Year : 2021

A Philosophical Study of Observation in Quantum Mechanics

Trotter Frida

UNIL | Université de Lausanne
Faculté des lettres

FACULTÉ DES LETTRES

SECTION DE PHILOSOPHIE

A Philosophical Study of Observation in Quantum Mechanics

THÈSE DE DOCTORAT

présentée à la

Faculté des lettres
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès lettres

par

Frida Trotter

Directeur de thèse

Michael Esfeld

UNIL | Université de Lausanne

Faculté des lettres

IMPRIMATUR

Le Décanat de la Faculté des lettres, sur le rapport d'une commission composée de :

Directeur de thèse :

Monsieur	Michael Esfeld	Professeur, Faculté des lettres, UNIL

Membres du jury :

Madame	Michela Massimi	Professeure, University of Edinburgh, Écosse
Monsieur	Anjan Chakravartty	Professeur, University of Miami, Etats-Unis
Monsieur	Eric Cavalcanti	Professeur associé, Griffith University, Gold Coast Campus, Queensland, Australie

autorise l'impression de la thèse de doctorat de

MADAME FRIDA TROTTER

intitulée

A Philosophical Study of Observation in Quantum Mechanics

sans se prononcer sur les opinions du candidat / de la candidate.

La Faculté des lettres, conformément à son règlement, ne décerne aucune mention.

Lausanne, le 30 septembre 2021

Léonard Burnand
Doyen de la Faculté des lettres

A Philosophical Study of Observation in Quantum Mechanics

Frida Trotter

September 29, 2021

Contents

Introduction

εἰ καὶ μὴ βλέπεις, φρονεῖς δ' ὅμως

You understand, though you don't see with the eyes

Sophocles, Oed. Ty. v. 302

When noticing a patch of rainbow on the wall in a sunny morning, and looking for the part of the window's glass that is acting as a prism, it is natural to become curious about what, more generally, brings about this particular phenomenon. How was that particular ray refracted so as to display all the visible spectrum? What is going on in that particular area of the glass, what, in its molecular composition, makes the photons undergo a modification of their wave lengths? *What is the physical process that underlies this patch of rainbow on the wall?* The answer that one would find by enquiring into physics would have a formal appearance, and be expressed in the language of mathematics. And the same theory vouching for this answer would be able to account for all the other phenomena involving light, and explain the basic mechanisms underlying every particular observable instance, independently of its complexity.

In order to have this powerful generality, so as to be able to unify an astoundingly broad range of phenomena under descriptions deriving from a narrow set of principles, a fundamental physical theory needs to be abstract, equipped with mechanisms to idealise, and able to yield predictions that can be tested in particular experimental instances. The indispensable tool that allows them to achieve this is the use of mathematical structures. The fact is that a very small number of fundamental physical theories, elegantly expressed in a short list of axioms are able to account for a wild variety of natural phenomena.

But *how* is it possible that a formal mathematical model describing an idealized and not realizable situation, is representing this and any other "patch of rainbow" on any surface? More generally, how is it possible for formal theoretical descriptions provided by physics to account for the particular and complex manifold of natural phenomena? Also, how is it possible for the formal apparatus put in place by fundamental physics to account accurately for this manifold, to the extent that the predictions it yields are characterized by such a high level of precision? Even more surprisingly, how is it possible that the elaboration of such a framework has also led to the discovery of previously unknown objects and properties of these phenomena? These questions have generated discussions in philosophy of science that have spanned over a century, and they haven't encountered definitive answers yet.

One particular side to this polymorphic set of problems is the question of how the connection between the formal description of a theory and a particular phenomenon is set up. A physicist who would notice the coloured patch on the wall could trace back what they see to their knowledge about this phenomenon, and understand it through the lenses of the idealised description derived from the theory's axioms. They would be able to do this also more generally, considering less contingent events such as for instance a set of measurements in a laboratory. And they could design an experiment in order to reproduce a phenomenon or some of its features, and verify whether their theoretical knowledge is able to capture it correctly. These two jobs could be described separately, namely, the development of the theory in its formal aspects on one side, and the construction of an optimized process to collect data about a specific phenomenon on the other. But one operation in which theoretical knowledge and its application to the natural world happen *at once* is scientific observation.

An observing physicist is the catalyst of two different sets of skills, as they need to master an abstract theory and also to single out, from the complexity of experience, the elements that are relevant for scientifically understanding a phenomenon, i.e. through the lenses of the theory. For its own structure, scientific observation is a process that *needs* to incorporate both the theoretical and the empirical components at the same time, in order to be scientifically meaningful. And the interplay of the theoretical and empirical components has been object of rich discussions since the beginning of modern philosophy of science. The charm of this fascinating side of the relation between humans and nature has influenced my research as an undergraduate already. It has been natural then to use the opportunity of this Ph.D. thesis to further explore the connection between theory and nature by enquiring into observation.

In general philosophy of science, observation is one of the concepts that have changed their meaning and role over the decades, depending on the development of correlated philosophical questions. If in the beginning of the last century observation was used to qualify the commonly agreed upon, immediately accessible manifold of human perceptual experiences in contrast with the abstractness of theories, the situation changed with the progressive acknowledgement that there is no immediacy in our experience of the world, and that perception is deeply biased by knowledge and beliefs of individuals. From placeholder of objectivity, it became something to be examined in the first place in connection with the scientific theories that encapsulated the knowledge possessed at a certain time. Perhaps more famously, debates on observation have been long time companions to the philosophical enquiry on scientific realism and antirealism, as terrain for the latter has been the dispute over what is observable and which kind of ontological commitment one should have towards observable and unobservable entities. For these reasons, observation has been defined in different manners, starting from a simple equation with human perception, to a minimal description in terms of process through which we gather information about a system.

The interesting side of these discussions is, I find, the moment in which observation is used as a vehicle for the exploration of the epistemic extent of scientific theories. On one extreme we find positions according to which scientific theories need to be empirically adequate, and everything they postulate should be understood instrumentally as tools to account for what we can observe. On the other, we find positions according to which not only the explanation of the observable is not the end of the story, but it may be only a stage of the scientific exploration of the universe, doomed to be relegated to the past in favour of a non-empirical future of the development of science. Either way, it seems that observation remains a resourceful locus for a fundamental discussion on the scientific epistemology of the world. Nevertheless, the present work is closer to the second class of positions, whereby the way in which observation is characterized in these pages directly aligns with how it is regarded in the current scientific practice.

Of course, these general considerations do not apply in the same manner to the whole of science in general. Behind their apparent ecumenical character there is a constant reference to the physical sciences that the informed reader hasn't missed. Well, the interest I have mentioned earlier, to enquire philosophically into the link between theories and nature, has materialized in this thesis as a research on observation in relation with physics, and with one physical theory in particular. The motivating question at the root of my research is general and eminently philosophical; methodologically, I have then decided to focus on physics, which is *locus classicus* for the debates in philosophy of science recalled above.

But why study quantum mechanics? Much of the discussion on observation has been carried out at a more general level as opposed to being restricted to a particular physical theory. If with generality comes unification and thus a gateway to comprehensive overviews, with detail comes depth, and the possibility to carve a problem possibly at its ultimate joints. Neither of these approaches is entirely satisfactory alone. Hence, I have decided to explore my general question by choosing a subject for which observation is relevant, and in relation to which the complexities related to observation could emerge with full intensity. Quantum mechanics displays precisely these two features. One could say that observers and observations have completely changed their meaning in 20th century physics, as they have lost their absoluteness in relativity theory, and their neutrality in quantum mechanics. There are many senses in which, for the combined teachings of these two macro-theories, observation has lost its alleged innocence and neutrality, and has become part of the physical description as well as the objects that these theories target. I chose to enquire into quantum mechanics given an original interest towards the fundamental ontology of the world and our epistemic access to it via empirical research, but continuing this research into another fundamental theory such as relativity could be a natural step forward.

But again, a general study of observation in relation with quantum mechanics would be an enormous task, as it would need to take into account a number of equally pressing and complex issues. Even by focusing solely on the measurement problem for instance, one could have gone into the problem of the disturbance of the system during measurement, on a discussion of the universality of the theory's unitary dynamics, on a study of decoherence; or, a different focus could have been put on the observers themselves, and on whether something new can be discovered by investigating on the human observer as able to trigger a special physical processes amongst others, for instance by appealing to the still mysterious character of consciousness. A further ramification of these issues would be provided by a study of how observation is described in the various interpretations of quantum mechanics and on the different quantum theories. This work has unavoidably touched upon some of these questions, but not as a main object of interest.

I have in fact further narrowed my scope of investigation down to one case study, towards which I have addressed my more general questions. To say that studying Bell's theorem amounts to a restriction of the scope of investigation however may sound ironic. First of all, it is not a theorem *of* quantum mechanics, as it is not derived from quantum principles, but its extent is notoriously more general than quantum mechanics. Second, the interpretive issues related to the implications of its violation by the latter theory are a large group that make this choice resemble all but a strategic move towards the detail. However, Bell's theorem is a unique result that points its finger on an eminently quantum effect that has been largely empirically confirmed, namely, the peculiar properties of entanglement. The reason why I have chosen Bell's theorem as a case study for my thesis is thus three-folded. On the one hand, the general applicability of this result constitutes the ideal ground for the discussion of the more general questions that underlie my research. On the other, the extremely rich history of experiments that confirmed the violation of the theorem by quantum mechanics, which spans over the period from 1950 to 2015, offers a great reserve of material to study precisely observation in the quantum context. All the so-called Bell tests are experiments on bipartite entangled systems that have consistently corroborated the quantum predictions, with the use both of a variety of physical systems and of different experimental techniques. I have seen in such a rich history a unique opportunity to explore in the first person the modalities, the assumptions and the results that have accompanied the observation of these quantum phenomena. This exploration has led to an understanding of the features of observation with respect to a core quantum phenomenon such as entanglement. Moreover, it has also fostered further philosophical questions that strictly exceed the role of observation in this story, such as the proposal for a metaphysical account of quantum mechanics as a physical theory, and an assessment of underdetermination among the various quantum theories and interpretations.

The thesis' set up is thus constituted by a general philosophical question addressed by looking at the specific practice of observation, within a particular physical theory that is quantum mechanics, and by exploring one case study represented by Bell's theorem; which results have been achieved in this exploration? As this is a philosophical work, I have the luxury of being able to move among arguments, and the freedom not to be demanded mathematical proofs. Nevertheless, there are three core results of the work made in the thesis.

The first is a reflection on observation: by studying the structure of quantum mechanics, the import of Bell's theorem at a purely theoretical level, and then looking at the concrete experiments that have incarnated these structures, it has emerged that the problems usually associated to the observation of quantum objects do not derive strictly from the lack of an adequate account of observation. Although this was a working hypothesis at the early stages of this research, finally it was abandoned. The issues related to observation in quantum mechanics derive directly from from the way in which this theory characterizes physical systems. For this reason, I have endorsed an existing definition of observation, namely, Dudley Shapere (1982), which I analyse and defend in the following chapters. This account provides a formula to classify *direct* observations.

The second is constituted by the proposal to envisage quantum mechanics as a *categorical toolbox* for understanding the world, as opposed to a theory that should directly model a fundamental ontology. This result derives from the study of entanglement in relation with Bell's theorem from a theoretical and experimental point of view. In the thesis however this proposal is developed only in a preliminary form, which I will use to set the stage for more extended future work.

Finally, the third result is related to an analysis of the main interpretations of quantum mechanics based on the significance of empirical evidence. I explore the features of the underdetermination of the interpretations, and assess the latter in terms of their theoretical virtues. What my analysis reveals is the "apple of discord" among the supporters of the various interpretations is represented by disagreement on normative assumptions on what science *ought to do*. Although the role of normativity in the development of science, and especially for the definition of progress has been acknowledged and argued for in general philosophy of science, assumptions of this character in the debate I am considering have not been sufficiently recognized. In connection with this point I advance some remarks on progress in the thesis' conclusions.

How does all this emerge in the current work? The thesis is divided in three macro parts that define the role of the chapters they contain. Part I contains the framework for the present analysis, and an introduction of the context of the philosophical issues touched in this work. Chapter 1 sets the philosophical stage for the exploration, by outlining the philosophical premises of the current enterprise. These are two methodological guidelines and a philosophical position on the one hand, and a clear minimal account of theory, evidence and explanation on the other. In chapter 2 I explore more closely observation in the philosophy of science, by outlining parts of the debate on the entities that can and that cannot be observed, and how these definitions have played a role in the classical contrast between the realist and the antirealist positions. I also enter in the matter of two examples that show how observation is special in the context of quantum mechanics, and flesh out a notion of principled unobservability. Part II encompasses the case study of my thesis, Bell's theorem. Chapter 3 provides an analysis of Bell's theorem from a conceptual perspective, by outlining its premises, its structure and implications depending on how one manages the theorem's premises. The experiments that have corroborated the correct predictions of quantum mechanics in violation of the theorem's result are explored in chapter 4, which also gives way to an analysis of the theoretical considerations surrounding the history of how the experiments were developed. Finally, part III contains the main philosophical analysis of the thesis. Chapters 5 takes stock from the work made in the previous chapters, and fleshes out the results outlined above. In particular, it focusses on the implication of reasoning on Bell's theorem starting from the evidence, and assesses the various interpretations of quantum mechanics based on their fundamental tenets. A consideration of the underdetermination in which they verse leads to acknowledging the resort to normative arguments on the basis of which their supporters endorse one interpretation over the others.

❈

The fact that an original question which is simple in its structure can lead to the investigation of complex and deep problems is an astounding commonplace in philosophy, a feature that captures its endless charm for the intellect. In good philosophy however, the peaceful simplicity of the original question is reached — no matter how long is the track to arrive there — in the renovated peace of a solution which is again simple in its structure. This work has taught me how important it is to envisage this ideal aim as the northern star, or the southern cross, while navigating under the constellations of thought.

