_z^b) = \frac{1}{\sqrt{12}}(|+\rangle_x^a|+\rangle_x^b - |+\rangle_x^a|-\rangle_x^b + |-\rangle_x^a|+\rangle_x^b - 3|-\rangle_x^a|-\rangle_x^b)$, yields inconsistent probabilities for measurements performed along different directions combinations. Berndl et al.'s theorem is actually based on a theorem proved by Hardy (1992).

> Lorentz invariance turns out to be a roundabout route to a more fundamental property: the essential fact about Lorentz invariant theories is that their dynamics depend only on the Special Relativistic metrical structure. And once put this way, all reference to coordinate systems and coordinate transformations may be dropped. Given, for example, a coordinate-free formulation of a theory, we may ask whether it postulates only the relativistic space-time structure or whether it posits more. (1996, 291)
>
> *Relativistic Constraint*: a theory is compatible with Relativity if it can be formulated without ascribing to space-time any more or different intrinsic structure than the (special or general) relativistic metric. (ibid, 292)

It is easy to see that BQT breaks this relativistic constraint. As we saw in section 3.5.3, in Bohm's theory the wave function ψ and its corresponding quantum potential Q for n-particles systems are defined in $3n$-configuration space. Accordingly, the momentum of the i^{th} particle in the system is given by $\mathbf{p}_i = \nabla_i S(\mathbf{x}_1, \dots, \mathbf{x}_n, t)$, so that the equation of motion is $\frac{d\mathbf{x}_i}{dt} = \mathbf{v}_i(\mathbf{x}_1, \dots, \mathbf{x}_n, t) = \frac{1}{m_i}\nabla_i S(\mathbf{x}_1, \dots, \mathbf{x}_n, t)$. That is, the position of the i^{th} particle in the systems at a certain instant depends on the position of the rest of the particles in the system at that same instant. Assuming that the particles are space-like separated, it is clear that the theory postulates a preferred foliation of space-time that allows to make sense of the expression 'at that same instant'. In other words, the non-local trajectory correlation mediated by the quantum potential Q as defined in configuration space presupposes a specific foliation of space-time that determines a time-sameness in $3n$-space:

> One of the problems which face relativistic quantum mechanics is that the notion of a configuration, as used in the non-relativistic version, already presupposes simultaneity. Configurations are *configurations at a time*, they specify where all the particles in a system are *at a given moment*. So the very notion of a configuration is not a Lorentz invariant concept. [...]
> So given the way simultaneity is presupposed in the very notion of a configuration, together with the fact that the wave-function is defined over configuration space, it is not surprising that some notion of simultaneity would find its way inextricably into Bohm's theory. (Maudlin 2010, 198)

This implicit postulation of a preferred frame, it must be underscored, manifests only at a fundamental level, but not at a phenomenological level. That is, even though the theory is not Lorentz-invariant, this feature is not reflected in the observable predictions of the theory. In this sense, we could even say that BQT is 'phenomenologically Lorentz-invariant'[172].

[172] As Cushing explains: 'An illustration of the non-Lorentz invariance of quantum mechanics at the level of individual events is provided by returning to the well-known EPRB experiment. Let the two electrons be emitted simultaneously (in a singlet state) from a source. We can calculate the marginal, conditional and joint probabilities for detection at the two distant stations in different Lorentz frames. That is, we can arrange (simply by choosing suitable Lorentz frames) to have either one of the detection events at a given station follow that at the other, or vice versa, or we can arrange to have them occur simultaneously. Even though the quantum state of the system between observations as seen by the observers in different Lorentz frames is different, the *predictions* for observations turn out to be the same (i.e., Lorentz invariant) upon explicit calculation. And, these different states are *not* the Lorentz transformations of each other. At the level of the state vector, all Lorentz frames are *not* equivalent, although all predictions for observables are. If one takes the wave function as being merely a mathematical device for calculation, then we need not see a conflict with Lorentz invariance (i.e., with "relativity", loosely speaking). However, if one takes the wave function as representing a physical reality (as for Bohm, where the wave function determines actual particle trajectories through the guidance condition), then all Lorentz frames are not equivalent at the level of individual processes. Such a result is not unexpected when we appreciate that the time evolution of the Schrodinger equation (which is itself not Lorentz invariant) will be different in different Lorentz frames. The dynamics governing the motion of a particle depends upon the wave function in a nonlocal fashion [...]. Although there is a unique (but not experimentally distinguishable) frame in which the nonlocal correlations are instantaneous, the statistical laws are covariant. [...]
> The basic idea that emerges naturally from all of these considerations is that Lorentz invariance is a (quantum) equilibrium symmetry, rather than a fundamental law of nature. At the most fundamental level (i.e., that of Bohm's trajectories), there is *no* Lorentz symmetry and this is indicated by the fact that the basic laws (such as the guidance condition for the Bohmian trajectories) are valid only in the preferred frame' (1996, 177-8).

This last point becomes relevant when one ponders about how bad is the need to introduce the preferred foliation of space-time that BQT requires: it has no observable effects, and, accordingly, it is impossible to determine which reference frame is associated to the preferred simultaneity foliation. To properly evaluate the seriousness of this issue, we may compare it with the case of the ether in the LPT. I argued in the previous chapter that the mere observability of the ether was not a definitive reason to dismiss the theory, or even to definitively conclude that the ether does not exist. There are plenty of examples in modern physics of concepts which refer to (directly) unobservable entities, but which—typically through the explanatory role they play—are nevertheless connected to observable features of the theory. In the case of the ether, we had that in the LPT theory empirical effects such as the v-dependence of mass observed in Kaufmann's experiments, and length-contraction and clock-retardation, could be understood as *traces* of the ether.

Something similar occurs in the case of BQT and the preferred foliation that corresponds to the non-local effects caused by the quantum potential Q: we could actually interpret the results of Aspect's experiments as observable traces of Q. Non-local correlations between the space-like separated components of entangled systems are observable quantum features, and, according to Bohm's theory, it is the quantum potential Q which mediates those correlations 'along' a preferred space-time foliation. Thus, all non-local effects could actually be understood as observable traces of Q, which in turn is associated to the preferred foliation we are considering. Furthermore, we have that the undetectability of the preferred foliation associated to non-local effects in BQT is grounded on, and *explained* by, an essential feature of the theory: the dynamics of measurement interactions. We saw above that we can have only a statistical knowledge of the exact trajectories of the particles: we do not know where the particles corresponding to a specific system in the ensemble associated to a wave function are, we can only know that the systems in the ensemble are distributed in their locations according to $P = |\Psi|^2$, and if we tried to find out that specific location by a measurement, we would disturb the wave function in a way such that the outcome could not be used in order to determine the future trajectory of the system. Now, given that in entangled systems the position of a particle at a certain time depends on the positions of the other particles at that same instant, if we were able to determine the trajectories with precision and without disturbing the wave function, then we could actually determine the preferred foliation.

I think then that the undetectability of the preferred foliation is not that bad. In the case of the ether in the LPT, we saw that Poincaré's proof that the Lorentz transformations are symmetric made the ether not only undetectable but superfluous, in the sense that it became formally idle for the derivation of the empirical consequences of the theory. In the case of BQT and the quantum potential Q—which is the entity associated to the foliation—no signs of this form of superfluity are present. Thus, I agree with Maudlin in that 'if the existence of empirically inaccessible physical facts is fatal, then Bohmian mechanics is a non-starter even before Relativity comes into play. Conversely, if one takes Bohm's theory seriously as a possibility (as one should!), then the fact that the foliation is hidden from view cannot by itself imply that the theory is untenable' (1996, 296).

When we turn to the way in which the preferred foliation relates to the chrono-geometric structure of space-time as postulated by SR, things become more serious and complex. In order to throw some light on this issue, it is useful to consider two different ways in which a theory can depart from relativity: by rejecting the Minkowski structure altogether, or by adding some structure that is not postulated by SR; and also two different ways in which a theory can introduce space-time structure not considered by SR: adding 'intrinsic' or 'non-intrinsic' structure. The question is now in which side of these two dichotomies we may locate BQT.

Let us consider first the distinction between 'intrinsic' and 'non-intrinsic' (or 'not-so-intrinsic') structure. A possible way to describe the difference is the following:

> In Minkowski space-time, the postulation of a preferred foliation into flat spacelike hyperplanes, unaffected by all mater fields, would clearly count as more intrinsic structure, indeed as the postulation of absolute simultaneity. It would help make such a foliation look less "intrinsic" if it were coupled to some matter fields somehow, so that the exact form of the foliation depended on the distribution of matter [...]. If the foliation is not intrinsic, then one expects that different initial conditions could have produced space-times with the same metric but different foliations, and one would like to know what the relevant initial conditions are. (Maudlin 1996, 292-3)

If the preferred foliation that BQT requires to be postulated is conceived as intrinsic space-time structure, as the quote indicates, this would amount to a postulation not only of a preferred reference-frame, but also of absolute simultaneity. Now, if we interpret that BQT obliges us to think that the space-time we inhabit includes a preferred foliation as a part of its *intrinsic structure* to which the non-local effects featured by the quantum potential Q 'accommodate', the upshot would be that BQT departs from SR by effectively rejecting the Minkowski structure of space-time. If this were the case, then the situation would be very dramatic. BQT would tell us something like 'SR is false, it only looks correct because of our limited knowledge of the initial conditions of physical systems', but then the unobservability of the physical facts that, according to this interpretation of BQT, render SR false would make the theory rather dubious: 'Won't any theory according to which Relativity is *false* involve a very odd sort of conspiracy of physical law so that a non-relativistic space-time appears to be relativistic?' (Maudlin 1996, 296-7)[173]. In order to illustrate the dramatic nature of this scenario we could simply notice that the kinematic explanations of the typical relativistic effects such as v-dependence of mass (and thus also $E = mc^2$), length-contraction and clock-retardation could not be accepted by the Bohmian (if our space-time yields absolute simultaneity, then none of these physical facts can be explained in the way that SR does, of course). But there is nothing in Bohm's theory that we could refer to in order to provide an alternative explanation! Thus, if a strong case for Bohm's theory is to be introduced, this scenario must be certainly avoided.

In my opinion, Bohm's theory does not *oblige* us to conceive the preferred foliation as the expression of intrinsic structure, and thus as a hyperplane of *absolute simultaneity*. First, BQT is not a *space-time* theory. Although the non-local features it contains lead us to introduce a preferred foliation, the main objective of the theory is to describe the dynamics of the quantum world, not the description of the chrono-geometric structure of the physical world. That is, the preferred foliation is an upshot of the specific quantum dynamics described by the theory, and the fact that the preferred foliation is dynamically determined is actually a first indication that the preferred foliation does not need to be thought of as intrinsic metric structure. If one thinks carefully about it, BQT does not say much about the structure of space-time. The non-local features explained just above and in section 3.5.3 give us to think about the structure of space-time granted that we conceive the dynamics of BQT as occurring within a certain space-time with a certain structure, but this description comes from a different theory. Suppose that we had no space-time theory at all. In this case, I think, it is clear that the non-local effects would not be taken as implying either an intrinsic or a non-intrinsic structure, we would take them only as a peculiar dynamical feature. This is what I mean when I state that BQT is not a space-time theory: the non-local effects purported are consequences of the dynamics of the theory, and BQT does not really describe space-time

[173] Maudlin actually evaluates the introduction of the preferred foliation as an introduction of intrinsic structure, that is, not as dynamically grounded but as a matter of the kinematics associated to the guidance equation: 'Since there are no real collapses in Bohm's theory, *the wave function has no need of the foliations*. So the wave function could evolve in accord with a covariant equation, such as the Dirac equation. *The guidance equation for the particles* (or other local beables) *would be framed with reference to the foliation*. The ontological duality already at the base of the theory would map onto similar duality of space-time structure, with the accuracy and effectiveness of Relativity being explained by the completely relativistic nature of the wave function. *The beables would be sensitive to absolute simultaneity*, but only in a completely unobservable way' (1996, 297). However, Maudlin thinks that this description yields that BQT simply adds extra-structure, but retains the relativistic metric at the fundamental level. He does not notice that this way to interpret the introduction of the preferred foliation, insofar as it introduces absolute simultaneity, implies a full rejection of Minkowski structure: if there is absolute simultaneity, the relativistic explanation of the length-contraction and clock-retardation cannot be correct!

structure at all. Depending on what space-time structure scaffolding we consider, we may interpret what chrono-geometric structure corresponds to the non-local effects.

Second, and now explicitly considering that we accept SR as our space-time theory, we have that the fact that the preferred foliation associated to the non-local effects is dynamically grounded—on the non-local description and action of the quantum potential Q—implies that we can certainly understand such a foliation as non-intrinsic structure, and in this case there would be no need to understand such a foliation as a hyperplane of *absolute simultaneity*. Imagine that Minkowski space-time would have been there before any physical processes occurred within it. When the physical processes begin, they obey the laws of BQT. Those laws feature non-local effects, of course. But this fact does not mean that the underlying chrono-geometric structure of space-time gets altered by quantum physical processes, and that a hyperplane of absolute simultaneity emerges, de-Minkwoskivizing space-time—if you allow me the expression. It only means that there are space-like separated events that are causally and dynamically connected. Dynamically understood, this connection would not imply an absolute temporal order between space-like separated events. We would be in a situation in which there are reference frames in which the cause (or the antecedent, if we want to remain more agnostic about the nature of the causal connection) occurs after the effect (consequent). But since the preferred foliation only corresponds to a causal connection between *space-like* separated events, this is not a big deal—after all, we already know that the form of causality involved in non-local correlations does not correspond to the notion of causality in classical physics. Under this interpretation, BQT would still yield a preferred reference frame: the reference frame in which the time sameness for events described in configuration space corresponds to the foliation that joins the correlated non-local effects, the frame in which the preferred foliation is a simultaneity hyperplane. However, and considering the remarks in the previous paragraph, this preferred reference frame would be privileged in a dynamical way, not in a chrono-geometric way—for the corresponding simultaneity hyperplane does not need to be thought of as an *absolute* simultaneity hyperplane.