Acknowledgements

Although everything that is written in these pages is my sole responsibility, I would like to wholeheartedly thank the people that made it possible for me to write any of this in the first place. I would like to thank Michael Esfeld, who not only helped me to navigate subtle and plainly hard topics such as those studied in this thesis, but who also believed in me and supported this work through moments of all sorts. I will not forget how original and unique these four years have been thanks to his equally original and unique way of being a philosophy professor, and a human being. To Eric Cavalcanti goes a big measure of my gratitude, for *really* teaching me the physics of Bell's theorem in a philosophically deep and extremely sharp manner. Perhaps this makes me a bit less of a newbie in the thick forest of its many subtleties, and it was a one in a lifetime experience to be welcomed by him to learn all this while surrounded by the subtropical forests and the Pacific Ocean. I acknowledge the traditional owners of the land at Griffith University, on which part of this work was completed, the Yuggera and Yugambeh peoples. I am also in debt with Michela Massimi and Anjan Chakravartty, who have read an earlier version of this manuscript in great detail, contributing directly to what I consider a significant improvement, thanks to their attentive comments.

For supporting me through all what takes place in the backstage of a thesis, such as reading, commenting and discussing countless drafts, I thank all my colleagues at the University of Lausanne; for also being real friends in all this, I thank in particular Christian, Tiziano, Guillaume, for sharing day-to-day experiences of all types, and the ability to transform complex issues of all sorts in long, entertaining conversations, and very often in liberating laughs. For proving that physical distance doesn't entail any sort of further distance, I thank Cristian and Federico, for all the invaluable things I have learned and shared with them, academically and outside. My life has been incredibly enriched by meeting all of you.

From the lively and stimulating group at the University of Geneva I thank Christian Wüthrich, for his hospitality at countless interesting seminars, and for his sensitivity towards the issues surrounding the life and duties of a philosophy doctoral student. I am in debt with some of his students and collaborators, Niels Linneman, Rasmus Jaksland, and Alberto Corti, for their precious insight on parts of my thesis, as well as with James Wills, with whom I was lucky enough to cross paths and benefit enormously from our discussions. At Griffith University I also had the occasion to discuss with Howard Wiseman and Michael Hall, on some arguments in this work that are particularly dear to me.

For having become a friend over the years, with whom I could exchange ideas and grow as a philosopher and as a person, I thank Antonio Nunziante, always a model of how I would like to become as an academic. Finally, I am blessed with amazing family and friends, who share the merits of this work in all the ways that haven't been already mentioned. My mother Ilde, my father Cristiano, my sister Verdiana, my soul-mother Tiziana and my soul-father Narci, it is incredible to be able to share equally joys and less joyful periods, and come out anyway stronger from either. My precious friends Elisabetta, Chiara, Sofia, Alessandro, Mattia, Michele, Francesco, Valeria, Evie, Vanessa, Justine, Laetitia, Jana, Elise, Amita, I can only look forward to more adventures with you! Lisa, for the part of life we have gone through together, and for how we have gone through it, on so many levels, this thesis is partially yours. Markus, thank you for making me understand that scientific ideas are an art. Before meeting you I didn't know that love, philosophy and physics could be lived fully at once.

Part I

Framework and Context

Chapter 1

Conceptual framework

The ideal of a starting point that has the most reasonable premises, or, even better, is unbiased, is and remains an ideal, and perhaps not even a very palatable one. Which kind of start would a complete absence of *any* type of bias provide for tackling themes that are intrinsically complex? Whether this question could be replied to in principle remains perhaps an open issue, but it is a certainty that this work does not approach this ideal, nor does it intend to do so. Hence, I start with a clear exposition of what are the tools for this analysis, which, for a philosopher, are premises and their justifications. With this I am not referring to the motivations and justifications that innervate the present research, but rather to something simpler. In philosophy of science, terms like *theory*, *explanation*, *realism*, *evidence*, *confirmation* just to name a few, are semantically quite rich, and for this reason they support many uses which are sometimes divergent from one another. Observation is one of these terms, and it is part of the aim of this thesis to open a window into its role and properties in fundamental physics. Since it is the central subject of the present work to focus on quantum mechanics, on Bell's theorem, and on what can be learned about observation from this study, the mentioned terms and other rich notions will appear rather often in these pages. Moreover, debates that refer to different definitions of these terms have arisen precisely around quantum mechanics and around some issues specific to the philosophical study of such a theory. My intention in this introductory chapter is therefore to provide a clarification of what notions I am embarking with in this enterprise.

The accounts schematically outlined in this chapter serve the purpose of a Wittgensteinian ladder that can be abandoned once the summit of the tree has been reached, and, as such, they are not definitive, nor immune to further questioning. In section 1.1 I outline two methodological guidelines — naturalism and fitness for quantum mechanics, that I intend to follow for the conduction of my analysis, and a philosophical position I adopt as a methodological principle, namely, scientific realism. In section 1.2 instead I provide the minimal definitions of three crucial relata of observation in science, i.e. the notions of theory, evidence, and explanation. The main aim of the present chapter is to set a minimal stage for the next steps; let us thus start then with the premises.

1.1 Two assumptions plus one

I have narrowed down the work's premises to three, two of which are methodological guidelines, while the third is a philosophical position with respect to theories and the scientific understanding of the world more broadly understood.

1 – ***Naturalism***. Perhaps along the same lines of the program of naturalized metaphysics, this first point stresses that the manner in which I am approaching the matter of observation is scientifically informed. How and to which extent a philosophical discipline ought to rely upon science has been argued with respect to metaphysics in Ross et al. (2013). The concept however had been previously developed in the philosophical tradition of American naturalism, see Kim (2003) and Papineau (2021). This premise may seem to be completely redundant in a thesis concerning this practice in connection with quantum mechanics, for obvious reasons. Therefore, let me specify why this clarification is due nevertheless. Reference to observation and observability has been a commonplace in the development of contemporary philosophy of science starting from the logical empiricists to the classical debates on scientific realism of the 80s. It has played a discriminatory role for the distinction among the entities that had to be endowed with ontological status as opposed to those that were left in the realm of postulation and that were attributed with a more functional than ontological role. Also, it has played a crucial role in the definition of anti-realism, or at least in the account developed by Bas van Fraassen, where the role of science is envisioned as to account for the *observable* phenomena. Well, within this particular debate, the definitions of observation employed have very often not drawn upon scientific directives, but have been formulated on independent, philosophically motivated grounds. The working idea of observation until the 80s has been to refer to *observable* in terms of *perceivable with unaided senses*. The definition has then evolved, as we will see in chapter 2, with the analysis of Shapere (1982). Problems of observation then have diluted into more specific questions that have moved the focus away from observation per se. The point remains that, at least in the classical study

of this subject, the starting point has been more commonsensical than scientifically informed; with the latter expression I mean both that it hasn't focused on particular physical or biological processes related to the manner in which humans perceive, and that it did not seem to take into account the manner in which scientists themselves envisage observation. It is a guideline of my research to start from a scientifically informed notion of observation; with this I do not intend that I am blindly following the scientific authority: the aim of this principle is to conceive scientific work as a clear source of concepts that are subject of my philosophical analysis.

2 – ***Fitness to approach quantum mechanics***. Despite its focus on the limit between observable and unobservable, and more specifically on how this distinction plays out with respect to microscopic fundamental entities, the traditional debate on observation has evolved on a fairly general level, by referring to "science" or to "physics" without qualifying these terms further. In a sense, it is by hovering over these large domains of knowledge that these analyses can reach a substantial level of generality. But on the other hand they fail to apply satisfactorily to more specific fields, or cases, especially when these display some exceptional peculiarities. As we will see in part in the following chapter, the traditional discussion on observation has not entered in the matter of any particular theory on these entities.

The second methodological guideline for my investigation is that I do not wish to maintain this non-committal approach; to the contrary, I start with a specific focus on quantum mechanics. Interestingly, observation and observer are concepts that have enjoyed separate attention in the realm of quantum mechanics. Classical locus for these considerations is the measurement problem, that prima facie questions the commonsensical neutrality of the interactions between measuring apparatus and physical systems, up to the point at which sometimes even the observers themselves are conceived as special elements in the quantum universe. Although these are matters that would appear of central importance in an investigation of observation, they are not primary catalysts of attention of my analysis. I study in fact observation as a physical interaction that enables transfer of information from a source to a receiver, and how this is characterized in quantum mechanics. Chapter 2 contains a detailed exposition of the account of observation adopted in this work. Given my focus is on quantum mechanics, I am not considering alternative theories of observation that have to do with different types of systems.

1.1.1 Plus one: scientific realism

The classical debate on observation in philosophy of science has traditionally come in pair with discussions on scientific realism. Interestingly, for a long time ontological commitment was reserved only to observable entities. Therefore, how to commit to the existence of fundamental entities, given that they were unobservable? Before conclusive empirical confirmation on the existence of the fundamental systems was available, these were envisaged as mere postulations, functional to build a scientific account of observable phenomena. The debate on realism has evolved since then, up to the point in which observability of the entities to which ontological commitment was not previously granted has ceased to be a problem. The third premise of my work consists in the realist commitments I am equipping myself with to embark in the following exploration.

As it is well known, in the last two decades work on scientific realism has been very prolific, to the point that the current realist accounts are very sophisticated, and especially fine-grained — an idea of the richness of the debate is offered by the recent Saatsi (2018). At this preliminary point of the thesis it would be premature and difficult to justify endorsing one particular account, and this is in fact not what the current premise is about. Rather, I have more parsimoniously referred to "realist commitments", an expression that refers to a take on scientific theories. These are: scientific theories are approximately true (Schurz, 2018), in the sense that the claims they make about objects and/or structures of the world describe a real state of affairs. If a theory postulates an ontology of objects that can be manipulated for the empirical testing of scientific hypotheses, these entities have ontological status. The entities that a theory may describe that cannot be subsumed under the latter category, e.g., abstract spaces, ought to be considered as true descriptions, in the sense that they serve to build an account for the empirical phenomena.

But what is the importance of this premise? Whether one endorses a realist or an antirealist position makes a difference for the construction of an account of observation. In particular, a realist take on scientific theories allows to formulate observation in a way that relies on the theories one endorses for the understanding of a specific set of objects and phenomena. This is particularly important in the case of theories that model also fundamental entities, such as quantum mechanics, which yields predictions that are empirically corroborated via tests on this kind of entities. A definition of observation that allows to rely upon these theories in order to classify a particular interaction as a direct observation is open in principle to the possibility of directly observing such an object. In contrast with this, an account of observation that purports to be based on "theory-independent" criteria, such as unaided human perception, and that does not support a realist take on scientific theories, sabotages at the outset the possibility to consider interactions carried out via complex technology — the use of which is supported by theories — to directly observe particular objects and phenomena. This is the core reason why, in this thesis, a realist take on scientific theories plays the role of a premise; it supports endorsing Dudley Shapere's definition of observation, which is introduced and discussed in the following chapter. Importantly however, this also delimits the extent of this realist premise: it is not the case that all the arguments presented in this work depend on a realist commitment, nor is it the aim of the thesis to produce a contribution to the debate on scientific realism via a study of observation. The only argument that is potentially connected to the realism debate in philosophy of science is the proposal of a view on quantum mechanics advanced in ch. 4, sect. 4.4 and in ch. 5, sect. 5.3.3. Let us consider the influential antirealist position defended by Bas van Fraassen, and what it entails for approaching observation.

At the heart of constructive empiricism (van Fraassen, 1980), is the idea that the scientific account of the world must be empirically adequate, namely, able to provide a coherent picture about the objects that we experience. Empirical adequacy is a desideratum for good scientific theories also from the realist standpoint, but in van Fraassen's constructive empiricism this property is the principal goal of scientific theories, and it does not lead to committing ontologically to the objects and structures described by the theory. The sole things that can be considered real are observable entities, which, for van Fraassen, are those that we can perceive or that we could perceive were we in the right conditions to do so — we will enter in the detail of this account in ch. 2, sect. 2.1.1.

This notion, which is purposefully extra-theoretical — the fact that humans have perceptual abilities is independent of any particular scientific theory — immediately sorts fundamental, microscopic entities among the unobservables. Rather than being the *real* constituents of the world, unobservable fundamental entities are seen by the antirealist as theoretical devices that play an instrumental role for the ability of science to account for the objects of experience. The fact that we do not have the type of empirical access to these entities that could allow us to verify what exactly they are is a limit upon which the antirealist argues that we are not compelled to endorse ontologically the theory's account of such entities. What the antirealist denies is not the *existence* of a fundamental ontology of the things in the universe; they can remain epistemic agnostics with respect to what such a level of reality is and to how to characterize it. What they deny is instead the ontological status of the fundamental entities described by scientific theories.

The equation between observation and unaided perception was a shared view amongst empiricists and realists alike, with some exceptions — representatively, see Maxwell (1962) — until the 1980s. However, in the case of van Fraassen's antirealism, this way of defining observation is also motivated by the fact that evidence that can be obtained through the use of complex technology requires us a realist type of belief also in the theories that underlie the usage of such a technology, that is precisely what is denied to scientific theories in general. The claim of the unobservability of fundamental entities based on such an account of observation plays a pivotal role in the argument of the antirealist.

In contrast to this, some realists have argued that it is totally uninformative, with respect to scientific practice, to limit the scope of observation to the scope of unaided perception. Observation is in fact a process which, in tandem with experimental practice, has the role of gathering and producing empirical data about the objects accounted for by the theory. Hence, it is not possible to make use of an idea of observation which is limited to the abilities of the perceptual apparatus. Rather, within the range of what is observable should be included also what is accessible through —- and only through —- the usage of complex technology. Some of the data about microscopic entities, such as those produced by cloud chambers or particle accelerators, require a great level of interpretation to be “deciphered”, interpretation that is available on the ground of theories. A realist-motivated line of argument includes these entities within the realm of the observable, in order to defend the solidity and reliability of the empirical access we have to such otherwise unreachable entities. In fact, what is called "entity realism" extends the predicate of "real" to the entities that are "manipulable" also by means of complex technology (Egg, 2018).

The argumentative structure of this type of accounts gives an idea of how tightly observation is connected to this group of problems. Such attempts have taken the form of different accounts of observation, and have been used by realists to produce arguments contrasting the antirealists' opposition. Classical reference for this kind of argument is Hacking (1983). For my analysis, to allow for a more permissive definition of observation such as Shapere's has the advantage of not imposing strong constraints as those of van Fraassen's definition at the start; and, traditionally, more open definitions have been elaborated within a realist framework.

Finally, it is true that the point of the antirealist may stand even irrespective of a particular notion of observation. Although the boundary between observable and unobservable can be moved, it still constitutes a boundary: hence, one may still maintain an antirealist stance with respect to whatever is considered unobservable following the new definition. Another reason then to opt for a realist take on theories is to not adopt a priori an ontologically negative position with respect to what is regarded as unobservable.

Virtuously theory-laden. Yet again, one may argue, even if we agree that a realist stance plays in favour of an analysis of observation, it comes with a higher price to pay: if we endorse a definition of observation that is more permissive with respect to the boundaries of the observable, as it is often the case within realist accounts of science, we may not be able then to escape the danger of committing to an intrinsically theory-laden conception of observation. The infamous negative views attached to the latter has its origin in works of Norwood Russell Hanson (1958) and especially of Thomas Kuhn (1996), who are commonly taken as promulgators of a completely relativistic view of science that would fundamentally undermine its objectivity. Letting aside my general disagreement with such a reading of these works, what I am going to highlight is the "other side of the coin" with respect to observation's theory ladennes. Namely, I am going to argue that not only in a sense scientific observation is necessarily theory-laden, but that this theory-ladenness is also a *desirable* feature of observation.

The motivation for stressing this virtuous influence of theory on observation is first linked to the relation between realism and observability. According to the realist, scientific theories provide insight about how the world *truly* is. Through observation it is possible to gain empirical data about some of this latter's fundamental constituents and their properties. The connection between theory and objects of observation is for the realist very strong, as it instantiates a relation of truth that holds between theory and its referents in the empirical world. For this empirical comparison to be possible in the first place, the observation process cannot be neutral and uninformed, but, quite to the contrary, it needs to be infused by the group of theories that are relevant for the object observed. The following are four points on how observation is theory-laden, and why this *ought to be the case*.

- The first is that observation both gives rise to and is supported by the development of theories. The relation between theory and empirical research is dynamical, and is mutually constructive for both parts. In this sense, observation is both grounded on the body of scientific theories, and it grounds in turn such theories too, by being one process through which empirical data is gathered. The influence that theory has on observation is thus part of the *conditions of possibility* of scientific observation. Importantly, this is the case also when observation is not specifically informed by the theory it is providing evidence for. For example, a set of observational data can contribute to the development of a new theory. In this case, the data would not have been laden by the new theory, but by the theoretical background upon which the community relied for carrying out these specific observations. Within the scientific enterprise, observation is always theory-laden, as it relies on theory broadly understood as theoretical background. When considering the level of specific theories, a set of observational data may be influenced by a particular theory while being independent of another.

- Secondly, to be fully accepted as valid, theories need to be empirically confirmed. Observation, by linking theories to the world, contributes to the procedure of theory testing, and hence it is informed by the theory that is being tested at a particular moment. Evidently then, in cases in which the reason for carrying out specific observations is to test the theory in question, observation must be informed very clearly by the theory considered.

- Analogously, the same precision is required within empirical research itself. In order to address observation to isolate and scrutinize accurately specific objects, phenomena and processes, it must be clear what the theory says about these objects in a first place. The informative role of the theory on observation is especially crucial with respect to theories that refer to phenomena or objects that can be observed only indirectly.

- Finally, it is fundamental that observation is informed by the theory to recognize anomalies. Anomaly here is intended as a discrepancy between theory and data. For instance, a phenomenon can be observed to have a *strange* behaviour when the description that the theory provides of this latter on one hand and the data about it on the other present a relevant deviation from one another. When such an area of divergence between theoretical description and data collected is observed repeatedly, it is classified as an anomaly. Another type of anomaly regards the cases in which predictive claims derived from the theory do not find confirmation in the experiments, and are actually falsified by these latter. Sometimes, predictions turn out to be *wrong* because there is an anomaly in the theory that needs to be fixed. Finally, sometimes it is the case that what at first is classified as an anomaly, can, at a later stage, turn out to be recognized as a *new* phenomenon: detected empirically, but not previously predicted by the theory.

⁂ ⁂ ⁂

The premises just outlined define the starting point of the thesis. Let us then pass to another part of this work's conceptual framework, namely, let us get a grip on some terms and concepts that we will encounter consistently throughout the following chapters: scientific theories, evidence, and explanation. The definitions presented here are not to be taken fully rigidly, as they will not be analysed in detail in the present work. The scope of this second half of the chapter is to trace a map where it is possible to locate crucial concepts for the work that comes after. Albeit briefly, exploring these concepts offers already the possibility to point at some of their interesting properties in relation with observation and quantum mechanics.

1.2 Observation's relata

1.2.1 Theory

Physical theories are systems of axioms that are interpreted in terms of models. Such models are connected to physical systems in the world by means of particular relations of idealization and abstraction. Theories may contain information about objects and phenomena of the world, but this is not their primary function. They must be able to yield explanations of physical phenomena on the one hand, and lead to possibly testable predictions concerning these phenomena on the other. Within the latter they incarnate physical principles that make them differ from mere mathematical constructions. It is in the nature of theories to be fallible, and thus subject to constant examination through time.

The nature and structure of theories, and the issue of how they relate to the physical world are perhaps the oldest and most fundamental debated subjects of 20$^{\text{th}}$ Century philosophy of science. The most adequate to characterize the structure of quantum mechanics seems to be by far the semantic view — see Winther (2021, sect. 3), specifically conceived to capture the mathematical nature of physical theories; classical references for this view are the works of Patrick Suppes (1967), and of Bas van Fraassen (1980, 1989) — see also Giere (1988, 2004). Such an account seems particularly fit to describe quantum mechanics, as the outline of the theory itself consists in a specification of particular mathematical structures designed to represent physical states. The formalism would constitute the theory's set of principles, the application of which is then made explicit in the formulation of mathematical models which describe via idealization the behaviour, the properties, and the interactions of systems. A clear outline of how quantum theory is to be understood in terms of the semantic view is done by Richard Healey (2017b, ch. 8).

The version of the semantic view that seems the most appropriate as a basis for the present work is the so called "model-theoretic approach" (Winther, 2021, sect. 3.1.2), which, differently from the "state-space approach", does not identify the theory with a set of mathematical models. According to the model-theoretic approach, a theory is formed by a number of abstract principles that are interpreted mathematically through models, which are in turn put in relation with physical systems. The difference between the two versions is that the "model-theoretic approach" includes a meta-mathematical structural level, which prescribes how the principles are to be interpreted through models.[1] This view is particularly fit for capturing some research in the foundations of quantum mechanics that investigates on the mathematics upon which the theory relies, the implications of using these structures, and the relations between this and other types of theories.

To see the semantic view of theory in the "model-theoretic approach" translated in an example, let us consider Ronald Giere (1988, 62-91). A theory contains a set of general principles, e.g. Newton's second law of motion, the application of which takes place in the construction of a mathematical model, such as model of the linear oscillator, which in turn can be connected to a real physical system by means of what he calls "hypotheses". Models have both the function of instantiating the application of the axioms, and of representing the physical system. Hypotheses work as the link between models and systems; they are linguistic entities that outline which type of relation there is between the physical system and the idealized mathematical model. In this view, hypotheses play the role that correspondence rules have in the syntactic view, where they are supposed to relate observational statements describing the physical system to theoretical statements. However, the character of hypotheses is quite different, as they instantiate a relation of *similarity* between model and system, while correspondence rules are, more appropriately, connectors between different types of propositions, evaluated in terms of truth-making relations.

Quantum mechanics as a theory has been fleshed out and challenged in several ways since its birth, and one result of this has been the development of several *interpretations*. There is an open debate about how the latter should be classified,

[1] The guidelines of this interpretations are discussed within model theory Hodges (2018).

both in relation with the original theory, and as theoretical constructions in their own right. In fact, the term "interpretation", although well established by use, does not do justice to all of these views equally. Importantly, according to the account of theories just recalled, some of these would classify as fully-fledged separate theories. I am referring in particular to those which contain a modification of the set of axioms, though most of them yield the same predictions of the original theory. Alternative quantum theories, in this light, would be Bohmian Mechanics, and the spontaneous collapse models as the Ghirardi-Rimini-Weber, and the Diósi-Penrose model. All of these diverge from standard quantum mechanics in their axioms, as we will see in a little more detail in chapter 5, sect. 5.2.1. Views that maintain the same postulates, but that complement the theory with specific world views on its meaning are QBism, Relational Quantum Mechanics and the pragmatist interpretation. The Many Worlds Interpretation dispenses of the collapse postulate, but maintains the use of Born's rule, and does not, therefore, strictly modify the original axioms of the theory. For this reason, we will consider it an interpretation in this classification. Given the view on theories adopted here, we will call these latter interpretations of quantum mechanics, and the former quantum theories. It must be clarified however that there is a specific debate on how to identify and classify theories within the physical community itself, and that there are alternative positions to the account I am recalling here, which can be defended equally validly. Just to consider one example, theories may be grouped according to the predictions they yield, which would lead to a different taxonomy of the quantum theories and interpretations listed above. The adoption of the account of theory presented in this section is not a stance in this larger and deeper debate.

1.2.2 Evidence

It is called evidence a set of data that leads to inferences in favour or against the theory's claims. Even the sole possibility of testing a theory is of utmost importance for the theory to be accepted in the scientific community. Theories that do not yield any testable prediction are not considered fully-fledged scientific theories by some.

To understand whether a theory correctly describes the world, claims derived from it must be confronted with evidence in order to be confirmed or disconfirmed. The processes that lead to the production of evidence are observation, measurement and experiments. Sometimes it is not possible to produce an uncontroversially decisive verdict about the confirmation or disconfirmation of a hypothesis or of a prediction, but it is possible to specify the degree to which such a claim is confirmed or disconfirmed by evidence. The notions of confirmation and disconfirmation have their limit cases respectively in the concepts of verification and falsification. A certain evidence is said to *verify* a claim when it is conclusive for attributing positive truth value to such a claim, while it falsifies it in the opposite case. The *conclusive* character of evidence that verifies or falsifies a claim might be subject to objections analogous to those moved against a fully correspondentist notion of truth with respect to the theory's claims and their referents in the world. In response to this, the notion of conclusiveness, as well as the notion of truth, may be conceived to come in degrees of approximation. Of course, the sense in which the process of theory confirmation is to be understood depends very much on which role is attributed to the evidence in such a process, and different accounts of confirmation ascribe different roles to the latter – see Crupi (2016).

For what concerns the case of quantum mechanics, confirmation by evidence is accompanied by additional considerations that aid the assessment of the evidential power of certain data. As for other theories that describe unobservable entities, data in quantum mechanical experiments are rather uninformative unless they are accompanied by an interpretation that clarifies their meaning, and, at the present moment, a plurality of quantum theories can account equally successfully for the very same set of data. Hence, each of these theories provides a different account that explains why the data are confirmatory. But in each case, it is not the evidence alone the basis upon which its efficacy is assessed: for instance, as we will see in chapter 4, the loophole-free Bell tests are considered conclusive evidence of a hypothesis that stands on a precise set of premises. Other interpretations of the phenomena involved provide different explanations of the experiments.

Given the space that will be dedicated to the discussion of actual experimental evidence in chapter 4, at this point we can limit ourselves to a general characterization of scientific evidence. At this preliminary level, it is possible to individuate at least five general requirements that data must fulfil in order to be considered fully-fledged scientific evidence; these were clearly outlined in this particular form by Jim Bogen and Jim Woodward in (Bogen and Woodward, 1988, 319-321).

(1) The data considered must be in a form that is accessible to our senses. Even in the case of evidence about unobservable entities, the data used to confirm or to undermine a hypothesis must be formulated in a form that makes them available to human experience.

(2) In order to constitute evidence, data must be reproducible in abundance, and with sufficient frequency. This warrants that it is possible to individuate which are the features that constitute a stable feature of the entities studied, or of particular

phenomena. The stability that some features display by appearing regularly in large ensembles of data-sets constitutes a strong reason for inferring that these features can in fact be considered real elements of the world. It goes without saying that natural phenomena that occur only sporadically, such as the passage of a comet or the extinction of a particular species, make exception to this requirement.

(3) Scientific evidence must be produced in a form that makes it agile to be shared within the scientific community. Data must be formulated and organized in a way that favours their readability, clarity, and that specifies the methodologies through which they have been obtained.

(4) It should be possible to produce a statistical analysis of an ensemble of data. Regularities in the world and the features of these regularities are not objects of experience; we can have access to them through a statistical analysis of an observation or of an experiment repeated a considerable number of times over an acceptably extended period. A statistical analysis of data allows to reconstruct an idealized description of the studied phenomenon and of its features in a purified form from the idiosyncratic factors that occur in the production of each singular data-point.

(5) Data that constitute evidence must be obtained through a control of the amount of error. The techniques of error-control should be well specified and made available together with the set of data that constitutes evidence.

1.2.3 Explanation

Part of a realist stance is to conceive science as a source of knowledge about the world, which allows us to explain natural phenomena and their features. Scientific theories can provide epistemic access to the world by *explaining* natural phenomena. A scientific explanation is a systematic argument that provides structured information about natural phenomena by focussing on the elements relevant for accounting for *how* or *why* a phenomenon comes about.