This way to understand BQT is certainly less dramatic when it comes to its relationship to SR. The dynamic interpretation of the preferred foliation entails that the way in which Bohm's theory departs from SR is by introducing some non-intrinsic structure that is not considered by Einstein's theory, but not by rejecting Minkowski structure altogether. This would still imply an important point of departure: the introduction of a dynamically preferred frame, which would amount to a weakening of the relativity principle. As we saw above, the quantum equilibrium postulate cannot hold in all reference frames, and since this postulate is law-like determined (remember that by inserting $\Psi = Re^{iS/\hbar}$ in the Schrödinger equation it follows that $P = R^2$), then there are laws that do not look the same in all reference frames. However, and once again, this case of Lorentz-invariance implies neither a divergence in terms of empirical predictions with respect to SR, nor that Minkwoski space-time structure is rejected. It only implies that a dynamically (not kinematically) preferred space-time foliation emerges from the dynamics of the theory.

We have then that if we compare SQM with BQT regarding the way they stand with respect to SR, we have that no definitive conclusions can be drawn in order to pick one of the theories and reject the other. The 'received view' that SQM and SR can peacefully coexist insofar as the former violates 'outcome independence' but not 'parameter independence' is mistaken. Violations of outcome independence certainly yield non-local causal connections that witness that some non-relativistic space-time structure is presupposed. Though this is especially clear when we assume the projection postulate, modal non-collapse interpretations of SQM must also face important difficulties associated to Lorentz-invariance. Thus, we cannot use an argument that claims that whilst BQT is in open conflict with SR, SQM peacefully coexist with it, in order to make a choice favoring SQM. Actually, if we notice that the source of conflict with SR in is in all cases grounded *in Bell's theorem* (in the case of SQM + projection postulate it is violation of factorizability (outcome independence) what brings the problem, whereas in the case of the modal interpretations and of BQT, both Myrvold's and Berndl et al.'s proofs rely on Bell's theorem) then the

conclusion that a conflict with SR is a general problem that quantum *physics* must face is hardly surprising—for Bell's theorem holds for any possible quantum theory. However, we have also seen that the friction between SR and BQT is not as bad as it seems at first sight if we interpret the preferred foliation as corresponding to a dynamical feature of the theory, and not as an expression of intrinsic metric structure postulated by it.

After this lengthy analysis of the case of SQM vs. BQT we can finally turn to the general conclusions that can be extracted from this work. We will look at the results of chapters 2 and 3 in perspective, in order to test the philosophical appraisal of the problem of EE and UD defended in chapter one. Besides, we will use this philosophical appraisal in order to provide an accurate evaluation of the two case-studies here addressed.

CONCLUSIONS

In the first chapter of this thesis I argued for a reassessment of Laudan & Leplin's views on EE and UD that resulted in an adequate appraisal of the status of the problem at issue. There I concluded that

i) even though algorithms and the Q-D thesis are ineffective in providing an EE rival theory T' given any theory T, and although EE is a time-indexed condition between two theories, EE between scientific theories that results in UD is still a possible scenario;

ii) that recourse to non-empirical features can work as a partial solution of the problem, in the sense that such features can provide rational motives in order to prefer one of the theories, but cannot ground a uniquely determined and fully objective choice;

iii) that Laudan and Leplin are right in that the UD of the choice to be made between EE theories can be removed by the regular practice of science: by new auxiliary hypothesis or observation methods that can break the EE, or by non-entailed empirical evidence (grounded on intertheoretic connections of the theories involved) that can break the UD;

iv) these possible solutions do not count as a complete removal of the problem: that the EE or the UD will be thus eliminated is a contingent matter, there is no warrant that the development of science will be such that the problem gets solved in every case, so that recalcitrant UD as a result of EE is a possible scenario; and

v) the problem of EE and UD, when present, is a problem for *science* to solve, and, just as in any other scientific problem, that a solution will be found is not guaranteed from the outset.

In chapters 2 and 3 I offered a detailed analysis of two examples of EE in modern physics. Equipped with the results of these analyses, we can now turn to an evaluation of these conclusions. We will see that the careful examination of both cases provided in the previous chapters clearly illustrates the adequacy of statements *i)-v)*. In turn, these conclusions allow an accurate evaluation of the assessed case-studies. That is, they establish both a sound description of the nature and status of the problem of EE and UD, and a rigorous and compelling appraisal of the cases of the LPT vs. SR, and SQM vs. BQT.

Conclusion *i)* is obviously exemplified by the case-studies. Both the LPT vs. SR and BQT vs. SQM are true instances of EE, in the sense that in neither case we had that the theories were just two different formulations of a single theory. As we saw in section 2.4.2, the rivalry between the LPT and SR consists in that the theories diverge in the chrono-geometric structure they postulate for the physical world. According to the former, we inhabit a Newtonian space-time in which some 'conspiring' dynamical effects deceive us in our chrono-geometric measurements, a deception that results in observable effects like clock-retardation, length-contraction, v-dependence of mass, and so on. On the other hand, Einstein's theory tells us that the chrono-geometric structure of the physical world is given by Minkowski space-time, and that all the mentioned effects are the outcome of the kinematics imposed by the chrono-geometric structure of space-time. Now, given that the theories postulate the same group of coordinate transformations, and granted that we include $E = mc^2$ and Poincaré's amendments, the predictions of the theories are the same.

In the case of SQM vs. BQT, the rivalry can be expressed in a twofold way (see section 3.5.4). First, the rivalry is given by the different ontology that BQT postulates with respect to (any of the interpretations of) SQM: according to Bohm's theory particles have always well-defined trajectories. Second, unlike (any

of the interpretations of) SQM, in Bohm's theory the description of a physical state given by the wave function Ψ is incomplete, Ψ describes the quantum 'field' associated to an ensemble of particles or it denotes a law-like term, but it does not completely determine the positions of such particles, so we have only a statistical knowledge of their distributions given by $P = |\Psi|^2$. In other words, whereas (the usual interpretations of) SQM tells us that Ψ is a complete description of quantum systems, BQT is a HVT. The predictive equivalence between the theories is insured if the distribution postulate is assumed in BQT. Thus, the EE between SR and the LPT and between SQM and BQT certainly involves UD. As we will see, in the first case the UD was removed, whereas in the second it remains—so we have only the partial solution provided by non-empirical features

Conclusion *ii)* is also exemplified by the analyses provided in chapters 2 and 3. Let us recall the reasons that could and have been invoked in the case of Einstein vs. Lorentz. We saw in sections 2.5.1 and 2.5.2, respectively, that accusations of ad-hocness on the length-contraction hypothesis are unfounded, and that recourse to the mathematic-aesthetic features of SR could be predicated, *mutatis mutandis*, also of the LPT. Thus, these two possible non-empirical features cannot be used to ground a preference. However, in sections 2.5.3 and 2.5.4 we found two non-empirical features that did work as a criterion according to which SR is a better theory than its rival. First, we saw that although Janssen's argument can be severely criticized, its underlying idea can nevertheless be used to show that the physical simplicity of SR—expressed in its observance of Earman's symmetry principles—reflects in more solid foundations than in the LPT: that the latter theory violates the principles implies that the ether becomes a highly suspicious entity, in the sense that it may represent nothing physical. I drew a similar conclusion based on the analysis of the implications of the symmetry of the Lorentz transformations. This symmetry entails that the velocity term involved is always the relative velocity between the frames, so that the velocity with respect to the ether is not needed in the derivations of the observable predictions of the theory, and this form of superfluity of the terms makes it even more suspicious. Therefore, the problematic character of the ether supports a preference for Einstein's theory.

However, just as conclusion *ii)* states, if non-empirical features were all we had in order to justify our choice, then such a choice could not be fully objective and uniquely determined. We can imagine a supporter of Lorentz's theory that, when confronted with the mentioned arguments concerning the explanatory superiority or the physical simplicity of special relativity, simply replies 'I accept that Einstein's theory scores better on those issues, however, Lorentz's theory—by means of the dynamic explanations in a Newtonian space-time it promises for effects like length-contraction, clock-retardation and the v-dependence of mass—offers a simple, classical and intuitive visualization that special relativity cannot provide'[1]. The Lorentzian could actually invoke Einstein's own comparative evaluation in terms of explanatory power between theories of principles and constructive theories: he regarded constructive explanations as clearly superior[2]. The supporter of the LPT could trade the problematic physical foundations on which the theory is built for the constructive explanations that it offers, explanations which are

[1] As I mentioned in section 2.5.2, Scott Walter (2010) has shown that the difficult intuitive spatio-temporal visualization involved in Minkowski space-time was noticed and considered by physicists like Laue and Wiechert. It is true that they traded the (classical) visualization issue for the elegance and simplicity of the Minkowskian presentation of Einstein's theory, but our Lorentzian physicist might not be willing to make that deal—and if it is to be argued that the only rational decision is to prefer special relativity, a previous demonstration that elegance and simplicity are more important features than intuitive visualization is required.

[2] 'The advantages of the constructive theories are completeness, adaptability and clearness, those of the principle theory are logical perfection and security of the foundations. The theory of relativity belongs to the latter class. In order to grasp its nature, one needs first of all to become acquainted with the principles on which it is based [...]. When we say that we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the process in question' (Einstein 1919, 228).

absent in Einstein's theory of principle[3]. Actually, something like this seems to have been what Lorentz had in mind when he evaluated the theories:

> Einstein simply postulates what we have deduced, with some difficulty and not altogether satisfactorily, from the fundamental equations of the electromagnetic field. By doing so, he may certainly take credit for making us see in the negative result of experiments like those of Michelson, Rayleigh and Brace, not a fortuitous compensation of opposing effects but the manifestation of a general and fundamental principle.
> Yet, I think, something may also be claimed in favour of the form in which I have presented the theory. I cannot but regard the ether, which can be the seat of an electromagnetic field with its energy and its vibrations, as endowed with a certain degree of substantiality, however different it may be from all ordinary matter. *In this line of thought it seems natural not to assume at starting* that it can never make any difference whether a body moves through the ether or not, and to measure distances and lengths of time by means of rods and clocks having a fixed position relatively to the ether. (Lorentz 1916, 229-30, my emphasis).

The fact that the LPT offers an account of the phenomena that allows us to remain within the framework of classic electrodynamics and Newtonian space-time might be enough for the Lorentzians to simply dodge the fact that SR is a better theory with respect to other non-empirical features. All that is needed to defend such a position is to underscore that the ether and the preferred frame of reference in Lorentz's theory allow having a physical seat for the electromagnetic fields—as Lorentz states in the quoted passage—and an intuitive, 'classic' visualization of length-contraction and clock-retardation. To highly value these epistemic features is a position that can be, of course, defended[4].

Thus, if non-empirical features were all we have for deciding the case of Einstein vs. Lorentz, then the decision could not be determined in a fully objective and unique way. We would be in a sort of Kuhnian conundrum, in the sense that the decision could be made only by presupposing certain (non-empirical) standards of theory appraisal, standards which may not be shared by the whole scientific community. And since the theories are predictively equivalent, the merits of the competing standards of theory appraisal cannot be evaluated in terms of empirical success. The upshot of conclusion *ii)* is that although non-empirical features can offer rational motives to state a preference, there may be different features of this kind that can ground rational preferences that select each of the theories—both choices could be rationally defended by recourse to theoretical virtues, and if this were the only way to assess the case of Lorentz vs. Einstein, then a uniquely determined choice could not be done.

Something similar occurs in the case of BQT vs. SQM. Non-empirical features can be certainly invoked in order to rationally ground a preference, but what preference is to be adopted depends on what specific features we consider, and, especially, on how one evaluates their importance. In section 3.6.1 we saw that

---

[3] There are places in which Einstein himself described SR as having gaps in terms of constructive explanations. For example, he expresses concerns of this kind in a letter to Arnold Sommerfeld, dated January 1908: 'So, first to the question of whether I consider the relativistic treatment of, e.g, the mechanics of electrons as definitive. No, certainly not. It seems to me too that a physical theory can be satisfactory only when it builds up its structures from *elementary* foundations. The theory of relativity is not more conclusively and absolutely satisfactory than, for example, classical electrodynamics was before Boltzmann had interpreted entropy as probability. If the Michelson-Morley experiment had not put us in the worst predicament, no one would have perceived the relativity theory as a (half) salvation. Besides, I believe that we are still far from having satisfactory elementary foundations for electrical and mechanical processes. I have come to this pessimistic view mainly as a result of endless, vain efforts to interpret the second universal constant in Planck's radiation law in an intuitive way. I even seriously doubt that it will be possible to maintain the general validity of Maxwell's equations for empty space' (quoted in Brown 2005, pp. 72-3).