In general there are two types of explananda: singular matters of empirical fact, i.e. singular instances of events that have taken place in a precise location at a precise moment in time, and regularities in the world, i.e. stable and re-occurring processes and events. The urge for producing an explanation derives from the presence of a question of how a certain phenomenon takes place, or why it does — good reviews of different accounts of explanation in philosophy of science are Salmon (1989) and Woodward (2017).

It has been increasingly acknowledged that it is unreasonable to conceive a comprehensive model of explanation able to fit science in general. The structure of different explanations is in fact relative to the specific discipline within which it is produced, and hence it is in many cases not transferable to other disciplines. Moreover, also different theories within the same discipline may lead to different types of explanation. Nevertheless several comprehensive accounts of explanation have been developed in the 20th century, which can be distinguished on the basis of the methodology through which the explanation would be outlined.

Starting with the renowned *deductive-nomological* model by Carl Hempel and Paul Oppenheim (1965), which worked as a blueprint for the understanding of explanation for decades, other canonical accounts are represented by the *causal* model — see Salmon (1998), and Woodward (2003), the *probabilistic* account, developed for cases in which causality is not a particularly helpful notion, by considering the elements that are statistically relevant for the occurrence of a phenomenon (Salmon, 1971), and the *unificatory* account of explanation (Kitcher, 1989). Despite some particular theories may be better suited with tailored types of explanation — an example is represented by the *geometrical* explanation that would be more properly capturing how general relativity explains (Nerlich, 1979) —, it is undeniable that the notion of causation is one of the most powerful tools we use to understand the natural world, and this is also the reason why the causal account and its variants are the most known and used explanatory models. Famously, however, precisely the notion of causation has been recognized early on as a problematic concept to implement in relation with quantum phenomena such as the correlations among entangled systems.

The reason for this clash is due to the fact that causality has traditionally been understood in relation with macroscopic regularities that can be observed and related to one another in a very extended range of disciplines; from a formal point of view, the scientifically most employed account of causality appears to be the interventionist notion, meticulously formalized for instance by Judea Pearl (2000), notion that is in turn embedded in the physical framework of classical field theory. Within the latter, the physical interactions that underlie the type of relations among events that can be regarded as causal are constrained by the structure of relativistic space-time. In a nutshell, there are particular restrictions on the amount of causal correlations among events, which derive from the fact that physical interactions are limited to occur at subluminal speed, and from the fact that classical causality also assumes that the probabilities of two events, given a common cause in their past, factorize — this is crucial for the derivation of Bell's inequalities, which will be explored in detail in ch. 3, sect. 3.1.1 and 3.1.2. Notions such as the principle of common cause and the Causal Markov Condition encompass these restrictions, and outline precisely the boundaries among what can be considered causally related and what cannot. Crucially, the correlations arising from entangled systems do not come out as causal in such a scheme, fact which has led to claims such as that quantum mechanics would be *non-explanatory* (Salmon, 1990, §6), or that the Bell-correlations would be indeed a type of natural correlation, but of a special type other than causal (Hausman and Woodward, 1999). Nevertheless, a very recent trend within quantum information and quantum foundations has produced a proposal for a generalization of the classical notion of interventionist causation in such a way as to capture the entanglement correlations — see Shrapnel (2017) and references therein. For this reason, a causal explanation of the latter could follow, closing this long-standing history of incompatibility between causality and entanglement.[2]

A declaration of which is the best account of explanation for the entanglement correlations is beyond the scope of the present section. In order to at least provide some orientation as to what an account of explanation should look for in this case, I am going to formulate three different questions that I think must be addressed in such a task.

First, it is necessary to individuate what exactly are the explananda upon which the theory can shed light. One central explanandum immediately related to the experiments we will consider in chapter 4 is constituted by correlations of Bell type, i.e. entangled bi-partite systems that can be measured at space-like separation. These give rise to measurement outcomes that display correlations that cannot be explained by means of classical variables. To which extent does quantum mechanics explain these correlations, and how?

Second, it is necessary to individuate what is the explanans of the considered phenomena. Quantum mechanics intended as the fundamental axioms plus the measurement postulate provides one sort of explanation. But it is well known that that is not the only available theory. Different interpretations of quantum theory and different quantum theories provide completely different explanations, based on the premises underlining each different account. Therefore, perhaps a more precise question to ask could be whether the assumptions involved in the various quantum theories have any consequence on the way in which the phenomenon to be explained is described. And hence, whether it is possible to individuate more than one explanans, and what the difference among those would be. An example as to how this actually is the case is provided in section 5.2.1, where I discuss what I call

[2]This is nothing more than speculative and it may be completely unrelated, but it is interesting to notice that the voice "Causation in Physics" has been added to the *Stanford Encyclopaedia of Philosophy* for the first time only a year before the time I am writing this thesis: Frisch (2020).

"the observation question" with respect to the "minimal observational units" of a Bell test.

Third, it is important to specify what one wants the explanation to accomplish. In other words, it seems that the type of explanation one designs would sensibly change depending on the type of questions asked, and on the type of expectations one has relative to the explanation itself. For instance, any interpretation should be able to explain which kind of correlations we could expect given a certain state, a particular type of system, and the experimental setting. However, answers to *why* these particular outcomes occur, or *how*, are harder to produce.

❊ ❊ ❊

The equipment for embarking in the exploration in front of us is minimal, but sufficient. Our mind is now endowed with a particular theoretical framework that will help to organize the events in the right perspective, preventing us from getting confused about why a certain path is on our track rather than another, and helping us to partially select the objects and scenarios we should look out for along the way. Our tools moreover will become handy when we will have to do with observation in this journey, first as a theoretical concept analysed in the philosophy of science, and then more pragmatically as an actual practice through which we can relate the elegant structures of quantum mechanics to a concrete abundance of phenomena matched with masterful precision by its predictions. Let us then roll up our sleeves, and start off by seeing what is the deal about observation in the quantum domain.

Chapter 2

What’s special about observation in quantum mechanics?

It is possible, in general philosophy of science, to sit comfortably in the ivory tower constituted by a general and rather abstract point of view. As much as it can be an incredibly useful task to address particular problems common to the scientific enterprise as a whole by developing ideas that provide a unified view on a subject, this is not what we are going to do in the present pages. In what follows we will exit this ivory tower, and venture on the lands of concrete and particular examples, where we will put to use the tools we just collected in the former chapter. In what follows I concentrate on observation both in philosophy of science and in relation with quantum mechanics. After having outlined in detail Shapere’s definition of observation, I will justify why I endorse it in this thesis. The aim then is to provide an idea of what are the classical philosophical issues surrounding observation, and how quantum mechanics puts forth additional and independent challenges to this concept. In order to address the thesis’ matter, I discuss directly two representative examples of what can make observation a hard game to play with respect to quantum phenomena, namely, Heisenberg’s uncertainty relations and entanglement. The latter example is particularly pertinent, as interpretational features of entanglement will also be discussed in relation with Bell’s theorem’s significance in the next four chapters.

The matters touched in this chapter have a *framing* role in the thesis, as they provide a more ample view on observation and on issues of observability in quantum mechanics. In this sense, the various points that are touched in the following pages do not constitute a list of the questions that will be addressed in the next chapters. In particular the explanation of the meaning of "theoretical entities", Heisenberg’s uncertainty relations and the notion of principled unobservability are functional to the construction of a broader context to understand the import of my research, but they are not central questions within this thesis. That said, they could of course be explored further in future work.

To tackle observation in philosophy of science I focus on the account outlined by Dudley Shapere in his seminal "The concept of observation in science and philosophy" (1982). Among the distinctive features of this account are the fact that the definition is not based on an anthropomorphic criterion, and that it phrases observation in terms of a transfer of information from object to detector. We will see in more detail all of its properties, but we can already notice how such a view fares naturally with the realist perspective embraced in the previous pages, and with the need of a permissive attitude towards the observable.

The choice of focussing directly on Shapere's definition is motivated by the very objects that are being considered in the present work, i.e. quantum phenomena, which have been studied experimentally via tests on microscopic systems such as photons and electrons. Shapere's definition of observation is the best available — albeit not the only one — among those able to include microscopic entities within the spectrum of the observable, because of its neutrality both with respect to the type of observable objects and in regards to the technology used to collect the data. Rather than focusing on the types of entities that can be observed, he fleshes out in a minimalistic guise the type of *process* that can qualify to yield observational data. As we will see, given the complexity of this matter in quantum mechanics, this type of approach has the advantage of not raising obstacles at the very beginning of this work. By endorsing this definition I wish to show with clarity what distinguishes cases of unobservability in quantum mechanics.

The chapter is structured as follows. Sections 2.1 and 2.2 address respectively the problem of theoretical entities, traditionally associated to the discussion of observation in philosophy of science, and Shapere's definition of observation. After the presentation of the latter, in section 2.3 I discuss two quantum phenomena that are interesting from the observational perspective. In relation with Heisenberg's uncertainty relations and entanglement I advance the claim that these may qualify as cases of "principled unobservability", of which I provide a definition in section 2.4, followed by a brief discussion of its possible implications in section 2.5. This argument is relevant in part for the discussion on how different quantum theories and interpretations characterize quantum data (see ch. 5, sect. 5.2.1), but it does not constitute the main point of focus of this thesis. In section 2.3.1 I introduce the fundamental axioms of quantum mechanics.

2.1 Theoretical entities

Even if it may sound more like a homage to a way to envisage the matter proper of traditional philosophy of science rather than of the way of expression of the contemporary debate, it may be interesting to explore the meaning and import of the expression *theoretical entities*, used to refer to objects postulated by a scientific theory, and not directly observed or about which there is no conclusive empirical evidence — see Andreas (2017). Since a very broad range of objects and their properties falls under the semantic extension of the expression "theoretical entities", it may be helpful to briefly clarify in which sense these quantum systems may be considered as such. Quantum systems usually employed in experiments that aim to display quantum properties and effects are an example of what would be called *theoretical entities*.

Physical[1] theoretical entities are both objects of microscopic or of submacroscopic scale, and entities that are defined by means of functions and relational terms that involve the just mentioned objects (e.g. forces, fields). The standard technical definition of "microscopic" scale uses the micron (10^{-6} meters) as unit of measure, where objects that have a size up to one micron are considered microscopic. My use of the term will be close to this convention, although at times it may be more liberal. The definition has also potential issues: first, also objects that measure a small number of microns would be considered as unobservable according to the same definition, although they do not technically classify as microscopic. Second, when one enters in the matter of different disciplines, the term microscopic assumes different meanings. In the case of quantum mechanics for instance the term is defined in contrast both with macroscopic, and with submicroscopic, i.e., of the order of magnitude of the Planck scale or smaller — see Jaeger (2014) for an interesting discussion of this precise point. Let us now see in more detail two senses of theoreticity.

2.1.1 I: unobservability

The traditional sense in which an object is defined as theoretical is when there is no direct empirical evidence that witnesses its existence or the theoretical description of its functions and properties. When this is the case, the access we have to these objects is given by indirect evidence, and by the characterization provided by the

[1]Also mathematical objects are theoretical entities, but in this section I concentrate exclusively on physical entities.

theory, and it remains a matter of debate whether it is justifiable to commit to the ontological status of these objects or not, regardless of their role for the theory's empirical adequacy. The theoretical nature of these objects is not due to the fact that they are *completely imaginary*; certain evidence may be considered indirect empirical evidence of such entities. Nevertheless, as there may be alternative, equally equally valid explanations for the same evidence, it is not possible to take it as a conclusive proof of the existence and properties of those objects. In this particular sense of "theoreticity" the absence of conclusive evidence is determined by observability. Objects were defined as theoretical entities in this sense within logical empiricism[2] in contrast with the objects of observation, i.e. those perceivable by a human observer. For example, the mean kinetic energy of a set of gas molecules would classify as theoretical as much as the molecules themselves, whereas the temperature of the gas correlated with that energy would be observational. This view is well known, although no longer endorsed in the community, and it helps to set the stage for the present point. As we will see in sect. 2.2 a few pages below, what counts as "direct" observation in the present work is sharply defined in terms of a transfer of information from a source to a receiver, which is not hampered by further disturbances in the process. This characterization does not abide by the restrictive distinction between perceivable and unperceivable. Moreover, later in the thesis — ch. 4, sect. 4.4 and ch. 5, sect. 5.3.3, rather than referring to theoretical entities, for the study of quantum mechanics I refer to the ontology of the theory.

An immediate sense in which fundamental objects are considered unobservable is when observation is equated with unaided perception. As already mentioned, the most relevant contemporary defender of this conception of observation is Bas van Fraassen, who introduced unaided perception as an *anthropomorphic* and *modal* criterion of observation, most famously in "The Scientific Image" (1980):

> The term 'observable' classifies putative entities (entities which may or may not exist). A flying horse is observable — that is why we are so sure that there aren't any — and the number seventeen is not. There is supposed to be a correlate classification of human acts: *an unaided act of perception, for instance, is an observation*. A calculation of the mass of a particle from the deflection of its trajectory in a known force field, is not an observation of that mass. (van Fraassen, 1980, p.15, emphasis added)

Van Fraassen's conception of observation emerges in the "rough guide" he provides in the next page, guide that he then consistently applies throughout both the

[2]Chapter 3 of Lutz (2012) contains a precise exposition of the theory-observation distinction as characterized in Rudolf Carnap's philosophy. For the distinction as formulated by Carnap himself, cfr. for instance Carnap (1956).

present and later works:

> X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it. This is not meant as a definition, but only as a rough guide to the avoidance of fallacies. (van Fraassen, 1980, p.16)

This account of observability has been extensively and precisely worked out in three papers by Frederick Muller (2004, 2005), Muller and van Fraassen (2008). One major quality of these works is that they contain a painstakingly exact extension of the domain of the observable in terms of unaided visibility. Muller (2005, pp.75–82) contains the definition of a scientific criterion of observability based on visual perception. According to this account of observation, entities like electrons and atoms are doomed to be and to remain unobservable, as there is no imaginable scenario in which they may become perceivable. They can be referred to exclusively within the domain of scientific theories, and, in van Fraassen's constructive empiricism, it is not possible to coherently endorse ontological commitment to such entities. Constructive empiricism does not utterly deny these entities' existence, but it rather maintains a neutral attitude towards them. As it was already recalled in ch. 1, sect. 1.1.1, van Fraassen's position is epistemological: it does not deny the existence or the features of a fundamental ontology as characterized in our theories, as its unobservability does not yield either conclusion. Rather, he claims that we do not have epistemic access to such an ontology, and advocates therefore a form of epistemic agnosticism with respect to this matter.

It is worth to underline that the couple "theoretical-unobservable" has not been exclusively paired with a correlated antirealist view. There are in fact traditional accounts of "hardcore" scientific realism that also endorse a view of observation based on human perception, such as Wilfrid Sellars — see in particular Sellars (1963, §440), Sellars (1965) and Sellars (1961). In his philosophy, the distinction theoretical-observational hinges upon a theory of meaning of Carnapian flavour — cf. Carnap (1956), whereby observational terms refer to objects and properties that can be ostensively defined, whereas the definition of theoretical terms depends on the scientific theory that includes them. Importantly, the original theory-observation distinction has the following two features. On the one hand, it was based on a *linguistic* type of investigation, on how scientific propositions were related to the world. On the other hand, it was connected to a specific view of science according to which this latter has the aim to account for the world *how we experience it*. This is a core tenet of the various forms of empiricism, such as logical empiricism and van Fraassen's constructive empiricism; But it was a popular view also among scientific personalities of the 20$^{\text{th}}$ Century, such as Albert Einstein (cfr. Einstein (1936)).

Observation in terms of unaided perception is nowadays no longer a popular view. In the majority of contemporary accounts of science, observation is usually implicitly conceived as extending also to objects that can be reached only by means of complex technology, as the type of entities we are concerned with. The emphasis on the fact that this looser sense of *observational* is eminently employed in the sciences is intentionally marking a distinction with the way in which observation was usually understood in the traditional philosophical literature.

2.1.2 II: theory-dependence

The second sense of theoreticity is related to the first, but it does not necessarily overlap with it. In this second sense, certain entities are called theoretical because in order to have epistemic access to them (i.e. know what they are, understand empirical evidence that involves them, and scientific explanations formulated in terms of them) one must rely upon scientific knowledge. The crux of the matter is that physical theories describe the properties and the dynamical behaviour of the fundamental constituents of matter, upon which depends what happens at greater orders of magnitude. However, it is not possible to describe, study, or have epistemic access through mere experience to these very entities, independently of these theories.

Remaining within the boundaries of physics, one can think of the following example. We all have access to the fact that bodies move, that they move at different velocities, and that their motion can start and be interrupted. Physics (e.g. Newtonian physics) provides a characterization of the phenomenon of motion in terms of the relative masses of the bodies considered, the forces that act between them and the Earth, the meaning of acceleration, momentum, work, etc. In principle it is possible to refer to these events by constructing two different descriptions, one by means of non-theoretical language and one by means of theoretical language. The second is a richer description, that introduces new terms such as force, and redefines existing ones formally, such as mass and energy. These terms refer to the objects and to some of the relations involved in the events of motion. What I intend to stress with this consideration is that in the hypothetical case one never came to know Newtonian mechanics, they would still have epistemic access to the fact that bodies can move at different speeds, and that their motion can be prompted or interrupted. The picture of these very events provided by physics enables one to form a unified picture of these and other seemingly unrelated events, to produce accurate predictions, and to provide an impersonal, intersubjectively verifiable perspective.

The situation is different in the case of microscopic entities, because it is not possible to have epistemic access to these unless one has also knowledge of the theories that refer to the objects one is considering. As we will see in the next section, it is important to reiterate that the data obtained via experiments on microscopic entities constitute our empirical access to these entities. Thus, it is not the case that their theoreticity in this second, linguistic sense, is a consequence of their empirical inaccessibility. In order to be able to consider these data as effective source of empirical access to these entities on the other hand, it is necessary to have a level of physical knowledge that enables one to understand the data as a function of their theoretical characterization.

2.2 Shapere's account

In this section I am going to recall the definition of observation elaborated by Dudley Shapere in 1982. After Shapere's paper, many other accounts have been formulated which can be considered as successors of that definition — for example see Kosso (1989). Some works dedicated to the relation between quantum theory and empirical research endorse an implicit definition of observation similar to Shapere's, such as for instance Dorato (2000), which contains an interesting distinction between measurability and computability; Cordero (2000) deals with the different interpretations of quantum mechanics and their possible empirical differences, while the more recent Jaroszkiewicz (2017), contains an analysis of observation within a much less empirically-oriented approach to quantum theory.

Since the present section has the purpose of providing the context for my analysis of observation in relation with quantum mechanics, I will not produce here a detailed examination of Shapere's work and of its implications. Nevertheless, as this is the definition of observation that I endorse in this thesis, it is worth to closely examine its properties. In earlier stages of the present work I pondered a conjecture for a while regarding the possibility that the apparent "oddness" or problematicity characterizing observation of quantum phenomena could be due to the inadequacy of the very definitions of observation available in philosophy of science. Nevertheless, attempts to modify or to implement a different definition have turned out to be redundant, as they did not add anything actually different from what was already implied in Shapere's definition. For this reason I understood that the problem of measurement and observation in quantum mechanics is not a matter of the definition of observation adopted, but rather of the way in which quantum mechanics pushes us to represent and understand physical systems and interactions.

There are three main reasons why I endorse Shapere's account of observation in this thesis. First, it is minimalistic, as it defines *direct* observation in terms of a physical interaction by relying on two distinct points. In the definition, this interaction is not further characterized because what constitutes an act of observation in each single case depends, as we will see, on what is the theory of the source, of the transmission and of the receptor of information. Different scientific disciplines and also different branches of the same field rely upon different theories, which qualify further these parts of the process. In a way then, the definition can be considered as "empty", quality which characterizes its very general extent. This generality and minimalistic character are particularly fit for a research on observation in a field where the experiments are as complex and debated as in fundamental physics, and in quantum mechanics. Second, Shapere insists on the non-anthropomorphic character of his account, in opposition with other definitions of observation where human perception plays a fundamental and discriminatory role. This point is particularly important in this work because the type of objects and the experiments studied in this thesis have nearly nothing to do with human perception. Electrons, atoms, photons, even the most massive systems used in Bell tests are largely out of the perceivable spectrum. To endorse a definition of observation that would release a priori a decree of unobservability derived from unperceivability did not appear as a philosophically interesting perspective, as it would be unable to capture the subtlety and complexity of the experiments, and of the issues revolving around these latter. Not to mention that it would have also clashed with the generally more permissive use physicists themselves make of "observable" — see the first two points of sect. 1.1 for the relevance of this remark. Finally, despite Shapere's definition is not the only one available, it is certainly among the most acute for its sensitivity towards actual scientific practices, for its precision, and for how accurately the philosophical issues surrounding it have been addressed in the 1982's paper. Its clarity and depth make it still stand out amid the other accounts.

In this section I focus on Shapere's definition and on two of its properties, to pass then to a consideration of what it means for something to be unobservable according to such definition. I proceed after to examine two examples that display some of the peculiarities of observation in quantum mechanics.

The definition:

x is directly observed (observable) if:

(1) information is received (can be received) by an appropriate receptor; and

> *(2) that information is (can be) transmitted directly, i.e., without interference, to the receptor from the entity x (which is the source of information).*
>
> (Shapere, 1982, 492, emphasis in the original)

Given that the pivot of this account is the emphasis on the fact that observation consists of a transfer of information mediated by an interaction, it is very important to get two aspects clearly. First, what qualifies as an observation as opposed to any other type of interaction, and, second, what constitutes a *direct* observation. The first question is answered in Shapere's paper at p. 510, namely, there is nothing in itself that distinguishes the interactions through which we gain direct observational evidence of something from interactions of other types. What qualifies some of these interactions as observational is the use we make of the data we obtain through them. This may appear as a misleading criterion, in light of the effort Shapere puts into making his account especially non-anthropomorphic, as we will flesh out in a minute. But this confusion should be avoided by considering the following: from an epistemological point of view, it is inevitable that the observational data are those that are used as such by the scientific community in the enterprise of scientific research, as it is the need for empirical data motivated by science that justifies carrying out scientific observations in the first place. Since scientific communities are composed by human individuals, the criterion to distinguish between observations and interactions more generally has this pragmatic, and hence epistemically anthropomorphic flavour. As it will be fleshed out in a few lines, the definition of Shapere is still not relying on an anthropomorphic criterion with respect to the *nature* of the interactions between sources and receivers of information. In particular, there is no special need for the receptor to be a perceiving human observer.

The second question of what constitutes a direct observation is answered in the exposition of the case study Shapere considers. A transfer of information from the source to the receiver is direct if it is no mediated by further interactions in between source and receiver. In the description of the Homestake experiment, of which a little bit more will be said below, Shapere clearly distinguishes between the case in which we were to gather information from the center of the sun by detecting electromagnetic radiation, as opposed to high-energy neutrinos. In the first case, the photons produced in the sun's core would have been absorbed and re-emitted several times before reaching the sun's surface and the earth, and the information they originally carried from the sun's center would have been highly altered, or lost (p. 491). High-energy neutrinos on the other hand are systems that have a very low probability of interacting with other systems, and they would thus reach the detectors having travelled through their trajectory undisturbed by further interactions. Therefore, *direct* observation of something is possible when the information that the experiment intends to collect reaches the predisposed detectors without having interacted further on the path towards the latter.

The context in which this definition was introduced is a critique of the classical theory-observation distinction within the philosophy of science. Such a philosophical (logical empiricist) conception of observation was modelled on perception, and what exceeded the realm of the perceivable, which was, *ipso facto*, non-observational, was thus regarded as theoretical. Shapere stresses that this distinction is not endorsed in science, where instead observation is not related to perception in any interesting way, if in any way at all. The first claim strongly put forth by Shapere is thus that the philosophical distinction between theory and observation is untenable. An empiricist conception of science, whereby experience is the measure of objectivity, and theories are built upon such alleged "neutral ground" is simply an incorrect view, which fails to capture what observation in science actually is. He underlines that

> My discussion will show, among other things, that *specification of what counts as directly observed (observable), and therefore of what counts as an observation, is a function of the current state of physical knowledge, and can change with changes of that knowledge.* ... More explicitly, *current physical knowledge specifies what counts as an "appropriate receptor", what counts as "information", the types of information there are, the ways in which information of the various types is transmitted and received, and the character and types of interference and the circumstances under which and the frequencies with which it occurs.* (Shapere, 1982, ibid., emphasis in the original)

I take this to be a clear statement of theory dependence. With the expression "physical knowledge" Shapere refers to the overall knowledge possessed by the scientific community at a certain point in time. What depends on the theory in virtue of being a "function" of it, is both what counts as a receptor of the information that constitutes observational evidence, and what counts as information coming from *the source*. The remainder of Shapere's paper is divided into an analysis of "*the theory of the source, the theory of the transmission and the theory of the receptor*, respectively, of the information" (ibid.). It may be of help to remember that Shapere applies his analysis to a concrete case study, i.e. the "Homestake experiment", performed between 1970 and 1994, that allowed to detect high-energy solar neutrinos through a recording of the decays of chlorine atoms into radioactive isotopes of argon, decays that were caused specifically by such neutrinos. The experiment was designed by physicists Raymond Davis and John Bahcall, and it constituted one of the first empirical findings of neutrinos. The reaction through which it was possible to infer the passage of neutrinos, and to count the neutrinos detected, is the weak interaction between an electron neutrino and an atom of chlorine, which causes the decay of this latter into an isotope of argon and an electron respectively ($\nu_e +^{37} Cl \longrightarrow^{37} Ar + e^-$). Bahcall and Davis Jr. (1982) is the historical account on which Shapere's survey is based (Shapere, 1982, pp. 493–505). Shapere considers the data obtained in this experiment as an instance of *direct observation of the center of the sun*.

In this experiment, the theory on which the observation process is based contains an account of the features of the sun in terms of the particles that constitute it, and their interactions. The theory describes the state of matter in the sun's nucleus, how the neutrinos that are detected in the experiment are produced, how they pass through the external layers of the star without interacting with the other particles, how they reach the earth, and finally in which manner they cause the decays detected in the experiment. According to Shapere, also the very claim that such neutrinos convey information about the center of the sun is sustained by the theory. The objectivity of the evidence obtained in the experiment is justified in terms of the compatibility of the data obtained on the one hand with the theory underlying such experiment on the other.

The theory-dependence that characterizes the observational process is more interesting on a more general level. Each observation is an individual process, defined each time by the theories that underlie it. More explicitly, each time an observation is carried out, the community has first asked *what* information is been retrieved in a particular experiment, which type of interaction will be employed to reach this end, and how the data will look like, given the experimental set up. In the case of the Homestake experiment, the question was not whether it was possible to observe neutrinos as such, or their positions, or their amount in a certain time-span. Rather, it was engineered to create the conditions such that a particular type of atomic decay would take place, which required specifically high-energy neutrinos produced in the center of the sun to interact with the chlorine. In that way, information about the physical conditions of the star's core was retrieved. This work of theory-based definition and agreement, within a community, about what an observation aims at and which data will be accepted as direct observations is crucial for the latter to take place. On the one hand, this means that *what* can be considered as the detector in a particular experiment directly depends on the theory considered, and on the type of information that is object of interest each time. The set up of an experiment that allows to perform direct observations of a certain object, property, phenomenon, is theory-laden in the sense that it is underlain by an agreed-upon theoretical basis that frames which information the community is interested in retrieving, which interaction will be planned, and what will be considered as data. This "basic" level would then constitute a common ground among members of the community, about which they would not disagree with one another. The *interpretation* of the data on the other hand, namely, how these would be explained, given the agreed-upon basis, could differ among members of the community. In other words, given the acceptance that a specific experiment is what ought to be done to gain a certain piece of information, different researchers could endorse different theories explaining the data within a broader theoretical context. As an example, we will see both in ch. 3 and 5 that, whereas the community agrees on the experimental modality to retrieve information about the entanglement correlations at space-like separation, members within it disagree as to how to explain the data, and, in hindsight, on what the data represent. Therefore, there are two levels of theory-ladenness in this scenario. A fundamental one, whereby the community agrees on the theories considered to set up the experiment, and an "epistemological" level, where the explanation of the same set of data may differ depending on which theory supports it.