[4] During lectures he gave in Leiden in 1910-12, published in 1922, Lorentz describes the advantages in his theory—the ether as the carrier of electromagnetic fields and classic chrono-geometry—in the following way: 'whether there is an aether or not, electromagnetic fields certainly exist, and so also does the energy of oscillations. If we do not like the name of "aether", we must use another word as to peg to hang all these things upon. It is not certain whether 'space' can be so extended as to take care not only of the geometrical properties but also of the electric ones.
One cannot deny to the bearer of these properties a certain substantiality, and if so, then one may, in all modesty, call true time measured by clocks which are fixed in this medium, and consider simultaneity as a primary concept' (quoted in Brown 2005, p. 66).

it can be argued that BQT scores better than SQM in terms of explanatory power: Bohm's proposal offers a clear and intuitive description in contexts where SQM is very unclear and mysterious—the double-slit experiments and neutron-interferometry experiments, for example. However, Bohrians could simply say 'OK, Bohm's theory certainly provides an intuitive picture, but, according to the way that we understand SQM—the theory tells us that the world is essentially indeterministic and that the wave function is a complete-but-symbolic description of physical states—a semi-classical explanation like the one offered by Bohm is beside the point'. That is, if we are going to consider the explanatory power of BQT as a feature that favors this theory over SQM, we must presuppose that the world is amenable to the kind of descriptions proposed by the theory, but Bohrians deny this—and they can invoke the theory itself (SQM) and the evidence supporting it to defend their position! On the other hand, the modal interpretation of SQM, by means of the values state that 'underlies' the dynamical state, makes room for explanations of quantum phenomena that consider 'a determinate story to be told', so that a supporter of this approach could claim that, after all, we can understand SQM as providing a satisfactory explanatory framework

In section 3.6.2 we obtained a similar conclusion, but this time based on the fact that, unlike in SQM, in BQT identical particles are always distinguishable. Bohmians could highly praise this feature—if particles are always distinguishable, then the philosophical perplexities concerning objective identity do not even come up. But from the Bohrian point of view, this feature may not be regarded as a virtue. If we consider the wave function as a symbolic representation of physical states, as Bohr did, then the question about the identity of indistinguishable particles becomes beside the point. Furthermore, there exist at least three strategies that *any* of the interpretation of SQM can follow in order to cope with the question of identity in the context of indistinguishable particles: *haeccity* arguments, weak discernibility, and to limit the concept of objective identity to quantum fields and restrict the tag 'individualized particle' to entities that under certain conditions emerge from quantum fields. That is, supporters of SQM could simply concede that in BQT there are no special difficulties regarding objective identity in (anti)symmetrized quantum states, but they could also say that these difficulties can be overcome in the standard theory. Again, whether particle-distinguishability is to be considered as an especially important feature depends on what are our specific philosophical credos. If we want a world in which the objective identity criterion is classical, then distinguishability in terms of trajectories will be highly praised. However, supporters of SQM that cherish a principle of ontological and theoretical economy—so that they would reject any HVT—may be willing to lower their requirements for objective-identity criteria.

In section 3.6.3 we dealt with the question of the classical limit. It turns out that the way that the classical description arises in BQT is very natural and clear, though in SQM there are both technical and conceptual gaps. In Bohm's proposal, the vanishing of the quantum potential Q is a necessary and sufficient condition for classical behavior and description, so that a clear continuity between the quantum and the classical world can be seen. In SQM things are not so easy, Bohr's interpretation essentially considers a somewhat arbitrary 'cut' applied to measurement interactions. Alternative interpretations do not include such a cut, but the technical issue of connecting the quantum theoretical description with the classical description is not completely spelled out. Though it is true that research on environmental decoherence looks promising in this sense, there are still gaps to be filled. Thus, when it comes to the question of the classical limit, it is an objective fact that BQT scores better than SQM. However, though this comparative advantage can be invoked by Bohmians to ground their preference, supporters of SQM could stand their ground and argue that the (to their minds) advantages of SQM in other contexts—ontological and conceptual economy, for example—are worth to keep the faith in that a satisfactory answer provided by decoherence will be fully spelled out.

As we saw in section 3.6.4, the account of measurement interactions is yet another issue in which BQT scores objectively better than SQM. When a quantum system interacts with a measurement device, its corresponding wave function 'ramifies' in several branches, but the particle(s) in the system take only one of these branches—which of them is actually taken depends on the initial position of the particles.

Thus, measurement outcomes only tell us which of the branches was actually followed by the particle(s), so that no collapse at all is needed in order to describe measurement interactions. Compared to BQT, thus, the von Neumann-Wigner view of SQM looks deeply problematic when it comes to measurement descriptions. Though it is true that providing a solution for the measurement problem is one of the essential goals of the alternative no-collapse interpretations, such solutions are not complete or involve questionable assumptions. Everettian approaches evade the measurement problem on the price of postulating highly bizarre ontologies—and they face some technical problems like the meaning of probability and the preferred basis. Modal interpretations offer a more down to earth solution, but there remain some technical gaps to be filled. However, even though it is an objective fact that BQT is a more satisfactory theory from the point of view of measurement interactions description, no-collapse interpretations are still available as a rational stance for those who reject the HVT approach. Everettians may argue that the almost literal interpretation of the 5 postulates of SQM offered by their approach is worth to pay the price of a relative-state or branched ontology—for those who highly value *theoretical* economy, the Everettian stance may look more tenable than BQT. Those who cherish *ontological* economy may opt for the modal outlook, and hope that future development will provide a complete solution of the measurement problem. Again, it is rationally defensible to prefer some interpretation of SQM in spite of the fact that BQT does not imply any kind of measurement problem.

We can also adopt the opposite point of view: though there is an element in Bohm's theory that is very hard to accept, Bohmians could argue that the benefits that this element allows are worth to pay the price of its inclusion. In section 3.6.7 we saw that the ontological status of the wave-function in BQT is very strange. If we interpret that the term Ψ denotes a physical entity, it has to be a very odd one. Though it looks like a wave-field, it cannot be a wave in the regular sense, for it does not have any source, it does not respect Newton's third law, and it is associated to explicit non-local effects. To make things worse, the wave cannot be directly detected, we can only interpret that quantum effects are observable *traces* of its reality. The nomological approach avoids an ontological commitment to such a strange entity, but I showed that it falls into deep trouble. The Bohmian explanation of some experimental contexts in neutron interferometry strongly indicates that Ψ must be understood as denoting a real entity. Besides, the nomological approach implies a rejection of a central principle in space-time theories: that free particles follow geodesic trajectories. The meaning of the term Ψ represents thus a clear dilemma for the Bohmian, either she commits to a very strange and directly unobservable entity, or rejects an essential principle in our space-time theories. However, Bohmians could still argue that since displays other important advantages—no measurement problem, distinguishability, classical limit, etc.—then we should be willing to pay the price of admitting the quantum wave (or a rejection of the geodesic principle).[5]

Therefore, both in the case of the LPT vs. SR and in the case of BQT vs. SQM we have that non-empirical features can be invoked to rationally argue that one of the theories is superior to its rival. However, the rationality of such arguments presupposes some epistemological or ontological commitments that may not be universally accepted by the whole scientific community. Thus, both cases clearly illustrate what conclusion *ii)* states: theoretical features provide a partial solution to the problem of EE and UD. Even if the problem is present, rationally grounded reasons can be invoked in order to prefer one of the theories. However, those reasons may not be enough in order to ground a fully objective and uniquely determined choice.

[5] I do not consider the remaining three non-empirical features evaluated in 3.6 because we saw that they do not really work as criticisms of Bohm's theory that could be invoked in order to regard SQM as a better theory. In 3.6.5 we saw that the Heinsenberg-Pauli objection is founded on a misinterpretation of the physical meaning of the momentum-position symmetry in SQM. In 3.6.6 I argued that although there are interesting questions that arise in BQT with respect to the foundations of probability—about the justification of the quantum equilibrium hypothesis—those questions have to do more with the foundations of probability *in physics* than with a specific problem within BQT. Finally, in 3.6.8 I argued that although there are worries about the compatibility between BQT and SR, similar worries hold for SQM.

Now turning to conclusion *iii)*, we can establish that the analysis offered in chapter 2 clearly illustrates it. We saw that the intertheoretic connections between the LPT and early quantum physics, and between SR and GR, show that non-entailed empirical evidence can be used in order to break the UD at issue, even though the EE remains. In section 2.5.5 I explained that Lorentz's model of the electron, a central part of the LPT, was at deep odds with the quantum hypothesis. Lorentz himself showed that if such a model was assumed in attempts to derive the blackbody radiation law, the result would unavoidably be the Rayleigh-Jeans law, which was empirically wrong. On the other hand, there was no friction between SR and the quantum hypothesis. Since the former did not assume any specific model about the ultimate nature of matter, there was no conflict.

Now, if we recall the principle of confirmation that we extracted from Boyd (1973), we have that the evidence confirming the quantum hypothesis counts as indirect evidence against the LPT. Boyd's confirmation scheme considers two predictively equivalent theories T and T', such that the theoretical core of T' is at odds with theory P, which is empirically well-confirmed, whereas T and P are compatible. Our principle indicates that, given the conflict, the empirical evidence supporting P counts as evidence against T', but is neutral with respect to T. Therefore, in spite of the empirical equivalence, there is empirical evidence that disconfirms only T'. After the analysis in section 2.5.5, we can simply replace T for 'special relativity', T' for 'Lorentz's theory', and P for 'early quantum physics'. In other words, the case of Lorentz vs. Einstein can be decided in terms of non-entailed empirical evidence.

An analogous conclusion can be drawn from the results of section 2.5.6. There we saw that after Einstein's formulation of his gravitational theory, SR became a special case of GR: Minkowski space-time represents either a flat space-time devoid of matter, or an infinitesimal region of a curved space-time. On the other hand, given that the LPT postulates a Newtonian space-time, it is incompatible with GR. That is, unlike Lorentz's theory, SR can be encompassed by GR. We also saw that GR makes some observable predictions that SR cannot entail on its own, most notably the light-bending effect produced by the presence of a massive object that curves space-time.

Now, given that in our empirically equivalent pair only SR can be encompassed by GR, it follows that the empirical evidence given by measurements of the light-bending effect flows to Einstein's theory but not to Lorentz's, so that the evidential tie can be broken by non-consequential empirical evidence. To see this we only need to recall Laudan & Leplin's confirmation pattern: if we have two empirically equivalent hypotheses H_1 and H_2, such that H_1, but not H_2, can be embedded in a more general theory T, and if T is such that it entails the hypothesis H, which in turn implies the observational statement e; then the truth of e counts as empirical evidence for H, for T , and for H_1, but not for H_2. After the analysis provided in section 2.5.6, it is clear that H_1 can be replaced by 'SR', H_2 by 'the LPT', T by 'GR', H by 'light gravitates', and e by 'the results of Eddington's expedition'. Therefore, the theoretical interconnections described imply that, despite the predictive equivalence involved, SR has more evidential support than the LPT[6].

[6] Elie Zahar (1973) refers to the connection between special and general relativity as *the* criterion to decide the case of Einstein vs. Lorentz. However, this author does not argue in terms of non-entailed empirical evidence. Presupposing the correctness of Lakatos account of scientific progress in terms of increasing empirical and heuristic power of scientific research programs, he states that it was *Einstein's research program*, not SR, that defeated Lorentz's theory. The argument presupposes that Einstein's road from special to general relativity can be described in terms of a Lakatosian research program, of course. I think that this presupposition is highly dubious. First, Zahar states that one of the heuristic principles that define Einstein's research program was the generalization of the principle of relativity by means of the attainment of general covariance. However, recent historiography on the formulation of general relativity shows that the quest for general covariance was a constraint that Einstein originally considered, then abandoned because of certain difficulties he found and which he expressed in the famous *hole argument*, and that he finally resumed—but only to realize that the physical meaning he had originally assigned to general covariance was unjustified. These zigzags in the relevance of the principle of general covariance make it dubious that its role in the formulation of general relativity can be described by means of Lakatos' concept of a scientific research program. Second, Zahar also states that the successful explanation of the behavior of Mercury's perihelion counts as a 'novel prediction' that testifies the empirical and heuristic power of Einstein's research program. Zahar's concept of 'novel prediction' asserts that even if the predicted fact was already known at the time of the formulation of the theory, the successful prediction was not a specific goal in the formulation of the theory.

The inter-theoretic connections between early quantum physics and the LPT, and between SR and GR, break the evidential tie between SR and the LPT. That is, we have a clear instance of conclusion *iii)*: problematic situations in which EE leads to UD can be solved by the regular practice of science. In the case of BQT vs. SQM we find that a solution like this is not currently available. All of the issues analyzed in section 3.6 are of a non-empirical nature, and none of them can be used to break either the EE or the UD between the theories, they can only ground a rational-but-subjective preference. Thus, we can certainly consider the case of BQT vs. SQM as a true and unsolved instance of EE and UD—in which only a partial solution based on theoretical features is available. This evaluation constitutes, in turn, a clear illustration of conclusion *iv)*: that a solution of the problem of EE and UD is a contingent matter, science may develop in a way such that either the EE or the UD breaks down, but it may not. This contingency is also reflected in the fact that, although the current state of science is such that the EE and UD remain, future developments could change this situation. I will briefly mention three possible scenarios that could lead to a solution.

First, we can ponder about the (relativistic) extension of BQT and SQM to the quantum field case. There are field versions of both theories, and the EE remains in those extensions[7]. Now, one of the main goal in current theoretical physics is to find a fundamental theory that encompasses both GR and QFT-QM. If this goal is finally accomplished, it may turn out that the resulting encompassing theory is compatible with only one of the field extensions of the quantum theories involved in our case of EE. In other words, it is a possible scenario that a solution for the case between SQM and BQT could be found, and such that it is analogous to the solution that the formulation and confirmation of GR provided in the case of SR vs. the LPT. More generally, future development of science may yield inter-theoretic connections between our EE theories and new theories, and these interconnections may in turn result in non-entailed empirical evidence that could break the UD.