One important feature of Shapere's definition is its non-anthropomorphic character, under two respects. On the one hand, the receptor is not tied to unaided perception in any important sense, although of course it does not exclude it. When a perceiver in normal conditions of light and with normal visual abilities sees a tree ahead, they are also observing the tree. There is a transfer of information from the source (the tree) to the receptor (the perceiver's eyes), which is caused by electromagnetic waves from the light source reaching the perceiver's eyes after having interacted with the tree. The perceiver's eyes record such information in such a way as to allow the perceiver, on the basis of their conceptual framework, to claim that they are observing a tree. What is important in this example is that the receptor involved, namely, the human eye, is one type of receptor which is sensitive to a determinate spectrum of electromagnetic waves, and that functions efficiently under certain conditions. Organs related to the other perceptual senses can be described as well in terms of their functions as receptors. In this manner, Shapere's move becomes clear when he notes that our sensory apparatus is only one of the many other receptors that can successfully record information coming from given sources. Optical microscopes or telescopes, or radars, antennas, up to complex technological instruments such as particle accelerators, or quantum tunnelling microscopes, can all be rightfully considered detectors *in the same manner* as the human perceptual apparatus. In the light of Shapere's criterion there is nothing privileged nor special about the latter. The usage of any type of technology to obtain information about a specific object as outlined in the definition is thus allowed within the scope of what constitutes an act of observation. It follows that data obtained through any such devices are to be fully considered observational evidence.

The second sense in which Shapere's definition is non-anthropomorphic is the fact that, in principle, a set of data constitutes observational evidence even in the case in which it were never to be used by a human observer. When the data has been recorded, what makes it observational is not its disposition of being perceived by a scientist. Even in the case in which an experiment were to be set up, run, and the data were recorded successfully by a computer and never read, such data would constitute observational evidence; the computer would have completed an act of observation.[3]

Shapere explicitly addresses this point, first by laying out the intuitive claim that

> ... after all, it is *we* human beings who have set up the "appropriate receptor", *we* who will use the received information as information. It follows that whatever information is received through the "appropriate receptor" must be transformed, in a final segment of the apparatus, into humanly-accessible form. Thus if the information comes in the radio region of the electromagnetic spectrum, or via weak interactions, it must be transformed into electromagnetic information in the visual wavelengths, or into audible clicks, or into readable printout, or the like. (Shapere, 1982, 508)

Therefore, it may seem necessary to add a third conditions to points (1) and (2) of the original definition:

> (3) the information is transformed by appropriate devices into humanly-accessible information which is (eventually) perceived (and used appropriately as information) by a human being. (Shapere, 1982, 509)

But he then explains that this ought not be the case:

> whatever the merits of this suggestion, it must be rejected: the fact that *we use* certain information in certain ways for certain purposes — as observations in the role of evidence — must be kept separate from the scientific (epistemic) considerations which lead to its being taken as observational evidence. (Shapere, 1982, ibid.)

[3]This definition of observer would be of primary importance in a recent discussion on an old paradox originally introduced by Eugene Wigner (Wigner, 1995), known as "Wigner's friend paradox", which uncovers possible problems deriving from the universal unitary character of the dynamics of quantum mechanics in a world in which our experiences seem better described by a non-unitary evolution. Cf. Frauchiger and Renner (2018), Brukner (2018), Bong et al. (2020) for some examples of the discussion and the interpretations of the problem.

The interactions that convey information that can eventually be classified as observational evidence are not different from other types of interactions in virtue of their privileged relation with human perception. With Shapere's words,

> The reasons that make certain interactions, and certain receptors, appropriate for use in an observation-situation — that determine what can count as observational — are scientific, and are given by the information detailing (1) and (2); even the problem the observation is to deal with is, in sophisticated science, posed in the light of current knowledge or well-founded belief. Our use of that information, *at least insofar as that use is scientific*, is determined by those considerations; and it is therefore those considerations; that reasoning, that must be emphasized if we are to bring out the role of observation in science.
>
> That conditions (1) and (2), taken by themselves, do not distinguish observations from other interactions is thus not a sign of their failure; on the contrary, *it is precisely the assimilation of observation to the general category of "interactions", and not its use by us, that constitutes the important point in understanding the role of observation in the search for knowledge and the testing of beliefs*; for that assimilation reflects the fact that "observation" has been, or at least has moved far toward being, integrated with the larger body of our best-warranted beliefs about nature. (Shapere, 1982, 509–510, emphasis in the original)

The point made in these paragraphs is that, if there is nothing special about the particular interactions that take place between source and receptor of information that then yield observational data, this special character is to be found within the theories. It is the theoretical framework within which scientists operate that provides the reasons for considering certain data as instances of observation. Ultimately, it is the *use* of this information that "makes it" observational. Therefore, one might add, in practice it is indeed necessary that the data do have an ultimately perceivable form. What needs to be kept in mind here is that it is not this latter bit that defines the observational character of the data according to Shapere.

2.2.1 Note on Shapere's unobservability

It is interesting to notice that in this paper Shapere does not concentrate on the *limits* of the observable that derive from his account. If we go back to the definition, it is possible to derive an analogous definition of unobservability by negation. Therefore,

> *Something is unobservable if it is not the case that information about object x can be received by an appropriate receptor, or it is not the case that it can be transmitted (with or without interference) to the receptor from object x.*

Let us focus on what makes object x unobservable. The first evident factor is represented by technological bounds: the detector may be limited in the sense that it is not able to receive information from the object under study. Since here the issue is not failure of observing object x, but rather the present impossibility to observe the object, the limits of technology to which I refer here are radical. For something to be unobservable in this sense means that the apparatus that may detect information about the presence or the properties of a certain object has not been built, or it has not been designed yet. In other words, this would mean that our current technology is not advanced enough for the study of the considered object. The latter expression can have two meanings. On the one hand, it might be clear in principle which kind of interactions could reveal information about the investigated object, but there might be problems in understanding how to reproduce and control these interactions experimentally. On the other hand, more radically, the theoretical description of a certain object might be clear, but it might be unclear how to obtain empirical information about such an object.

The second factor involved regards Shapere's analysis of the theory-dependence of his criterion. It will be remembered that to be completely non-anthropomorphic meant that the only determinant for declaring something as an instance of observation is the theory itself. This means that a set of empirical data is observational data of a certain object *in function of* the theory that underlies the experimental apparatus, and that provides knowledge of the interactions involved in the transfer of information. In the example of Shapere, the theory that constituted the ground for declaring those detections as atomic decays caused by neutrinos was Enrico Fermi's theory of interaction, which was later substituted by the theory of weak interactions.

In this particular case of theory-dependence the theory has the last word in declaring a set of evidence observational. But it is not the case that the theory contains a clause about, say for example the unobservability of a specific particle of the standard model. If something is considered unobservable in this sense, it is not due to a decision that has been taken at the level of the theory. Rather, it may be due to — perhaps insurmountable — technological limitations. The crucial difference with quantum mechanical examples regards exactly this point: in the quantum realm, certain cases of unobservability seem indeed to be traceable to the axiomatic body of the theory. This is what we are going to investigate now, by means of two representative examples.

2.3 Two cases from quantum mechanics

In the light of what has been treated in the previous pages, I would like to sketch now one of the special properties that characterize observation in relation with quantum mechanics. I call this property *principled unobservability* because the impossibility to directly probe empirically some features of quantum systems appears to be a matter settled in the very heart of the theory. After having outlined two examples, I am going to propose a definition of principled unobservability. Before passing to the outlines of these examples, in the next section I briefly recall the fundamental axioms of quantum mechanics.

2.3.1 Quantum mechanics — the "bare formalism"

Textbook quantum mechanics, standard quantum mechanics and orthodox quantum mechanics are expressions that refer to the set of axioms and postulates that constitute the formal body of the theory. There are multiple ways in which it is possible to spell the theory out, depending on which mathematical structures one privileges — see for instance Landsman (2017) for a comprehensive exploration of the formalism from Hilbert spaces to the algebraic formulation. The most commonly used in philosophical circles is the Hilbert spaces axiomatization, to which I also refer in the present work. The way of exposition follows Wallace (2008),[4] with integrations from Le Bellac (2011) and Isham (1995).

The bare formalism is constituted by the following elements. For simplicity I refer to pure states, but all these points can be easily expressed also for the more general cases. Moreover, it must be noted that the following points refer to isolated systems to allow a simpler outline but that, with the due adjustments, these axioms can be generalized to represent open systems.

- *State space* – A complex vector space, the Hilbert space ($\mathcal{H}$), constitutes the state space of quantum systems. When normalized, vectors in this space represent the possible states of an isolated system.

- *Structure of state space* – An "additional structure" on Hilbert space $\mathcal{H}$ is required to allow the description of systems. Such a structure can be provided in two

[4]The number of references for the axiomatization of quantum mechanics is an embarrassment of riches, to say the least. I have chosen Wallace's reference as his introduction of the formalism is philosophically motivated: since in this piece he is then going to argue for the advantages of the Everettian interpretation, it is in his interest to provide the purest outline of the original formalism. As this is also what I need, to keep my reference to "quantum mechanics" separate from the theory's interpretations, this seemed a convenient choice.

different manners. Either via a preferred set of basis vectors on the space, or by a preferred decomposition of the system into subsystems. The structure can also be provided by both these specifications together.

- *Quantum observables* – The observable quantities of a system are represented by self-adjoint operators that act on $\mathcal{H}$.

- *States and observables* – The spectral decomposition of a quantum observable operator shows the relation between the values of a certain observable and the state of an isolated system. For instance, observable operator $\hat{A}$ with eigenvalues a_n has the following spectral decomposition:

$$\hat{A} = \sum_n |n\rangle a_n \langle n|,$$

where $|n\rangle \in \mathcal{H}$ represents the operator's eigenstates. If a system is in state $|n\rangle$, then if we measure observable $\hat{A}$ of a system in that state, the measurement will yield exactly the value a_n. Empirically, this means that if a measurement of the quantity represented by $\hat{A}$ is performed on a system in state $|n\rangle$, the outcome will be equal to value a_n with certainty.

- *Dynamics* – These are the rules that determine how an isolated system evolves through time. The dynamical law is specified on $\mathcal{H}$, instantiated by a set of unitary transformations, and it is constituted by Schrödinger's equation, here recalled in its time-dependent form, and with reference to generic state $|\varphi\rangle$, that evolves the state of quantum systems unitarily:

$$i\hbar \frac{\partial |\varphi(t)\rangle}{\partial t} = \hat{H}|\varphi(t)\rangle. \qquad \text{(Schrödinger's equation)}$$

The unitary transformations are usually represented by means of the Hamiltonian operator ($\hat{H}$), that represents the total energy of the system.

- *Empirical predictions* – the bare formalism needs to be paired with a notion of measurement, which has the role of connecting what happens at the quantum microscopic scale to macroscopic empirical processes. In other words, given a specification of what is supposed to take place at the scale of electron and photons, a notion of measurement has to be able to show what are the correlated macroscopic phenomena, or measurement outcomes, that can be registered and analyzed. Measurements and their outcomes are respectively the empirical process and the empirical data that can be accessed in practice in the context of quantum mechanics. Observation is strictly connected to such processes.

Quantum mechanics yields predictions about what is the probability, upon measurement, of obtaining a certain outcome given the state of the system. There is a probability amplitude a of finding a system that is in state $|\varphi\rangle$ in a particular state $|\chi\rangle$, denoted as a scalar product on Hilbert space $\mathcal{H}$: $a(\phi \rightarrow \chi) = \langle\chi|\phi\rangle$. The probability of finding the system that is in state $|\varphi\rangle$ in a particular state $|\chi\rangle$ upon measurement, namely $p(\phi \rightarrow \chi)$ is found by applying *Born's rule*, which consists in calculating the square modulus of the probability amplitude defined above:

$$p(\phi \rightarrow \chi) = |a(\phi \rightarrow \chi)|^2 = |\langle\chi|\phi\rangle|^2. \tag{2.1}$$

There are two possible formal representations of measurement, respectively called "projection-valued measurement" PVM and "positive-operator-valued measurement" POVM. Both represent the act of measurement by means of particular self-adjoint operators, which assign a probability to the possible measurement outcomes. The technical differences between these two formal constructions of are not of focal interest for the present context, but what is important to notice is that this notion of measurement does not intrinsically have an immediate physical meaning.

The interaction between quantum systems and the measuring device is characterized by peculiarities. I am here going to mention only two types of elements that seem to mark a discontinuity between the macroscopic world and the quantum scale. First, quantum systems seem to display types of behaviors that cannot be connected by means to analogy to similar phenomena at the macroscopic scale. Phenomena such as superposition, quantum jumps, or entanglement, just to mention three examples, are not something that was thought to be part of the empirical world before quantum mechanics was formulated. This means that the theory purports to account for physical phenomena that are not only empirically directly inaccessible, but that also require a set of special concepts in order to be fully understood.