Second, let us recall that in section 3.6.3 we saw that the description that BQT provides for the particle-in-the-box thought experiment proposed by Einstein is very different from the account that (any interpretation of) SQM offers. According to Bohm's theory, since the quantum 'wave' that corresponds to the particle is a standing wave ($\nabla S = 0$), then the particle is at rest within the box (for $v = \nabla S/m$). We know that the EE between SQM and BQT is guaranteed by the location-distribution postulate $P = |\Psi|^2$. In the case of Einstein's thought experiment we have that, according to SQM, a momentum measurement would yield the equiprobable results $\pm \hbar n\pi/L$—the particle would be found moving back or forth with equal probability. In BQT the momentum of the particle in the box is 0, but the prediction for the result of momentum measurement is the same as in SQM, for such a measurement would imply a disturbance of the quantum potential (by removing one of the walls, for example) such that the particle would be set in motion with a momentum $\pm \hbar n\pi/L$—with the direction depending on the specific initial position of the particle, that we do not know, for we cannot go further than the location distribution postulate.

The important point in this context is that the predictive equivalence guarantee given by the quantum equilibrium postulate presupposes that all (quantum) measurements are, in the end, position measurements. This assumption is highly reasonable of course, but I do not see any reason why it should be

---

Again, historiography shows that the correct prediction regarding Mercury's perihelion was an explicit and specific goal that Einstein considered in the formulation of the theory (see Earman and Janssen 1993). More generally, even though there is a continuity in Einstein's road from special to general relativity, it is quite difficult to make sense of it in terms of Lakatos' model. The non-entailed evidence standpoint, on the other hand, does not require to accept any specific model for the rationality and development of science. Non-entailed evidence settles the matter independently of whether one is a Lakatosian, a Kuhnian, or what have you.

[7] Whereas QFT, the field extension of SQM, is a Lorentz-covariant theory—as long as measurements are not considered—the Bohmian version is not. However, the non-covariant equations of BFT yield *covariant predictions* when the quantum equilibrium hypothesis enters the picture. For a simple sketch of BFT see (Cushing 1996), for detailed treatment see (Holland 1993, Ch. 12).

elevated to an *a priori* principle. That is, there may be a way to (perhaps indirectly) measure the momentum of the particle within the box that does not involve a measurement of its position—in a way such that the walls need not to be removed, for example. If we recall the description of the particle as at rest holds even in a macroscopic context (a particle with diameter 1 mm in a box with walls 1 m apart), this hypothetical possibility becomes even more interesting. Actually, we saw in section 1.4.1 that Laudan & Leplin compellingly argue that the EE between scientific theories is a time-indexed condition that may be broken by *a)* the introduction of new methods or instruments of observation, or by *b)* the availability of new auxiliary hypotheses (or the rejection of old ones) that change (enhance or reduce) the class of observable consequences of one of the theories in the EE pair. The hypothetical possibility just mentioned could work as an EE-breaker of the type *a)*. If we had a method to determine the momentum of the system in the box without opening it, we could have a crucial experiment in which BQT predicts the value zero, whereas SQM predicts $\pm \hbar n\pi/L$.

The third issue I will consider could also be the source for the breakdown of the evidential tie between the theories. A remarkable prediction of quantum physics is that, given an ensemble of particles that hit or 'climb' a barrier whose potential value is greater than the kinetic energy of the particles, some of them will be reflected by the barrier and some of them will be transmitted[8]. For experiments on an ensemble of particles that correspond to the same Ψ, both SQM and BQT make the same statistical predictions concerning the relative frequency of particles that bounce off and particles that cross the barrier. However, if we consider what the theories have to say about the *time* that the particles spend inside the barrier before being reflected or transmitted—the so called 'tunneling-time'—diverging predictions may, in principle, result. Consider a barrier with a potential V_0 and thickness d. A beam of N identically prepared particles (or, in Bohmian words, an ensemble of N particles associated to the same Ψ) is incident on the barrier from left to right, and we set detectors 1 and 2 at x_1 (somewhere to the left of the barrier) and x_2 (somewhere to the right of the barrier), respectively. We know that some of the particles will go through the barrier and some will be reflected. Suppose that we knew the time $t_j^{(1)}$ at which the jth particle passed x_1, and either the time $t_j^{(2)}$ at which it passed x_2 if it was transmitted, or the time $t_j^{(3)}$ at which it passed x_1 again if it was reflected. We could then define the time $\tau_j = t_j^{(2)} - t_j^{(1)}$ (or $\tau_j = t_j^{(3)} - t_j^{(1)}$) that each particle spent in the region comprehended between x_1 and x_2, and we could also define a *dwell time* $\tau_D = \frac{1}{N}\sum_{j=1}^{N}\tau_j$, that is, the average time spent by the particles $j = 1,2,\dots,N$ in the region (x_2, x_1). Similarly, we could define an average *reflection time* $\tau_R = \frac{1}{N_R}\sum_{\{N_R\}}\tau_j$ for N_R reflected particles, and an average *transmission time* $\tau_T = \frac{1}{N_T}\sum_{\{N_T\}}\tau_j$ for the N_T transmitted particles.

Now, in SQM the expression $\tau_D(x_1, x_2) = \int_0^\infty dt \int_{x_1}^{x_2} dx|\Psi(x,t)|^2$ can be derived[9], so a prediction for the average dwell time can, in principle, be obtained. However, in SQM the trajectory of a particle is not a well-defined concept, and the probability density $|\Psi(x,t)|^2$ inside the barrier cannot be divided into 'to be transmitted' and 'to be reflected' components, so that there is no (fully consistent) way to determine τ_R and τ_T[10]. Though SQM can determine the average time between the emission of the particles in the beam and the position measurements in x_1 and x_2, to speak of the times τ_R and τ_T presupposes reflection-*trajectories* and transmission-*trajectories*, so these terms cannot be defined. On the other hand, in BQT the trajectory of a particle is well-defined in all contexts, and it depends only on its initial position. Thus, the

[8] This *quantum barrier tunneling effect* is yet another surprising feature of the quantum world. It is as if a ball climbing a hill could reach to the top and continue to the opposite ladder, even if its kinetic energy is not enough to defeat the corresponding gravitational pull. The following treatment of tunneling times in BQT is based on (Cushing 1995).

[9] See (Leavens & Aers 1993, 107-8).

[10] This is a rather technical issue, but attempts to derive transmission and reflection average times within SQM lead either to ambiguous results or to negative times. For a detailed critical analysis of such attempts, see (Leavens & Aers 1993) and (Leavens 1996).

time τ corresponding to the trajectory of a transmitted or a reflected particle is a function of its initial position: $\tau = \tau(x_0)$; so a derivation of the expressions for the average transmission and reflection times is straightforward. In Bohm's theory, given the different initial positions x_0 distributed according to $|\Psi|^2$, we have that $\tau_D = \int \tau\,(x_0)|\Psi(x_0)|^2 dx_0$, $\tau_R = \frac{1}{R}\int_{\{R\}}\tau(x_0)|\Psi(x_0)|^2 dx_0$, and $\tau_T = \frac{1}{T}\int_{\{T\}}\tau(x_0)|\Psi(x_0)|^2 dx_0$. This is simply an instance of the Bohmian account of measurements. The measurement interaction 'splits' the quantum wave in different channels that are associated to possible experimental results, whereas which specific outcome obtains depends on which of those channels the particle actually occupies, and, in turn, this depends on x_0. In the formulas above, $R = \int_{\{R\}}|\Psi(x_0)|^2 dx_0$, and $T = \int_{\{T\}}|\Psi(x_0)|^2 dx_0$, where $\{R\}$ and $\{T\}$ represent those subsets of initial positions that determine that the corresponding particle will occupy the reflection and transmission channels, respectively[11].

Thus, we can have a source for a possible breakdown of the evidential tie between SQM and BQT. As Cushing puts it, 'if it should turn out that the predictions of Bohmian mechanics for the τs can be compared with experiment and are found to be in error, then that would count as a refuting instance, a failure of Bohm's program. Copenhagen [SQM] would be neither supported nor refuted here, since it is unable to make any unambiguous predictions. Bohm's program is clearly at more risk here (i.e., falsifiable)' (1995, 272). However, this is only an *in principle* line of thought, for the available data yielded by quantum tunneling experiments performed so far cannot be interpreted in a way such that transmission or reflection tunneling times become observationally revealed (see ibid, 274).

These experiments, though, were not designed with the goal of measuring tunneling times, so we may speculate about the possibility of a direct tunneling time measurement. Cushing has outlined an experiment like this. Current technology allows the fabrication of a barrier of thickness $d \approx 10^{-9}m$. We then prepare a beam of electrons and direct it to the barrier. We cannot really put a detector at x_1 to determine the time $t_1$when the particle passes that point, for it would 'collapse' the wave function and the incident electrons would be useless. Thus, we should include a state-preparation device at x_1 that allows only one electron at a time to go through, and keep detector 1 switched off. If Δx is the width of the wave-packet prepared at x_1, we could determine a time $t_1 \approx 0 \pm \Delta x/v_0$, where v_0 is the velocity of the incident packet. In principle, the margin of error $\Delta x/v_0$ can be made observationally negligible (for large v_0). After the state has been prepared by the device, we turn on the detectors at x_1 and x_2 to determine the instant t_2 when the electron passes x_2 if it is transmitted, or passes x_1 if it is reflected. Now, the distance from x_1 to the left 'wall' of the barrier must be a few times Δx so that the initial wave packet is completed before appreciable interference with the barrier begins, and x_2 must be far enough from the right wall so that the detector is out of the interference region. Besides, if we want a fairly defined energy, the width of Δx must be several times d. Give this setup, the distance $(x_2 - x_1)$ would be 50-100d, so that the free transit time of a transmitted particle would be of the order $100d/v_0$. Thus, in order to be observationally detectable, the tunneling time inside the barrier of width d for a transmitted particle would have to be nearly an order of magnitude greater than d/v_0. It is clear that if the experiment is to work as an evidential tie-break, the wave packet prepared must be such that $T \approx R \approx 1/2$ or such that $T \ll 1$ (if $T \approx 1$, then $\tau_T \approx \tau_D$, so that the predictions of SQM and BQT would completely agree).

This experimental sketch is illustrative in that there are no *in principle* considerations that preclude observations of transmission tunneling times. That is, the situation is not like an attempt to determine

[11] For a detailed and technical derivation of transmission and reflection times in BQT see (Leavens & Aers 1993) and (Leavens 1996). Interestingly, Leavens & Aers evaluate that transmission and reflection times can be clearly determined by BQT as an expression of the clear ontological picture the theory offers, and as a non-empirical virtue to prefer Bohm's proposal over SQM: 'the beauty of the Bohm interpretation, in this context, is that the concepts of transmission and reflection times follow directly from the basic postulates and consequently one can talk about them in an internally consistent way. Furthermore, due to the nature of these postulates, one can use the language of classical mechanics without apology [...]. for reasons such as these, which are largely ones of taste, we prefer the Bohm trajectory solution of the tunneling time problem' (1993, 138).

the initial position of a particle without disturbing its wave-function. However, when the numerical issues are considered, it becomes clear that the practical feasibility of such an experiment is not possible with the current technology and experimental methods. For wave packets such that $T \approx 1/2$ (a significant part of these packets would enter the barrier), the values of the parameters involved (the mass of the electron, $\hbar$, d, V_0) imply that the velocity of the incident electrons would be of about 10^5-$10^6\, m/s$, so that the transit time (the time the particle spends inside the barrier) would be of the order of $t \approx d/v_0 \approx 10^{-16}$-$10^{-14} s$. However, the border of the doable for a direct measurement of time is given by a resolution of $10^{-12} s$. Furthermore, calculations of the time τ_T that the particle spends in the $(x_2 - x_1)$ region yield values of the order $10^{-15} s$, that is, the time τ_T for a run of the experiment with the barrier removed would scarcely differ from a run in which the barrier is present, so that the sought effect would not be detectable. Using wave packets such that $T \ll 1$, we could obtain times which are distinguishable when barrier-in and barrier-out comparisons are made (at the cost of a great loss in intensity of the incident beam and detectability of the electrons). However, the τ_Ts obtained are of the order $10^{-14} s$, still far away from the current limits of observability.

It is important to underscore that the way in which a tunneling time experiment like this could break the evidential tie between the theories we are considering is rather subtle. Assume that the experiment is performed and the predictions of BQT for τ_T are confirmed. Would this count as a full victory of this theory over SQM? Not necessarily. If the predicted values for τ_T (and τ_R) are actually obtained in the experiment, that means that the predicted value for τ_D that *both* theories predict is confirmed. Now, since supporters of SQM do not accept that 'particle-trajectory' is a well-defined concept, they would simply deny that the experiment counts as an observation of τ_T (and τ_R)—after all, theoretical frameworks determine what is observed. Thus, whether the EE is removed is a complex issue: Bohmians would say that the mere prediction of well-defined values for τ_T and τ_R does imply an EE removal; whereas supporters of SQM, if convenient, may (rationally) argue that it does not. They could deny that τ_T and τ_R as obtained from BQT can correspond to any observational data, and that, as long as the predictions for τ_D are confirmed, SQM will be evidentially safe[12]. However, with an experimental outcome against BQT, the predictions and the corresponding experiment could indeed break the UD in a fully objective way. If the predictions of BQT for τ_T (and τ_R) are *refuted* by the experiment, then the case can be decided (granted that the predicted values for τ_D are not experimentally disproved, of course). If all the technological and theoretical specifications of the experiment are fully clear and accepted by both sides, a result like this would certainly break the evidential tie in favor of SQM. That is, a time tunneling measurement could work as a semi-crucial experiment: some of its possible outcomes can settle the case, some others cannot.

Bohm formulated his theory in 1952. Sixty two years later, the development of science has not yet provided a way to remove the EE or the UD between SQM and BQT, so we have a clear instance of persistent UD. However, I have outlined three possible ways in which the future development of science may allow an evidence-based decision between the theories. Just as conclusion *iv)* asserts, that a solution for a problematic case of EE will be found is a contingent matter. If future developments in, say, string theory, are such that it becomes clear that only one of the theories involved in the problem we are treating gets encompassed by it, and if the theory gets empirically confirmed, we could solve our problem using non-entailed empirical evidence, even if the EE remains. On the other hand, if we could measure the

[12] As Leavens & Aers point out, 'given sufficiently accurate measured values of $|\Psi(x,t)|^2$ for the scattering problem of interest, the mean Bohm trajectory transmission and reflection times [τ_T and τ_R] are readily obtained by carrying out the integrations appearing in $\left[\tau_T = \frac{1}{T}\int_{\{R\}}\tau(x_0)|\Psi(x_0)|^2 dx_0 \text{ and } \tau_R = \frac{1}{R}\int_{\{R\}}\tau(x_0)|\Psi(x_0)|^2 dx_0\right]$. However, the physical meaning attached to $\left[\tau_T = \frac{1}{T}\int_{\{R\}}\tau(x_0)|\Psi(x_0)|^2 dx_0 \text{ and } \tau_R = \frac{1}{R}\int_{\{R\}}\tau(x_0)|\Psi(x_0)|^2 dx_0\right]$ is based on a set of postulates unique to the Bohm interpretation of quantum mechanics. Hence, to a proponent of any conventional interpretation, there is no reason to identify the experimental 'times' based on these equations with actual mean transmission and reflection times' (1993, 136-7).

momentum of the particle in Einstein's box without removing its walls, or future developments of experimental physics are such that a direct measurement of quantum tunneling times becomes possible, (semi)crucial experiments could be set and an evidentially based decision may be allowed. But we do not know if science will develop along this lines—the problem may remain.

These remarks also illustrate what conclusion *v)* tells us: EE as leading to UD is mainly a *scientific* problem. Only the work of theoretical and/or experimental physicists can result in a way out of the UD that stands between BQT and SQM. It is true that, just as many other scientific problems, the problem of EE and UD has a clear philosophical import. For example, the arguments based on non-empirical features supporting each of the theories in the case-studies here offered certainly presuppose and motivate epistemological and ontological views. However, the analysis offered in chapter 1 shows that the philosophical terror that the problem seemed to imply is not founded. EE and UD is not a *universal* feature in science, and it does not jeopardize the *rationality* of theory choice. The problem we are treating is not a problem that threatens (the foundations of) science as a whole. It is true that stubborn (local) cases of EE and UD are such that a unique and fully objective decision cannot be done; but, again, this is mainly a scientific problem: we have two good rival theories for a certain realm of phenomena, but let us (*physicists* would say) try to pick *the* right one. If things go well, even the most radical supporters of the defeated theory should have to accept that the rival theory must be accepted—this is just what happened to supporters of the LPT, of course. If things do not go well, we should keep on waiting, and scientists will certainly keep on working, and, in the meantime, we can (rationally) justify a preference based on the non-empirical features we praise according to our epistemological and ontological creeds—and this is the situation in the case of SQM vs. BQT.

In short, we have that the statements *i)* to *v)* above provide a precise description and a correct evaluation of the problem of EE and UD. Besides, and in the light of these statements, we can accurately evaluate the cases of SR vs. the LPT and of SQM vs. BQT. The UD in the former case-study was soon dissolved by inter-theoretic connections. In the latter case, we have a persistent example of UD (that can be partially dealt with by means of non-empirical features), but for which we can rationally hope that the future development of science will provide a solution.

BIBLIOGRAPHY

Acuña, P. (2014a). "Artificial examples of empirical equivalence". To appear in Galavotti et al (eds.), *New Directions in the Philosophy of Science*. Springer.

Acuña, P. (2014b). "On the empirical equivalence between special relativity and Lorentz's ether theory". *Studies in History and Philosophy of Modern Physics* **46**: 283-302.

Acuña, P. & Dieks, D. (2014). "Another look at empirical equivalence and underdetermination of theory choice". *European Journal for Philosophy of Science* **4**: 153-180.

Adlam, E. (2011). "Poincaré and special relativity", http://arxiv.org/pdf/1112.3175.pdf.

Albert, D. (1992). *Quantum Mechanics and Experience.* Cambridge, Ma: Harvard University Press.

Albert, D. (1996). "Elementary quantum metaphysics", in Cushing, J., Fine. A & Goldstein, S. (eds.). *Bohmian Mechanics and Quantum Theory: an appraisal*. Springer, 277-84.

Albert, D. & Loewer, (1988). "Interpreting the many-worlds interpretation". *Synthese* **77**: 195-213.

Anandan, J. & Brown. H. (1995). "On the reality of space-time geometry and the wavefunction". *Foundations of Physics* **25**: 349-60.

Bacciagaluppi, G. (1998). "Bohm-Bell dynamics in the modal interpretation", in Dieks, D, Vermaas, P. (eds.), *TheModal Interpretation of Quantum Mechanics.* Dordrecht: Kluwer Academic Publishers, 177–211.

Bacciagaluppi, G. (2012). "The role of decoherence in quantum mechanics". *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/win2012/entries/qm-decoherence/.

Baggott, J. (2004). *Beyond Measure: modern physics, philosophy and the meaning of quantum theory*. Oxford University Press.

Balashov, Y. & Janssen, M. (2003). "Presentism and relativity". *British Journal for the Philosophy of Science* **54**: 327-46.

Bangu, S. (2006). "Underdetermination and the argument from indirect confirmation". *Ratio* **19**: 269-77.

Barrett, J. (1999). *The Quantum Mechanics of Minds and Worlds*. Oxford University Press.

Barrett, J. (2011). "Everett's relative state formulation of quantum mechanics". *Stanford Encyclopedia of Philosophy*, http://staford.edu/archives/spr2011/entries/qm-everett/.

Belinfante. F. (1973). *A Survey of Hidden-Variables Theories*. Pergamon Press.

Bell, J. S. (1964). "On the Einstein Podolsky Rosen paradox". *Physics* **1**: 195-200.

Bell, J. S. (1966). "On the problem of hidden variables in quantum mechanics". *Reviews of Modern Physics* **38**: 447-52.

Bell, J. S. (1982). "On the impossible pilot wave". *Foundations of Physics* **12**: 989-99.

Bell, J. S. (1987). "How to teach special relativity", in *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press, 67–80.

Bell, J. S. (1990). "La nouvelle cuisine", in *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press, 232-50.

Bell, J. S. (2004). "Are there quantum jumps?", in *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press, 201-12.

Bene, G., & Dieks, D. (2002). "A perspectival version of the modal interpretation of quantum mechanics and the origin of macroscopic behavior". *Foundations of Physics* **32**: 645-71.

Berndl, K., Dürr, D., Goldstein S., & Zanghi, N. (1996) "Non-locality, Lorentz-invariance, and Bohmian quantum theory". *Physical Review A* **53**: 2062-73.

Bergström, L. (1993). "Quine, underdetermination and skepticism". *The Journal of Philosophy* **90**: 331-58.

Bird, A. (2007). "Underdetermination and evidence". Chapter V of Monton, B. (ed.) *Images of Empiricism: essays on science and stances, with a reply from Bas C. van Fraassen*. Oxford University Press.

Bohm, D. (1952). "A suggested interpretation of the quantum theory in terms of "hidden" variables" I & II. *Physical Review* **85**: 166-93.

Bohm, D. (1953). "A discussion of certain remarks by Einstein on Born's probability interpretation of the ψ-function". In *Scientific Papers Presented to Max Born*, New York: Hafner, 13–9.

Bohm, D. (1953). "Proof that probability density approaches $|\Psi|^2$ in causal interpretation of quantum theory". *Physical Review* **89**: 458-66.

Bohm, D. & Hiley, B. J. (1993). *The Undivided Universe: an ontological interpretation of quantum theory*. Routledge.

Bohr, N. (1934). *Atomic Theory and the Description of Nature*, reprinted as *The Philosophical Writings of Niels Bohr*, Vol. I, Woodbridge: Ox Bow Press.

Bohr, N. (1958). *Essays 1932-1957 on Atomic Physics and Human Knowledge*, reprinted as *The Philosophical Writings of Niels Bohr*, Vol. II, Woodbridge: Ox Bow Press.

Bohr, N. (1963). *Essays 1958-1962 on Atomic Physics and Human Knowledge*, reprinted as *The Philosophical Writings of Niels Bohr*, Vol. III, Woodbridge: Ox Bow Press

Bonk, T. (2008). *Underdetermination: an essay on evidence and the limits of natural knowledge*. Dordrecht: Springer.

Boyd, R. (1973). "Realism, underdetermination, and a causal theory of evidence". *Noûs* **7**: 1-12.

Brown, H. (1986). "The insolubility proof of the quantum measurement problem". *Foundations of Physics* **16**: 857-70.

Brown, H. (2001). "The origins of length contraction: I. The FitzGerald-Lorentz deformation hypothesis". *American Journal of Physics* **69**: 1044-54.

Brown, H. (2003). "Michelson, FitzGerald and Lorentz: the origins of special relativity revisited". *Bulletin de la Société des Sciences et des Lettres de Lódz, Volume LIII; Série: Reserches sur les Déformations*, Volume XXXIX; 23-35.

Brown, H. (2005). *Physical Relativity: space-time structure from a dynamical perspective*. Oxford University Press.

Brown, H., Dewdney, C. & Horton, G. (1995). "Bohm particles and their detection in the light of neutron interferometry". *Foundations of Physics* **25**: 329-48.

Brown, H. & Pooley, O. (2000). "The origin of the spacetime metric: Bell's 'Lorentzian pedagogy' and its significance in general relativity", in Callender, C. & Huggett, N. (eds.), *Physics Meets Philosophy at the Planck's Scale*. Cambridge University Press, 2000.

Brown, H. & Pooley, O. (2006). "Minkowski space-time: a glorious non-entity"; in *The Ontology of Spacetime*, Dennis Dieks (ed.), Elsevier B. V. (2006), 67-89.

Brown, H. & Wallace, D. (2005). "Solving the measurement problem: de Broglie-Bohm loses out to Everett". *Foundations of Physics* **35**: 517-40.

Brush, S. (1999). "Why was relativity accepted?" *Physics in Perspective* **1**: 184-214.

Bub, J. (1969). "What is a hidden variable theory of quantum phenomena?" *International Journal of Theoretical Physics* **2**: 101-23.

Bub, J. (2010). "Von Neumann's 'no hidden variables' proof: a re-appraisal". *Foundations of Physics* **40**: 1333-40

Bub, J. (1997). *Interpreting the Quantum World*. Cambridge University Press.

Bunge, M. (1961). "The weight of simplicity in the construction and assaying of scientific theories". *Philosophy of Science* **28**: 120-41.

Busch, J. (2009). "Underdetermination and rational choice of theories". *Philosophia* **37**: 55-65.

Butterfield, J. (1992). "Bell's theorem: what it takes". *British Journal for the Philosophy of Science* **43**: 41-83

Callender, C. (2007). "The emergence and interpretation of probability in Bohmian mechanics". *Studies in History and Philosophy of Modern Physics* **38**: 351-70.

Callender, C. & Weingard, R. (1997). "Trouble in paradise: problems for Bohm's theory". *The Monist* **80**: 24-43.

Carrier, M. (2011). "Underdetermination as an epistemological test tube: expounding hidden values of the scientific community". *Synthese* **180**: 189-204.

Clauser, J. F., Horne, M. A., Shimony, A., & Holt, R. A. (1969). "Proposed experiment to test local hidden-variable theories". *Physical Review Letters* **26**: 880–4.

Clendinnen, F. J. (1989). "Realism and the underdetermination of theory". *Synthese* **81**: 63-90.

Cushing, J. (1991). "Quantum theory and explanatory discourse: endgame for understanding?" *Philosophy of Science* **58**: 337-58.

Cushing, J. (1993). "Bohm's theory: common sense dismissed". *Studies in History and Philosophy of Science* **24**: 815-42.

Cushing, J. (1994). *Quantum Mechanics: historical contingency and the Copenhagen hegemony*. Chicago: The University of Chicago Press.

Cushing, J. (1995). "Quantum tunneling times: a crucial test for the causal program?" *Foundations of Physics* **25**: 269-80.

Cushing, J. (1996). "What measurement problem?", in R. Clifton (ed), *Perspectives on Quantum Reality: non-relativistic, relativistic and field theoretic*. Springer, 167-82.

Cushing, J., Fine. A & Goldstein, S. (eds.) (1996). *Bohmian Mechanics and Quantum Theory: an appraisal*. Springer.

Cuvaj, C. (1968). "Henri Poincaré's mathematical contributions to relativity and the Poincaré stresses". *American Journal of Physics* **36**: 1102-13.

Darrigol, O. (1994). "The electron theories of Larmor and Lorentz: a comparative study". *Historical Studies in the Physical and Biological Sciences* **24**: 265–336.

Darrigol, O. (1995). "Henri Poincaré's criticism of *Fin de Siècle* electrodynamics". *Studies in History and Philosophy of Modern Physics* **26**: 1-44.

Darrigol, O. (1996). "The electrodynamic origins of relativity theory". *Historical Studies in the Physical Sciences* **26**: 241-312.