2.3.2 Heisenberg's uncertainty relations

The first example is represented by a core theorem of quantum mechanics, namely, Heisenberg's uncertainty relations. The original idea was introduced by Werner

Heisenberg in 1927, in the article "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik" (Heisenberg, 1927),[5] but the formalization in which the relations are commonly known derives from Earle H. Kennard (1927), who formulated rigorously the inequality describing the measure of uncertainty between position and momentum, and from Howard P. Robertson (Robertson, 1929), who generalized Kennard's inequality to all self-adjoint operators.

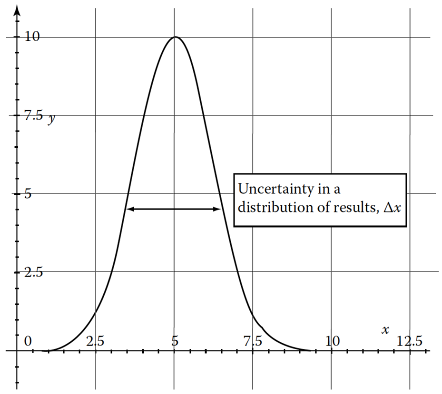

Figure 2.1: The uncertainty in a set of position measurements.

The theorem applies to specific pairs of quantities, i.e. "canonical conjugate quantities", which are mathematically related by Fourier transformations. To understand in which sense the relations define a notion of uncertainty, let us consider the following. If a series of measurements of position is performed on a system in a position superposition, the measurement outcomes will display a range of different values for the position of the system, with different relative frequencies associated to each value — some will occur more frequently than others. When a system is in a superposed state, it is usually said that the value of the observable considered is indefinite.

Figure 2.1 (Allday, 2009, p. 246) shows a possible probability density of an ensemble of position measurements performed on systems in a superposed state. On the x-axis are the position values that have been measured, while on the y-axis are indicated the number of occurrences of each value. Value 5 is associated with the greatest probability density. It is crucial to notice that such distribution of occurrences for the position values of this hypothetical system can be formally predicted by means of the quantum formalism *before* any experiment is carried out. The procedure is eminently mathematical, and its backbone structure consists in applying Born's rule to the module of the wave-function that represents the state of the system under study. In the case of a superposed state, the application of Born's rule permits to derive a probability distribution of the values that are going to be measured if an experiment is performed. In this example, value 5 would be the expectation value for the position observable of the system.

However, also all the other values comprised between one and nine have been measured a certain amount of times. The data displayed in the diagram can be

[5]"The Physical Content of Quantum Kinematics and Mechanics", English translation from J. A. Wheeler and W. H. Zurek, see Wheeler and Zurek (1983, pp.62–84).

considered consistent with a superposition of position eigenvalues. The crucial point of such a representation is what is referred to as the "uncertainty" of the distribution of the results, which is used to refer to the standard deviation of the distribution of position values, indicated with σ_x, where x indicates the position variable.

Generally, in an experiment it is possible to perform measurements of more than one quantity at once: for instance, it is possible to measure at the same time the spin and the position of an electron. In the case of classical systems there is no limit to the precision with which such measurement can be carried out: one can measure at the same time many of the physical properties of, say, a body moving at a certain velocity. If one performs measurements that are sufficiently accurate, one could find the exact value for the position and for the momentum that a system had at a specific moment in time. However, this is not the case for quantum systems, and the radical difference pertaining to these latter is indicated by Heisenberg's uncertainty relations.

For any couple of self-adjoint operators that represent canonical conjugate quantities, there is a specific relation of commutation such that the uncertainty in the values of each of the two observables can reach only a certain level of precision at the same time, which cannot be overcome. Position and momentum are one of these couples of canonical conjugate quantities. Let us consider the diagram above. The position of the particle is characterized by a certain standard deviation σ_x. If we performed also a measurement to determine the momentum of such a particle, we would find a certain standard deviation for the momentum too, indicated by σ_p.

Now, what Heisenberg's uncertainty relations determine is that it is not possible to reduce the standard deviations of both of these quantities at the same time to an arbitrarily small value. Mathematically, this means that there is a lower bound for the commutation relation of the two quantities. Empirically, this means that if one tries to increase the precision with which the position measurement is performed on a system, one will obtain a correlated more imprecise value for the momentum of the same system, and *vice versa*. The values of position and of momentum of the same system can be determined at the same time with a maximal precision regulated by parameter $\hbar/2$:

$$\sigma_x \sigma_p \geq \frac{\hbar}{2} \tag{2.2}$$

To appreciate in which sense this inequality sets a boundary to the precision with which each of these observables could be determined, let us consider the following. When a system is in an eigenstate of a precise value for the observable considered, the standard deviation for that observable is equal to zero, which means that there is no uncertainty with respect to the value of that observable. For instance, going back to the diagram above, if ten position measurements were performed on a system in a position eigenstate with eigenvalue equal to five, the graph representing the results would display an area on the plane with coordinates of a value closely approximating 5, and 10 — evidently, in this idealized example we do not take into account the error of the measurement apparatus. There would be no significant standard deviation of the measurement results.

If, however, for the same system we were to measure the momentum with arbitrary precision, things would be different: more precisely, in a subsequent measurement of position, the system would not be found exactly in the position predicted on the basis of its being in a position eigenstate, but the measurements would have revealed a range of values around the most precise one. Privileging the momentum measurement has as a consequence a loss of precision in the determination of position. The interesting question for us is: what does this theorem imply for observation? The *meaning* of the uncertainty relations is a matter of dispute that has not been settled unanimously among the philosophers of physics. In this case too, different quantum theories provide different accounts for what the relations imply at a physical level. There are mainly two classes of possible interpretations. On the one hand, it seems that the uncertainty relations individuate an *epistemic* limit: we have access to the physical properties of quantum systems only with a maximum level of precision. This means that quantum systems do have precise values of these properties at all times, but the theory determines the exact limits of our epistemic access to these values. On the other hand, it has been claimed (this was also Heisenberg's position), that the uncertainty relations describe an *ontological* feature of quantum systems: these latter do not have definite values for these properties at all times. In other words, quantum systems do not occupy a precise position, they do not move with a precise momentum, etc.

For what concerns the peculiar consequences that this theorem has with respect to observation it is not necessary to engage in the debate and to side with one of these two classes of interpretations. The limits *of observation* entailed by this theorem are in fact independent of an epistemological or ontological interpretation. First, what is observed, what *can* be observed, is precisely limited according to the parameter defined by the uncertainty relations. Regardless of the measuring apparatus one may wish to use, the precision gained in the localization of one of these properties will irreducibly correspond to a loss of precision in the experimental determination of the other property. The "picture" that we can get of matter at the quantum scale is irremediably out of focus, with a *fuzziness* proportional to $\hbar/2$. Second, and more importantly, Heisenberg's uncertainty relations are not derived from inductive reasoning based on empirical results, but they are derived from the purely mathematical formalism of quantum mechanics. What this implies for observation, *as a consequence*, is that it is not possible to obtain arbitrarily precise observational data of the canonical conjugate quantities, if measurements on them are performed at the same time. What is special of this particular limitation is that it is not due to any fault of the various measuring apparatuses we can employ to obtain these data. Nor is it a matter of factual limitations we may have for building more powerful or more precise measuring apparatuses. Sharp values for the non-commuting observables of quantum systems are *unobservable in principle*, as they derive from the theory.

2.3.3 Entangled systems

Empirically, two or more quantum systems become entangled when, after having interacted in a specific manner, once they are separated in space their individual states cannot be defined independently from one another. As a consequence of this particular state, the properties of these systems become correlated. As for the case of the uncertainty relations however, entanglement arises in the first place from the mathematical formalism. As we will see also in chapter 5, the formation of entangled states derives from the way of formalizing quantum mechanics. The correlations among entangled systems are not classical, and their features have been object of discussion for decades, starting with Einstein et al. (1935). The expression "entanglement" was coined by Schrödinger in his Schrödinger (1935, 1936). For the purposes of the present example, I focus on the experimental effects that characterize entanglement, namely, the correlations among the measurement outcomes and the fact that they can be reproduced also when the entangled subsystems are at space-like separation.

Let us take the simplest example of an entangled state, namely, the *singlet state* commonly used to define two non-identical systems A and B, of spin$-1/2$. Their entangled wave-function is the following (from Popescu and Rohrlich (1998, p.33)):

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_A|\downarrow\rangle_B - |\downarrow\rangle_A|\uparrow\rangle_B)u(x_A)v(x_B) \quad (2.3)$$

where $u(x_A)$ and $v(x_B)$ represent the two systems' wave-functions with respect to their spatial coordinates. When a measurement of spin is performed on this pair along any axis, here we will refer to the z-axis for convention, if measurement on A yields the outcome spin "up", B will be necessarily found in a spin "down" state, and vice versa. Whereas, if A is found in a spin "down" state, B will yield a spin "up" outcome. The two systems are perfectly anti-correlated. However, it is not possible to predict with certainty which one of the two possibilities will arise. Each possible value, namely, $|\uparrow\rangle_A$ and $|\downarrow\rangle_A$ for A, and analogously for B, have in fact the *same probability* of obtaining upon measurement.

The fact that the two entangled systems can be separated in space is relevant, because it shows that the correlations are not dependent on the distance at which the systems are located. The influence that the modification of the state of A exerts on the state of B is instantaneous, which, as we will see in chapter 3, can be cast out in a variety of interpretations.

The special character of quantum correlations resides in the fact that there is no way of reproducing them via the use of classical systems under the same conditions. The modification of state that the systems undergo when they interact with the measuring apparatus is quite relevant. Once a certain value of spin has been found, if a new measurement is made along the same axis, the outcome will be the same. If sufficient time passes however, and, after this time, another measurement is carried out, the system may display another spin value. If this scenario is mirrored for the other system, it will be seen that such a system displays values that are not correlated anymore with the other member of the pair. For this reason, it is generally claimed that the action of measurement on entangled systems causes the disruption of the relation of entanglement that had connected them up to that point.

The major issue is that it is not clear why, in which manner, and at which stage of this process, the act of measurement causes the instantiation of definite properties on systems that were described as having indefinite values for these very properties. The radical difference in the evolution of quantum states seemingly caused by the act of measurement represents a classical problem in quantum mechanics, which falls under the umbrella expression of *measurement problem*.

Let us now consider what is observed in experiments on entangled particles. An experiment can be conducted on pairs of electrons prepared in a singlet state, i.e., a state in which the two electrons' spins along the z-axis have a total value equal to zero, and the values for the spin-observables ("up" and "down") are perfectly anti-correlated (see eq.(2.3)). Ideally, a measuring apparatus that can yield results confirming the nature of the singlet state is composed by a source in which systems are "prepared" in an entangled state, two detectors, which are the parts of the measuring apparatus that interact separately with the two electrons, and pointers connected to the detectors, that display values linked to a particular spin state (e.g. value equal to zero for outcome spin "down", and value one for spin "up").

The scope of running the experiment is to understand whether the electrons are in an entangled state or not, and in which state they are. Therefore, the setting must be arranged in such a way as to allow the understanding of (a) the type of correlations between the two systems and (b) whether the modification in the evolution of a particle's state influences instantaneously the state of the other.

For verifying (a), it is important to know that the source is going to release each pair of electrons at the same time, and that this can be done with assessment of the preparation. If this point is verified, then for electrons in a singlet state the pointers' positions at the end of the experiment will need to always be either on value 0 on one side and on value 1 on the other, or vice-versa — of course, in the idealised scenario in which we do not take into account experimental error. Moreover, to show that the ensemble is effectively in the singlet state, it is also necessary to measure along other axes as well, and to analyse the outcomes of measurement along those axes, together with the outcomes of measurements along z. This secures that the outcomes are not due to a classical state, rather than quantum. For assessing (b), one can decide to locate the detectors at the same distance from the source, and at space-like separation from one another. Our final set of data is composed of a great number of spin measurement counts, which we will find to be perfectly anti-correlated, as each reading of the two pointers' positions will constitute the ordered pair either of value "0, 1" or "1, 0".

If the experiment is performed correctly, the empirical data are solid and reliable empirical evidence that confirms the features of the entangled state as outlined by quantum mechanics. The two electrons are both in a superposed state of spin "up" and "down", and their individual superposed states are also correlated via entanglement. The set of data described above can be considered as an empirical confirmation of all these features of entangled electrons.

But can this set of data be considered also *observational* evidence of the entangled state in which the pairs of electrons are? I argue that they are not. The structure of my argument is very simple: the entangled state, as described in quantum mechanics, has the constitutional feature that it cannot be preserved unaltered in an interaction with the measuring apparatus. The electrons in each entangled pair, described by a unique wave function that assigns a range of values to the property of spin before measurement, behave as singular systems when they interact singularly with the apparatus, and what is measured is a definite value for the property considered. If the act of measurement causes the disappearance of the entanglement relation among the elements of the pairs, what is observed consists *only* of the values registered upon measurement, and not of the entangled state that describes the pairs of electrons before measurement. It seems reasonable to claim that information provided by the measurement outcomes warrants indeed an inference regarding information about the state of the pairs of electrons before measurement. But I do not think that this *ipso facto* justifies concluding that thereby we are also obtaining *observational* data of the pairs before measurement. My contention is that what is observed is only what is effectively measured, and that the conclusions that can be drawn about the entangled state before-measurement have only theoretical nature.