Darrigol, O. (2005). "The genesis of the theory of relativity", in T. Damour, O. Darrigol, B. Duplantier, V. Rivasseau (eds.), *Einstein 1905-2005: Poincaré seminar 2005.* Basel: Birkhäuser, 2006, 1-31.

D'Espagnat, B. (1999). *Conceptual Foundations of Quantum Mechanics*. Perseus Books Publishing.

De Regt, H. W. and Dieks, D. (2005). "A contextual approach to scientific understanding". *Synthese* **144**: 137-70.

De Witt, B. S. & Graham, N. (eds) (1973). *The Many Worlds Interpretation of Quantum Mechanics*. Princeton University Press.

Deutsch, (1996). "Comment on Lockwood". *British Journal for the Philosophy of Science* **47**: 222-8.

Dewdney, C. (1985). "Particle trajectories and interference in a time-dependent model of neutron single crystal interferometry". *Physics Letters A* **109**: 377-84.

Dewdney, C., Holland. P & Kyprianidis A. (1986). "What happens in a spin measurement?" *Physics Letters A* **119**: 259-67.

Dewdney, C., Holland. P & Kyprianidis A. (1987). "A causal account of non-local Einstein-Podolsky-Rosen spin correlations". *Journal of Physics A* **20**: 4717-32.

Dieks, D. (1989). "Quantum mechanics without the projection postulate and its realistic interpretation". *Foundations of Physics* **19**: 1397-1423.

Dieks, D. (1998). ''Locality and Covariance in the Modal Interpretation of Quantum Mechanics''. In Dieks, D. & Vermaas, P (eds.) *The Modal Interpretation of Quantum Mechanics*, Dordrecht: Kluwer Academic publisher, 49–67.

Dieks, D. (1990). "Quantum statistics, identical particles, and correlations". *Synthese* **82**: 127-55

Dieks, D. (2005). "Quantum mechanics: an intelligible description of objective reality?" *Foundations of Physics* **35**: 399-415.

Dieks, D (2009). "Bottom-up versus top-down: the plurality of explanation and understanding in physics", in H. de Regt, S. Leonelli, K. Eigner (eds), *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

Dieks, D. (2010). "Are 'identical quantum particles' weakly discernible objects?", in Suárez, M., Dorato, M. & Rédei, M. (eds.), *Philosophical issues in the Sciences*, Springer, 21-30.

Dieks, D. & Lubberdink, A. (2011). "How classical particles emerge from the quantum world". *Foundations of Physics* **41**: 1051-64.

Dieks, D. & Veerstegh, M. (2008). "Identical particles and weak discernibility". *Foundations of Physics* **38**: 923-34.

Dorling, J. (1968). "Length contraction and clock synchronization: the empirical equivalence of Einsteinian and Lorentzian theories". *The British Journal for the Philosophy of Science* **19**: 67-9.

Douven, I. (2000). "The anti-realist argument for underdetermination". *The Philosophical Quarterly* **50**: 371-5.

Duhem, P. (1906). "Physical theory and experiment" (chapter VI of *The Aim and Structure of Physical Theory*), in Harding, Sandra (ed.) (1976).

Dürr. D., Goldstein, S, & Zanghi, N. (1995). "Bohmian mechanics and the meaning of the wave function", arXiv:quant-ph/9512031.

Earman, J. (1989). *World Enough and Spacetime: absolute versus relationist theories of space and time*. Cambridge: MIT Press.

Earman, J. & Janssen, M. (1993). "Einstein's explanation of the motion of Mercury's perihelion". In J. Earman et al. (Eds.), *The Attraction of Gravitation*, Boston: Birkhouser, 129-72.

Einstein, A. (1905a). "Zur Elektrodynamik bewegter Körper". *Annalen der Physik* **17**: 891-921. Reprinted in Lorentz et al. 1952.

Einstein, A. (1905b). "Ist die Trägheit eines Körpers von seinem Energieinhalt abhänging?" *Annalen der Physik* **18**: 639–641. Reprinted in Lorentz et al. 1952.

Einstein, A. (1906). "Das Prinzip von der Erhaltung der Schwerpunktsbewegung und die Trägheit der Energie". *Annalen der Physik* **20**: 627–633. Reprinted in Stachel, J., Cassidy, D., Renn, J., and Schulmann, R. (eds.). *The Collected Papers of Albert Einstein, Vol. 2, The Swiss years: writings, 1900-1909*. Princeton: Princeton University Press, 1989.

Einstein, A. (1919), "My theory", *The London Times*, November 28, 13. Reprinted as "What is the theory of relativity?" in Einstein (1954), 227-32.

Einstein, A. (1920). *Äther und Relativitätstheorie. Rede gehalten am 5. Mai 1920 an der Reichs-Universität zu Leiden*. Berlin: Springer. English translation: "Ether and the theory of relativity". In G. B. Jeffery and W. Perrett (transl.), *Sidelights on Relativity*. London: Methuen; New York: E. P. Dutton, 1922, 1–24.

Einstein, A. (1954), *Ideas and Opinions*. New York: Crown Publishers.

Einstein, A. (1949). "Autobiographical notes", in Schilpp 1949, 1–95.

Einstein, A., Podolski, B. & Rosen, N. (1935) "Can quantum mechanical description of physical reality be considered complete?" *Physics Review* **47**: 777-80.

Esfeld, M., Lazarovici D., Hubert, M. & Dürr, D. (2013). "The ontology of Bohmian mechanics". *British Journal for the Philosophy of Science*, doi:10.1093/bjps/axt019.

Everett, H., III. (1957). "'Relative state formulation of quantum mechanics". *Reviews of Modern Physics* **19**: 454-62.

Faye, J. (2009). "Copenhagen interpretation of quantum mechanics". *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/spr2009/entries/qm-copenhagen/.

Fine, A. (1986). *The Shaky Game: Einstein realism and the quantum theory*. University of Chicago Press.

French, S. (2011a). "Identity and individuality in quantum theory". *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/sum2011/entries/qt-idind/.

French, S. (2011b). "Metaphysical underdetermination: why worry?" *Synthese* **180**: 205-221.

French, S. & Krause, D. (2006). *Identity in Physics: a historical, philosophical and formal analysis*. Oxford University Press.

Frisch, M. (2005). "Mechanisms, principles, and Lorentz's cautious realism". *Studies in History and Philosophy of Modern Physics* **36**: 659-79.

Frisch, M. (2011). "Principle or constructive relativity". *Studies in History and Philosophy of Modern Physics* **42**: 172-183.

Galison, P. L. (1979). "Minkowski's space-time: from visual thinking to the absolute world". *Historical Studies in the Physical Sciences* **10**: 85–121.

Gearhart, C. (2002). "Planck, the quantum, and the historians". *Physics in Perspective* **4**: 170-215.

Ghirardi, G. (2011). "Collapse theories". Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/qm-collapse/.

Giedymin, J. (1982). *Science and Convention: essays on Henri Poincaré's philosophy of science and the conventionalist tradition*. Pergamon Press.

Goldberg, S. (1967). "Henri Poincaré and Einstein's theory of relativity". *American Journal of Physics* **35**: 934–944.

Goldstein S., & Zanghi, N. (2013). "Reality and the role of the wavefunction in quantum theory", in A. Ney & D. Albert (eds.) *The Wavefunction: essays on the metaphysics of quantum mechanics*. Oxford: Oxford University Press, 91-109.

Goldberg, S. (1969). "The Lorentz theory of electrons and Einstein's theory of relativity". *American Journal of Physics* **37**: 498–513.

Greenberger, D., Horne, M & Zeilinger, A. (1989). "Going beyond Bell's Theorem". In Kafatos, M. (ed.) *Bell's Theorem, Quantum Theory and Conceptions of the Universe*. Springer, 69-72.

Grünbaum, A. (1959). "The falsifiability of the Lorentz-FitzGerald contraction hypothesis". *The British Journal for the Philosophy of Science* **37**: 48–50.

Grünbaum, A. (1960). "The Duhemian argument". *Philosophy of Science* **27**. Reprinted in Harding, Sandra (ed.) 1976.

Grünbaum, A. (1973). *Philosophical Problems of Space and Time*. Dordrecht: Reidel.

Grünbaum, A. (1976). "*Ad hoc* auxiliary hypotheses and falsificationism". *The British Journal for the Philosophy of Science* **27**: 329–362.

Halliwell, J. J. (1995). "A review of the decoherent histories approach to quantum mechanics", http://arxiv.org/pdf/gr-qc/9407040.pdf

Harding, S. (1976). *Can Theories be Refuted? Essays on the Duhem-Quine thesis.* Dordrecht: Reidel.

Hardy, L. (1992). "Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories". *Physical Review Letters* **68**: 2981-4.

Harman, P. M. (1982). *Energy, Force and Matter: the conceptual development of nineteenth century physics.* Oxford University Press.

Healey, R. (1989). *The Philosophy of Quantum Mechanics: an interactive interpretation*. Cambridge University Press.

Heilbron, J., & Kuhn, T. (1969). "The genesis of the Bohr atom". *Historical Studies in the Physical Sciences* **1**: 211-90.

Heisenberg, W. (1955). "The development of the interpretation of the quantum theory". In Pauli, W. (ed.) *Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of his Seventieth Birthday*, New York: McGraw-Hill, 12–29.

Heisenberg, W. (1958). *Physics and Philosophy*. New York: Harper & Row.

Hempel, C. (1945). "Studies in the Logic of Confirmation (II)" *Mind* **54**: 97-121.
Hermann, A. (1971). *The Genesis of Quantum Theory (1899-1913)*. Cambridge, Massachusetts; London: MIT Press.

Hiley, B. J. & David Peat, F. (eds.) (1987). *Quantum Implications: essays in honor of David Bohm*. London & New York: Routledge.

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.

Hoefer, C. & Rosenberg, A. (1994). "Empirical Equivalence, Underdetermination, and Systems of the World. *Philosophy of Science* **61**: 592-607.

Holland, P. (1988). "Causal interpretation of a system of two spin-1/2 particles". *Physics Reports* **169**: 293-327.

Holland. P. (1993). *The Quantum Theory of Motion: an account of the de Broglie-Bohm causal interpretation of quantum mechanics*. Cambridge University Press.

Holton, G. (1969). "Einstein, Michelson and the "crucial" experiment". *Isis* **60**: 133–197. Reprinted in Holton 1988.

Holton, G. (1988) [1973]. *Thematic Origins of Scientific Thought: Kepler to Einstein*. Cambridge: Harvard University Press.

Hughes, R. I. G. (1989). *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press.

Hund, F. (1974). *The History of Quantum Theory*. London: Harrap & Co.

Hunt, B. (1988). "The origins of the Fitzgerald contraction". *The British Journal for the History of Science* **21**: 67-76.

Illy, J. (1981). "Revolutions in a revolution". *Studies in History and Philosophy of Science* **12**: 173-210.

Illy, J. (1989). "Einstein teaches Lorentz, Lorentz teaches Einstein: their collaboration in general relativity, 1913–1920". *Archive for History of the Exact Sciences* **39**: 247-288.

Isaacson, W. (2007). *Einstein: his life and universe*. New York: Simon & Schuster.

Ives, H. (1952). "Derivation of the mass-energy relation". *Journal of the Optical Society of America* **42**: 540-543

Jammer, M. (1966). *The Conceptual Development of Quantum Mechanics*. Ney York: McGraw-Hill Book Company.

Jammer, M. (1974). *The Philosophy of Quantum Mechanics*. Wiley.

Janssen, M. (1995). *A Comparison Between Lorentz's Ether Theory and Special Relativity in the Light of the Experiments of Trouton and Noble*. Dissertation. University of Pittsburgh, 1995. Available at http://www.mpiwg-berlin.mpg.de/litserv/diss/janssen_diss/TitleTOC.pdf.

Janssen, M. (2002a). "Reconsidering a scientific revolution: the case of Einstein versus Lorentz". *Physics in Perspective* **4**: 421-446. Available at https://netfiles.umn.edu/users/janss011/home%20page/HALvsAE.pdf.

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 = mc2, and a slice of Minkowski space-time"; in Abhay Ashtekar et al. (ed.), *Revisiting the Foundations of Relativistic Physics: Festschrift in honor of John Stachel*. Dordrecht: Kluwer, 2003. Available at https://netfiles.umn.edu/users/janss011/home%20page/trouton.pdf.

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. & Mecklenburg, M. (2007). "From classical to relativistic mechanics: electromagnetic models of the electron"; in: V. F. Hendricks, K. F. Jørgensen, J. Lützen, and S. A. Pedersen (eds.), *Interactions: Mathematics, Physics and Philosophy 1860–1930.* Dordrecht: Springer, 2007. Available at https://netfiles.umn.edu/users/janss011/home%20page/electron.pdf.

Janssen, M. & Stachel, J. (2004). "The optics and electrodynamics of moving bodies", preprint, Max Planck Institute for the History of Science. Available at http://www.mpiwg-berlin.mpg.de/Preprints/P265.PDF.

Jarret, J. (1984). "On the physical significance of the locality conditions in the Bell arguments". *Noûs* **18**: 569-89.

Katzir, S. (2005). "Poincaré's relativistic physics: its origins and nature". *Physics in Perspective* **7**: 268-292.

Kox, A. J. (1988). "Hendrik Antoon Lorentz, the ether, and the general theory of relativity". *Archive for History of Exact Sciences* **38** (1988): 67–78.

Kox, A. J. (2013). "Hedrik Antoon Lorentz struggle with quantum theory". *Archive for History of Exact Sciences* **67**: 149-70.

Kragh, H. (1999). *Quantum Generations: a history of physics in the twentieth century*. Princeton: Princeton University Press.

Kragh. H. (2000). "Max Planck, the reluctant revolutionary". *Physics World* **13**: 31-5.

Kuhn, T. (1978). *Black-Body Theory and the Quantum Discontinuity, 1984-1912*. Chicago: University of Chicago Press.

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. (1994). "Non-empirical theoretical virtues and the argument from underdetermination". *Erkenntnis* **41**: 157-170.

Kukla, A. (2001). "Theoreticity, underdetermination, and the disregard for bizarre scientific hypotheses". *Philosophy of Science* **68**: 21-35.

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes. Philosophical Papers.* Vol. 1. Cambridge: Cambridge University Press, 1978

Laudan, L. (1965). "Grünbaum on the Duhemian argument". *Philosophy of Science* **32**. Reprinted in Harding, Sandra (ed.) (1976).

Laudan, L. (1977). *Progress and its Problems: towards a theory of scientific growth*. Berkeley: University of California Press.

Laudan, L. & Leplin, J. (1991). "Empirical equivalence and underdetermination". *The Journal of Philosophy* **88**: 449-472.

Laudan, L. & Leplin, J. (1993). "Determination underdeterred: reply to Kukla". *Analysis* **53**: 8-16.

Leavens, C. "The 'tunneling time problem' for electrons". In Cushing, J., Fine A. & Goldstein, S. (eds), *Bohmian Mechanics and Quantum Theory: an appraisal*, Springer, 111-30.

Leavens, C., & Aers, G. (1993). "Bohmian trajectories and the tunneling time problem". In Wiesendanger, R. & Güntherodt, J. (eds), *Scanning Tunneling Microscopy III*, Springer, 105-40.

Leplin, J. (1975). "The concept of an *ad hoc* hypothesis". *Studies in History and Philosophy of Science* **5**: 309–345.

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.

Lewis, P. (2007). "How Bohm's theory solves the measurement problem". *Philosophy of Science* **74**: 749-60.

Lombardi, O & Dieks, D. (2012). "Modal interpretations of quantum mechanics". *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/win2012/entries/qm-modal/.

Lorentz, H. A. (1886). "Over den invloed, dien de beweging der aarde op de lichtverschijnselen uitoefent". *Koninklijke Akademie van Wetenschappen* (Amsterdam). *Afdeeling Natuurkunde. Verslagen en*

Mededeelingen **2**: 297–372. French translation: "De l'influence du mouvement de la terre sur les phénomènes lumineux". *Archives Néerlandaises des Sciences Exactes et Naturelles* **21**: 103–176. This translation is reprinted in Lorentz 1934–39, Vol. 4, 153–214.

Lorentz, H. A. (1892a). "La Théorie électromagnétique de Maxwell et son application aux corps mouvants". *Archives Néerlandaises des Sciences Exactes et Naturelles* **25**: 363–552. Reprinted in Lorentz 1934–39, Vol. 2, 164–343.

Lorentz, H. A. (1892b). "De relatieve beweging van de aarde en den aether". *Verslagen van de gewone vergaderingen der wis– en natuurkundige afdeeling, Koninklijke Akademie van Wetenschappen te Amsterdam* **1** (1892–1893): 74–79. English translation in Lorentz 1934–39, Vol. 4, 219–23.

Lorentz, H. A. (1895). *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern.* Leiden: Brill. Reprinted in Lorentz 1934–39, Vol. 5, 1–138.

Lorentz, H. A. (1899). "Simplified theory of electrical and optical phenomena in moving bodies". *Proceedings of the section of sciences, Koninklijke Akademie van Wetenschappen te Amsterdam* **1**: 427–42. Reprinted in Schaffner 1972, 255–73.

Lorentz, H. A. (1902). "Théorie simplifiée des phénomènes électriques et optiques dans des corps en mouvement". *Archives Néerlandaises des Sciences Exactes et Naturelles* **7**: 64–80. Translation of Lorentz 1899a. Reprinted in Lorentz 1934–39, Vol. 5, 139–55.

Lorentz, H. A. (1904a). "Weiterbildung der Maxwellschen Theorie. Elektronentheorie". In Sommerfeld, A (ed), *Encyclopädie der mathematischen Wissenschaften, mit Einschlussiher Anwendungen*, vol.5, *Physik*, part 2. Leipzig: Teubner, 1904–1922, 145–288.

Lorentz, H. A. (1904b). "Electromagnetische verschijnselen in een stelsel dat zich met willekeurige snelheid, kleiner dan die van het licht, beweegt". *Verslagen van de gewone vergaderingen der wis– en natuurkundige afdeeling, Koninklijke Akademie van Wetenschappen te Amsterdam* **12**: 986–1009. English translation: "Electromagnetic phenomena in systems moving with any velocity less than that of light". *Proceedings of the section of sciences, Koninklijke Akademie van Wetenschappen te Amsterdam* **6**: 809–831. Reprinted in Lorentz 1934–39, Vol. 5, 172–97, and (without the final section 14) in Lorentz *et al.* 1952.

Lorentz, H. (1916). *The Theory of Electrons and its Applications to the Phenomena of Light and Radiant Heat*. Leipzig: Teubner. Reprinted in facsimile by Dover: New York, 1952.

Lorentz, H. A. (1934–39). *Collected Papers*. 9 Vols. The Hague: Nijhoff.

Lorentz, H. A., Einstein, A., Minkowski, H., and Weyl, H. (1952). *The Principle of Relativity*. New York: Dover.

Magnus, P. D. (2003). "Underdetermination and the problem of identical rivals". *Philosophy of Science* **70**: 1256-1264.

Maudlin, T. (1995a). "Three measurement problems". *Topoi* **14**: 7-15.

Maudlin, T. (1995b). "Why Bohm's theory solves the measurement problem". *Philosophy of Science* **62**: 479-83.

Maudlin, T. (1996). "Space-time I the quantum world". In Cushing, J., Fine. A & Goldstein, S. (eds.). *Bohmian Mechanics and Quantum Theory: an appraisal*. Springer.

Maudlin, T. (2011). *Quantum Non-locality and Relativity: metaphysical intimations of modern physics*. Wiley-Blackwell.

McAllister, J. (1989). "Truth and beauty in scientific reason". *Synthese* **78**: 25-51.

McCormmach, R. (1970a). "Einstein, Lorentz, and the electron theory". *Historical Studies in the Physical Sciences* **2**: 41–87.

McCormmach, R. (1970b). "H.A. Lorentz and the electromagnetic view of nature". *Isis* **61**: 459–497.

Mermin, N. (1990). "Simple unified form for the major no-hidden-variables theorems". *Physical Review Letters* **65**: 3373-6.

Mermin, N. (1993). "Hidden variables and the two theorems of John Bell". *Reviews of Modern Physics* **65**: 803-15.

Miller, A. I. (1973). "A study of Henri Poincaré's *Sur la dynamique de l'électron*". *Archive for History of Exact Sciences* **10**: 207–328. Reprinted in Miller 1986.

Miller, A. I. (1974). "On Lorentz's methodology". *The British Journal for the Philosophy of Science* **25**: 29-45. Reprinted in Miller 1986.

Miller, A. I. (1980). "On some other approaches to electrodynamics in 1905", in Miller (1986).

Miller, A. I. (1990). *Sixty-two Years of Uncertainty*. New York: Plenum Press.

Miller, A. I. (1996). "Why did Einstein not formulate special relativity in 1905?", in Greffe, J.-L. et al. *Henri Poincaré: Science and Philosophy*. Ed. Akademie Verlag, Berlin, and Albert Blanchard, Paris

Miller, A. I. (1998) [1981]. *Albert Einstein's Special Theory of Relativity: emergence (1905) and early interpretation (1905–1911)*. New York: Springer.

Miller, A. I. (1986). *Frontiers of Physics: 1900–1911*. Boston: Birkhäuser.

Miller, R. (1987). *Fact and Method: explanation, confirmation and reality in the natural and the social sciences*. Princeton University Press.

Minkowski, H. (1909). "Raum und Zeit". *Physikalische Zeitschrift*, **10**, pp. 104–111. Reprinted in Lorentz et al. 1952.

Monton, B. (2002). "Wave function ontology". *Synthese* **130**: 265-77.

Monton, B. (2006). "Quantum mechanics and 3*N*-dimensional space". *Philosophy of Science* **73**: 778-89.

Mormann, T. (1995). "Incompatible empirically equivalent theories: a structural explication". *Synthese* **103**: 204-249.

Muller, F. A. (1997). "The equivalence myth of quantum mechanics—part I". *Studies in history and Philosophy of Modern Physics* **28**: 35-61.

Muller, F. A. (2013). "Circumveiloped by obscuritads: the nature of interpretation in quantum mechanics, hermeneutic circles and physical reality, with cameos of James Joyce and Jacques Derrida", to appear in Synthese, preprint available at http://philsci-archive.pitt.edu/10166/1/Obscuritads-FAMuller2013.pdf.

Muller, F. A. & Saunders, S. (2008). "Discerning fermions". *British Journal for the Philosophy of Science* **59**: 499-548.

Muller, F. A. & Seevinck, M. (2009). "Discerning elementary particles". *Philosophy of Science* **76**: 179-200.

Murdock, D. (1987). *Niels Bohr's Philosophy of Physics*. Cambridge University Press.

Myrvold, W. (2002). "Modal interpretations and relativity". *Foundations of Physics* **32**: 1773-84.

Myrvold, W. (2003). "On some early objections to Bohm's theory". *International Studies in the Philosophy of Science* **17**: 7-24.

Nersessian, N. J. (1984). "Aether/or: the creation of scientific concepts". *Studies in History and Philosophy of Science* **15**: 175-212.

Nersessian, N. J. (1986). "Why wasn't Lorentz Einstein? An examination of the scientific method of H. A. Lorentz". *Centaurus* **29**: 205–242.

Norsen, D. (2009). "Local causality and completeness: Bell vs. Jarrett". *Foundations of Physics* **39**: 273-94.

Norsen, D. (2010). "The theory of (exclusively) local beables". *Foundations of Physics* **40**: 1858-84.

Norton, J. (1994). "Why geometry is not conventional: the verdict of covariant principles", in Majer, U., and Schmidt, H-J., *Semantical aspects of spacetime theories*, pp. 159–167. Meinheim/Leipzig/Wien/Zurich: Wissenschaftsverlag.

Norton, J. (2008). "Must evidence underdetermine theory?", in Carrier, M., Howard, D., and Kourany, J., *The Challenge of the Social and the Pressure of Practice: science and values revisited*, pp. 17-44. Pittsburgh: University of Pittsburgh Press, 2008.

Nugayev, R. (1985). "The history of quantum mechanics as a decisive argument favoring Einstein over Lorentz". *Philosophy of Science* **52**: 44-63.

Okasha, S. (1997). "Laudan and Leplin on empirical equivalence". *The British Journal for the Philosophy of Science* **48**: 251-256.

Okasha, S. (2002). "Underdetermination, holism and the theory/data distinction". *The Philosophical Quarterly* **52**: 303-319.

Omnès, R. (1999). *Understanding Quantum Mechanics*. Princeton University Press.

Pagonis, C. & Clifton, R. (1995). "Unremarkable contextualism: dispositions in the Bohm theory". *Foundations of Physics* **25**: 281-96.

Pais, A. (1982). *'Subtle is the Lord...': the science and the life of Albert Einstein*. Oxford: Oxford University Press.

Park, S. (2009). "Philosophical responses to underdetermination in science". *Journal for General Philosophy of Science* **40**: 115-124.

Pauli, W. (1952). "Remarques sur le proble`me des parame`tres cache´s dans la me´canique quantique et sur la theòrie de l'onde pilote. In *Louis de Broglie: Physicien et Penseur,* Paris: Éditions Albin Michel, 33–42.

Philippidis, C., Dewdney, C. & Hiley, B. J. (1979). "Quantum interference and the quantum potential". *Il Nouvo Cimento B* **52**: 15-28.

Plotnisky, A. (2013). *Niels Bohr and Complementarity: an introduction*. Springer

Poincaré, H. (1900a). "Sur les rapports de la physique expérimentale et de la physique mathématique". In: *Rapports Presentés au Congrès International de Physique Réuni à Paris en 1900.* Vol. 1, pp. 1–29. Paris: Gauthier–Villars. Translated as chs. 9–10 in Poincaré 1952.

Poincaré, H. (1900b). "La théorie de Lorentz et le principe der réaction". In: Bosscha 1900, pp. 252–278. Reprinted in Poincaré 1934–54, Vol. 9, pp. 464–488.

Poincaré, H. (1904). "L'état actuel et l'avenir de la physique mathématique". *Bulletin des Sciences Mathématiques* **28**: 302–324. Translated as chs. 7–9 in Poincaré 1958.

Poincaré, H. (1906). "Sur la dynamique d'électron". *Rendiconti del Circolo Matematico di Palermo* **21**: 129–175. Reprinted in Poincaré 1934–54, Vol. 9, pp. 494–550. English translation of sections 6–8 in Miller 1973.

Poincaré, H. (1934–54). *OEvres de Henri Poincaré*. 11 Vols. Paris: Gauthier-Villars.

Poincaré, H. (1952a). *Science and Hypothesis*. New York: Dover.

Poincaré, H. (1952b). *Science and Method*. New York: Dover.

Poincaré, H. (1958). *The Value of Science*. New York: Dover.

Poincaré, H. (1963). *Mathematics and Science: last essays*. New York: Dover.

Popper, K. (1959). "Testability and ad hocness of the contraction hypothesis". *British Journal for the Philosophy of Science* **10**: 50.

Popper, K. (2002). *The Logic of Scientific Discovery*. London: Routledge.

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

Quine, W. V. O. (1951). "Two dogmas of empiricism". *The Philosophical Review*: **60**. Reprinted in Harding, Sandra (1976).

Quine, W. V. O. (1975). "On empirically equivalent systems of the world". *Erkenntnis* **9**: 313-328.

Reichenbach, H. (1956). "The genidentity of quantum particles". In *The Direction of Time*. Dover, 224-36. Reprinted in Castellani, E. (ed.), *Interpreting Bodies: classical and quantum objects in modern physics*. Princeton University Press, 61-73.

Reichenbach, H. (1958). *The Philosophy of Space & Time.* New York: Dover.

Reignier, J. (2004). "Poincaré synchronization: from the local time to the Lorentz group", in *Proceedings of the Symposium Henri Poincaré (Brussels, 8-9 October 2004).* International Solvay Institutes for Physics and Chemistry. Available at
http://www.ulb.ac.be/sciences/ptm/pmif/ProceedingsHP/Reignier.pdf.

Ruetsche, L. (2003). "Modal semantics, modal dynamics and the problem of state preparation". *International Studies in the Philosophy of Science* **17**: 25-41.

Rice, D. (1997). "A geometric approach to nonlocality in the Bohm model of quantum mechanics". *America Journal of Physics* **65**: 144-7.

Rindler, W. (1991). *Introduction to Special Relativity*. Oxford: Oxford University Press.

Sarkar, H. (2000). "Empirical equivalence and underdetermination". *International Studies in the Philosophy of Science* **14**: 187-197.

Sartori, L. (1996). *Understanding Relativity.* University of California Press.

Saunders, S. (1995). "Time, quantum mechanics, and decoherence". *Synthese* **102**: 235-66.

Saunders, S. (2003). "Physics and Leibniz's principles". In Brading, K. & Castellani, E. (eds.), *Symmetries in Physics: philosophical reflections*. Cambridge University Press, 289-307.

Saunders, S. (2006). "Are quantum particles objects?" *Analysis* **66**: 52-63.

Schaffner, K. F. (1969). "The Lorentz electron theory of relativity". *American Journal of Physics* **37**: 498-513.

Schaffner, K. F. (1972). *Nineteenth Century Aether Theories*. Oxford, New York: Pergamon Press.

Schaffner, K. F. (1974). "Einstein versus Lorentz: research programmes and the logic of comparative theory evaluation". *The British Journal for the Philosophy of Science*, **25**: 45–78.

Schaffner, K. F. (1976). "Space and time in Lorentz, Poincaré, and Einstein: divergent approaches to the discovery and development of the special theory of relativity". In: Machamer, Peter K., and Turnbull, Robert G. (eds.). *Motion and Time, Space and Matter: interrelations in the history of philosophy of science*. Ohio State University Press.

Schilpp, P. A. (ed.) (1949). *Albert Einstein: philosopher-scientist*. Evanston, IL: Library of Living Philosophers.

Schlosshauer, M. (2004). "Decoherence, the measurement problem, and interpretations of quantum mechanics". *Reviews of Modern Physics* **76**: 1267-1305.

Schlosshauer, M. (2007). *Decoherence and the Quantum-to-Classical Transition*. Berlin: Springer.

Scribner, Ch., Jr. (1964). "Henri Poincaré and the principle of relativity". *American Journal of Physics* **32**: 672–678.

Seevinck, M. (2010). "Can quantum theory and special relativity peacefully coexist?", arXiv:1010.3714v1.

Segrè, E. (1980). *From X-Rays to Quarks: modern physicists and their discoveries.* San Francisco: W. H. Freeman and Company.

Shimony, A. (1990). "Exposition of Bell's theorem", in Miller, A (ed), 33-43.

Sklar, L. (1974). *Space, Time, and Spacetime*. Berkeley, Los Angeles, London: University of California Press.

Stachel, J. (1982). "Einstein and Michelson: the context of discovery and the context of justification". *Astronomische Nachrichten* **303**, I, 47–53.

Stachel, J. (2002). "What song the syrens sang?: how did Einstein discover special relativity", in Stachel, J., *Einstein from "B" to "Z"*. Boston: Birkhäuser.

Stanford, K. (2001). "Refusing the devil's bargain: what kind of underdetermination should we take seriously?" *Philosophy of Science* **68 (**Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers): S1-S12.

Stanford, K. (2009). "Underdetermination of scientific theory". *Stanford Encyclopedia of Philosophy,* http://stanford.library.usyd.edu.au/entries/scientific-underdetermination/.

Szabó, L (2011). "Lorentzian theories vs. Einsteinian special relativity—a logico-empiricist reconstruction" in András, M., Rédei, M. & Stadler, F. (eds); *The Vienna Circle in Hungary*. Springer. Available at http://philsci-archive.pitt.edu/5339/.

Teller, P. (1998). "Quantum Mechanics and Haecceities", in E. Castellani (ed.), *Interpreting Bodies: classical and quantum objects in modern physics.* Princeton University Press, 114–41.

Timpson, C. & Brown, H. (2005). "Proper and improper separability". *International Journal of Quantum Information* **3**: 679-90.

Tipler, F. (2000). "Does quantum nonlocality exist? Bell's theorem and the many-worlds interpretation", arXiv:quant-ph/0003146.

Torretti, R. (1996) [1983]. *Relativity and Geometry*. New York: Dover.

Torretti, R. (2003). *Relatividad y Espaciotiempo.* Santiago: RIL.

Toyama, F.M & Matsuura, K. (2006). "Non-local correlations in Bohm trajectories". *Physica Scripta* **73**: 17-22

Vaidman, L. (2009). "May-worlds interpretation of quantum mechanics". *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/spr2009/entries/qm-manyworlds/.

Valentini, A. (2008). "De Broglie-Bohm pilot-wave theory: many worlds in denial?", arXiv:0811.0810v2.

Van Dongen, J. (2009). "On the role of the Michelson-Morley experiment: Einstein in Chicago". *Archive for History of Exact Sciences* **63**: 655-663.

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

Vigier, J. P., Dewdney, C., Holland, P. & Kyprianidis, A. (1987). "Causal particle trajectories and the interpretation of quantum mechanics", in Hiley, B. J & David Peat, F. (eds.), *Quantum Implications: essays in honor of David Bohm*. Routledge, 169-205.

Von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton University Press.

Wallace, D. (2002). "Worlds in the Everett interpretation". *Studies in History and Philosophy of Modern Physics* **33**: 637-61.

Wallace, D. (2003). "Everett and structure". *Studies in History and Philosophy of Modern Physics* **34**: 87-105.

Wallace, D. (2012). *The Emergent Multiverse: quantum theory according to the Everett interpretation*. Oxford: Oxford University Press.

Walter, S. (2010) "Minkowski's modern world", in Petkov, V. (ed), *Minkowski Spacetime: a hundred years later*. Springer.

Warwick, A. (1995). "The sturdy protestants of science: Larmor, Trouton and the earth's Motion through the aether". Chapter 11 in Buchwald, J. Z. *Scientific Practice: theories and stories of doing physics.* The University of Chicago Press, 1995.

Whitaker, A. (1996). *Einstein, Bohr, and the Quantum Dilemma*. Cambridge University Press.

Whittaker, E. (1953). *A History of the Theories of Aether and Electricity*. 2 Vols. London: Nelson.

Wolff, J. (2014). "Heisenberg's observability principle". *Studies in History and Philosophy of Modern Physics* **45**: 19-26.

Yalçin, Ü. (2001). "Solutions and dissolutions of the underdetermination problem". *Noûs* **35**: 394-418.

Zahar, E. (1973). "Why did Einstein's programme supersede Lorentz's?" *The British Journal for the Philosophy of Science* **24**: 95-123, 223-262. Reprinted in Zahar 1989.

Zahar, E. (1977). "Mach, Einstein and the rise of modern science". *The British Journal for the Philosophy of Science* **28**: 195–213. Reprinted in Zahar 1989.

Zahar, E. (1978). "Einstein's debt to Lorentz: a reply to Feyerabend and Miller". *The British Journal for the Philosophy of Science* **29**: 49–60. Reprinted in Zahar 1989.

Zahar, E. (1989). *Einstein's Revolution: a study in heuristic*. La Salle, IL: Open Court.

Zeh, H. D. (1999). "Why Bohm's quantum theory?" *Journal of Genetic Counseling* **12**: 197-200.

Zurek, W. (2002). "Decoherence and the transition from quantum to classical—*revisited*", arXiv:quant-ph/0306072.

SUMMARY IN DUTCH

Empirische Gelijkwaardigheid en Onderdeterminatie van Theorie Keuze

Wetenschappelijke theorieën worden geaccepteerd of afgewezen in termen van waarneembare voorspellingen. Indien de voorspellingen van een theorie geverifieerd worden in experimenten en waarnemingen, dan wordt de theorie, op basis van bewijs, bevestigd, maar indien de resultaten van waarnemingen en experimenten niet overeenstemmen met de voorspellingen wordt de theorie afgewezen. Echter, wanneer twee rivaliserende en strijdige theorieën exact dezelfde voorspellingen leveren en indien de resultaten van alle relevante waarnemingen en experimenten overeenstemmen met deze voorspellingen, dan worden beide theorieën in gelijke mate bevestigd en is er geen criterium op basis van bewijs om de correcte theorie te selecteren. En omdat de theorieën rivaliserend en strijdig zijn kunnen we niet beide accepteren. Deze situatie staat binnen de wetenschapsfilosofie bekend als het probleem van de empirische equivalentie en onderbepaaldheid van theoriekeuze.

Dit probleem is het onderwerp van mijn dissertatie. Ik bied een conceptuele evaluatie van de implicaties van dit probleem en ik verdedig een oplossing die gebaseerd is op de gebruikelijke praktijk van de wetenschap: Ik betoog dat het probleem eerder een moeilijkheid voor de wetenschap is dan een onoplosbaar filosofisch woordraadsel. Ik pas de voorgestelde evaluatie en de oplossing van het probleem toe op twee case studies van empirische equivalentie in de moderne natuurkunde: Einsteins speciale relativiteitstheorie tegenover Hendrik Lorentz' ether theorie, en de standaard kwantummechanica tegenover David Bohms alternatieve kwantummechanica.

CURRICULUM VITAE

Pablo Acuña L.

Address: Carlos Antúnez 1856, dp. 101, Santiago de Chile.
Date of birth: May 6th, 1980.
Mobile phone: +56(0)92899120
E-mail: p.t.acunaluongo@uu.nl
Citizenship: Chilean/Italian

Academic Background

2012-2014 *PhD in History and Philosophy of Science*, Utrecht University. Thesis title: *Empirical Equivalence and Underdetermination of Theory Choice: a philosophical appraisal and two case studies*. Supervisor: Dennis Dieks.

2010-2012 *MSc in History and Philosophy of Science* (cum laude), Utrecht University.

2008-2009 *Magíster en Filosofía de las Ciencias* (Master in Philosophy of Science), Universidad de Santiago de Chile.

2001-2006 *Licenciado en Filosofía* (Bachelor in Philosophy), Pontificia Universidad Católica de Chile.

1998-2004 *Licenciado en Letras con mención en Lingüística y Literatura* (Bachelor in Linguistics and Literature), Pontificia Universidad Católica de Chile.

Teaching Experience

2014 Pontificia Universidad Católica de Chile / Universidad Nacional Andrés Bello.

2011/2013 Teaching assistant in the graduate course *Philosophy of Space & Time*, dictated by Dennis Dieks.

2007-2009 Pontificia Universidad Católica de Chile / Universidad Nacional Andrés Bello.

2003-2009 High School teacher (*philosophy* and *Spanish language and communication*)

Publications

- 'On the Empirical Equivalence between Special Relativity and Lorentz's Ether Theory' (2014), *Studies in History and Philosophy of Modern Physics* **46**: 283-302, http://dx.doi.org/10.1016/j.shpsb.2014.01.002.

- 'Another Look at Empirical Equivalence and Underdetermination of Theory Choice' (2014), in collaboration with Dennis Dieks, *European Journal for Philosophy of Science* **4**: 153-180, http://dx.doi.org/10.1007/s13194-013-0080-3.

- 'Artificial Examples of Empirical Equivalence' (2014). In M. C. Galavotti *et al*. (eds.) *New Directions in the Philosophy of Science*, Springer. *Preprint* available at http://philsci-archive.pitt.edu/9893/.

Conferences

- 'Empirical Equivalence and Underdetermination of Theory Choice: a philosophical appraisal and a case study', in *New Directions in the Philosophy of Science,* Bertinoro, Italy, October 2012. http://www.pse-esf.org/Bertinoro-ScientificReport.pdf.

- 'Incommensurability and the Rationality of Science', in *Third Young Researchers Days in Logic, History and Philosophy of Science*, Brussels, September 2012. http://www.bslps.be/yrd3_schedule.pdf.

Prizes

- *Best Graduate Thesis* (2012-2013), Utrecht University. http://www.uu.nl/university/education/en/topteachersandtopstudents/studentawards/bestgraduatethesis/Pages/default.aspx.

- Second place in the international contest *Hanneke Janssen Memorial Prize 2012* for master theses in philosophy of physics, Radboud University. Jury report available at http://www.ru.nl/publish/pages/669795/jury_report_for_the_hanneke_janssen_memorial_prize_2012.pdf.

References

- Dennis Dieks. Institute for History and Foundations of Science, Utrecht University, d.dieks@uu.nl.

- Roberto Torretti. Emeritus Professor, Universidad Diego Portales, roberto.torretti@gmail.com.