First, the interaction between entangled system and apparatus is a complex issue in quantum mechanics, which is at the core of the renowned "measurement problem". Previously I have referred to the fact that in order to gain empirical significance, the quantum formalism needs to be integrated with a postulate regarding the behaviour of quantum systems when they interact with the measurement apparatuses. By means of specific mathematical procedures it becomes possible, given the quantum description of systems, to predict which measurement outcomes are going to obtain and with which probability value. The important aspect is the already noted fact that before the measurement is carried out, the two electrons' entangled state formally assigns more than one possible value ("up" and "down") for the same property (spin) at the same time; the value of the state before measurement is thus considered indefinite. The interaction between the entangled system and the apparatus leads to the instantiation of one possible outcome which *was not there* before the measurement was carried out, and when the measurement is completed, the systems are no longer entangled. Hence, as the presence of such sharp values is possible only when the relation is disrupted, it is not possible that measurement outcomes represent an instance of observation *of the entangled state*. Rather, such measurement outcomes only convey information about it, and they constitute evidence of the presence of an entangled system.[6]

Second, there is a logical aspect of the observability of the entangled state that

[6]It must be noted that it is possible to construct an operator for which the singlet state itself is an eigenstate, and that it is possible to measure such an operator as well, e.g., by keeping the two subsystems close to one another, i.e. not at space-like separation, or by verifying the perfect anti-correlations. This challenges the idea that the entangled state is unobservable in principle. Such an operator is non-local, which may add further interpretational problems of how to understand such a state from an empirical perspective. Nevertheless, it must be made clear that the argument presented in these pages applies clearly to measurements where the entangled subsystems are spatially separated, and that the case just introduced shall be discussed in separate work.

supports the claim of its unobservability. It is undeniable that the observed correlations are predicted with high accuracy by quantum theory, and that it is possible to derive very precise probabilities of the outcomes based on the very mathematical notion of the entangled state. Hence, the fact that the observational data confirm the predictions may imply that the mathematical notion of entanglement is a correct part of the theory. But the fact that the theoretical description of the entangled state yields observable predictions does not bear any implication concerning the *observability* of the state itself. Rather, since the production of the measurement outcomes requires the break of the entangled system, either the system is in an entangled state, or it is not any more, when such outcomes obtain. Therefore, in the hypothetical eventuality that it were possible to observe the entangled state, it seems not logically possible that observational data of it would have been represented by the data that we obtain in an experiment of the type sketched above, because the presence of the entangled state implies the *absence* of definite values as those that are instead obtained upon measurement. In chapter 5 we will also see another aspect of the unobservability of entanglement, namely, the fact that if measurements are made at space-like separation, the theory of relativity prevents the possibility of observing both at once, as the relativistic structure of space-time implies that all interactions are bound to take place at subluminal speed.

So, what are the data? They are not direct observations of the entangled state because the information of the original state is not carried to the receiver. What each single measurement reveals is a spectral value for the state of the ensemble, which says nearly nothing about the state *as a whole*: only the set of final data can be used to then reconstruct the original state. But, more importantly, the state reduction takes place at space-like separation in the experiment we have considered and, since all evidence we have about possible interactions in the world is local, the alleged non-locality involved in this phenomenon is not empirically attainable, but can only be postulated as a means to explain the data. We will return on these notions in ch. 3, while discussing the interpretations of Bell's theorem, and in ch. 5, while focussing on the nature of the entanglement correlations. For the purposes of the present point what is important is that the entangled state's direct observability seems to be hindered in principle by the way in which the theory itself characterizes entanglement and interactions with quantum systems.

2.4 Principled unobservability, a definition

In section 2.2 I have introduced Dudley Shapere's definition of observation and a correlated notion of unobservability. I have then provided two cases of unobservability borrowed from quantum mechanics. In the light of these latter, I am now going to finally introduce an explicit notion of "principled unobservability", which I think is the type of unobservability that characterizes specifically the quantum examples outlined above. My claim is that both exact simultaneous values for the couples of non-commuting observables to which Heisenberg's uncertainty relations apply, and the entangled state, are properties of quantum systems that are unobservable in principle. What I mean is that such unobservability, which is the impossibility of obtaining observational data of these properties, is a direct consequence of the theory's formalism.

The following is a proposed definition of principled unobservability.

Object X or property P is unobservable in principle if

(a) it is not possible to produce observational evidence of X or of P, and

(b) this impossibility is directly implied by the theory itself.

Point (b) is what distinguishes this type of unobservability from the unobservability implied by Shapere's definition, and it is the particular feature that characterizes the two examples from quantum mechanics outlined above.

The first example considered above regards Heisenberg's uncertainty relations, which imply that it is not possible to measure simultaneously and with arbitrary precision quantities such as the position and the momentum of quantum systems. Point (a) is related to this example as follows. There is indeed empirical evidence concerning an approximate value of the position and of the momentum of the system considered, but they are in a form analogous to the diagram reported above, i.e., they are affected by a margin of imprecision. Sharp simultaneous values regarding these particular quantities are in principle unobservable, and this point stands still regardless of which interpretation one may choose to ascribe to the uncertainty relations, namely, whether epistemic or ontological. In a paper where they discuss about the *epistemic* limits implied or imposed by quantum mechanics, Charles Cowan and Roderich Tumulka (2014) consider two interpretations of quantum theory, namely, the two forms of GRW theory and Bohmian mechanics, and provide a thorough analysis of what the formalism implies with respect to what can be measured of quantum systems at the microscopic level, given the ontology assumed in each of these interpretations. With respect to the Heisenberg's uncertainty relations (point 10, pp. 411-412), they underline that whether such a theorem implies an epistemic limit to our access to the world depends on the quantum theory one endorses. In particular, this theorem sets an epistemic limit in the case of Bohmian mechanics, where particles do have definite values for each property at all times, but it does not imply the same in the case of GRW theory, because in this case, between spontaneous collapses quantum systems do not in fact have definite values for such properties. This discussion is related to the present discussion on observability, but it does not overlap with it: whereas it seems impossible to discuss about *epistemic* access to the world without also referring to a clear ontology about the world, in the case of *observation* the consideration of what the fundamental ontology of the world may be is not a prerequisite for carrying out the analysis.

The second example regarded entangled systems, and it was argued that what is unobservable in that case is the entangled state. Point (b) is again clear with respect to this example as, it was argued, it is a consequence of how entanglement is characterized in the theory that it is not possible to obtain observational evidence of the entangled state; interaction with a measurement apparatus causes in fact the disruption of the entanglement relation between the systems involved. The data we obtain are *indirect* evidence of the entangled state because they oblige to make an inference about what it was that caused them, which is, the systems in the entangled state as described in the theory. The reason for this is that *the theory* predicts that the interaction between the systems and a measuring device causes a radical modification in the evolution of the state, such that this state is not preserved intact. This case differs from the scenario in which indirect evidence of something is the only option because of *technological* limitations: in the case presented here, it is the theory itself to impose that direct observation of something is unattainable, regardless of the technology one may put in place.

2.5 Avenues to general consequences: a speculation

Does principled unobservability imply an "absolute impossibility" of observing the object or the property under study? And, what does the answer to this question imply? The answer to the first question is, intuitively, "no". This type of impossibility is defined as a consequence of the theory, and, since the beginning of the past century, when quantum mechanics was formulated, it has been true that the properties described in the two examples above have remained unobserved and unobservable, and that it is not clear whether it may be possible at all to overcome this limitation, and how. Given the content of the theory and what it implies for how we observe quantum systems, the idea to overcome the boundaries of observation set by it seems to be something not even possible to be grasped in imagination, unless in contradictory terms. What seems to be the relation between principled unobservability and the theory is a rather strong disjunction: either the theory is true, and X/P is unobservable in principle, or the theory is *false*, and X/P is not unobservable in principle. Of course, this is not the only reason why a theory would be declared false, but only one of them. And, the strong alternative presented here could occur in cases in which the theory's take on object's unobservability were derivable from the theory's core structure, as I argued is the case for the uncertainty relations and entanglement. The disjunction brings attention to the fact that there may arise a stark contrast between the object of observation and the theory, specifically in a case in which such an object would be declared to be unobservable in principle. Let us take the example of the uncertainty relations: if, at some point, an experiment was designed that did in fact allow the simultaneous measurement of the pairs of conjugate observables — position and momentum, energy and time, spin along different axes, etc. — with arbitrary precision, it would not just mean that the theory contained an imprecision that could be "fixed" via a simple update. This empirical result would generate a contradiction with the manner in which the theory envisages the relation among the observables considered, which would mean that the theory could be falsified. In other words, the observability of an object previously declared unobservable in principle is *incompatible* with the theory from which this unobservability was previously derived. The implications of this fact may be interesting from a philosophical perspective for a study of theory assessment, and for a sharper understanding of the relation between particular theories and the phenomena they account for.

Another consideration that arises from the themes of this chapter connects with a more general question at the basis of this thesis to which I hinted in the introduction. The notion of principled unobservability shows that the dialectic between theories and the world is somehow "reversed": when something comes out as unobservable in principle as the consequence of a theory, the latter is directly modelling an *empirical* limitation to our relation with the world. Bear with me: of course, if this notion of unobservability is derived from the theory, it is not just a matter of an empirical impossibility; as much as squaring the circle in Euclidean geometry is not a matter that is up to my free will or to my hard work, managing to measure the conjugate observables with arbitrary precision is not a matter of using a more accurate detector. Nevertheless, the fact that the latter mathematically do not commute because each function is the Fourier transform of the other *does have empirical consequences*. The reason why this fact is interesting is that it sheds light on the richness of the relation between scientific knowledge and the world. The genealogy of scientific theories wants that in the beginning was the observation of empirical regularities, from which general considerations were derived via induction and could then be refined in different ways. Our theories would then be tested, via the derivation of predictions and hypothesis, and the confirmation or falsification of these would contribute to the thories' refinement and to an increase of our knowledge of the world. But this notion of principled unobservability seems to hint at the fact that theories are not limited to a form of a somehow "passive" description of the world, but to a very active operation of tracing boundaries, defining properties, and *moulding* the empirical realm from which they had been generated in the first place. Whether these limits and details of the world are revealed, modelled, made possible, or if they just mirror the abilities as well as the limitations of the human way of thinking is not something that I am trying to stipulate here. What I want to underline is that the relation between scientific theories and the world seems to be not completely captured by the discovery of the empirical world as something that enables and enriches our theories, but it may need to integrate also the fact that the study of our theories reveals aspects that sharpen and, in turn, trace the areas and boundaries of the world.

One may wonder however if there is something particular about quantum mechanics that allows to build cases of principled unobservability. Perhaps this theory is affected by specific interpretational issues that facilitate reference to examples that can be phrased in terms of this type of unobservability. Recent studies in the philosophy of science about indirect methodologies for theory confirmation should control this doubt. In particular with respect to more fundamental theories such as different proposals for a quantum theory or gravity and string theory, and with respect to theories in the cosmological domain, recent research has developed models of *non-empirical theory confirmation* — see Dardashti et al. (2019) and Dawid (2013) for a general overview and a particular account targeted to string theory respectively. The fact that the objects modelled by the type of theories studied in these works may *never* be empirically reachable has required a revision of the criteria for theory assessment, that are crucially based on a particular use of indirect evidence as a sufficient support for the corroboration of hypotheses. One representative example of what this indirect evidence could consist of is constituted by analogue experiments, the potential and epistemic power of which are assessed in recent works Thébault (2019), Evans and Thébault (2020); for a critique see Crowther et al. (2019). What these works show is that a deeper study of principled unobservability may become an interesting contribution in alignment with these recent research programmes, the potential of which has not yet been fully explored.

❊ ❊ ❊

The chapter started as an exploration of the terrain of concrete examples that extends outside the "ivory tower" of general arguments. Nevertheless, whereas the philosophical analysis of observation has indeed led us to a certain level of deepness, the examples of quantum phenomena considered at this stage have admittedly been discussed still on a fairly general level. The curious reader who may wish to gather more precise and more tangible examples from the heart of quantum theory will soon be satisfied. In the following two chapters in fact we will fully enter in the quantum realm, by studying a theorem of foundational character that has contributed to define the unique and somehow alien profiles of entanglement, a pervasive feature and phenomenon of the theory that elegantly eludes the familiarity of our classically oriented intuitions about the natural world. After having outlined the theoretical aspect of the theorem, we will explore how the quantum predictions have been connected to the natural world. We will come back to the now known lands of the philosophical analysis after having walked the sharp trails of Bell's theorem.

Part II

The Case Study

Chapter 3

Bell's theorem — Theory

3.1 What is the theorem?

Anyone who has become curious about Bell's theorem has become familiar with some of the slogans usually associated with it. For instance, it is said that the theorem proves that the world is non-local, or that we ought to give up some of our common intuitions, that we share especially with classical physics, among which is the idea that a unified description of all matters of fact at once is possible, and to the very least it does not depend on whoever may produce it. But who then ventures more deeply into the details of the theorem will soon find themselves in a thick forest, where more knowledgeable venturers would look with nostalgia at the naivety of the newbie who was fascinated by those slogans, and sigh "would that it t'were so simple".

Despite the formal result is praised for the simplicity of its derivation, it would be an understatement to say that its interpretation does not match with this mathematical elegance. A great deal of research has gone into trying to understand both he assumptions and the implications of the theorem, and this is only part of the story. Given its logical structure — Bell's theorem belongs to the family of no-go theorems — the *meaning* of the result is not automatically suggested by the logic of the derivation. The reason is very simple: a set of assumptions is specified, from which is derived a conclusion which is contradicted by another theory. Well, the mere rejection of the conclusion points at the failure of the conjuncted assumptions, which in turn means that *at least one* among the latter is rejected. But which one is not indicated by the contradiction, hence the melancholic sighing of the expert venturer of what discloses more and more as a theoretical forest. Alternatively, one could challenge the correctness of the theory that generated the contradiction in the first place. Such a theory, in our case, is quantum mechanics. In the early days of the theorem this was perhaps ventilated as a possibility, but an increasing number of confirmations of the success of the theory has made this option non-viable, at least until quantum mechanics is falsified. Nevertheless, the violation of the Bell inequalities would withstand the "fall" of quantum mechanics, and would thus continue to be an explanandum of future theories.

Until then, the task of physicists and philosophers is to understand what the theorem stands on and what it implies, depending on which of the assumptions is given up, which varies depending on the theorem's interpretations. The task is made particularly hard because the derivation does not start by assuming utterly unreasonable premises in the first place. To the contrary, Bell's assumptions are rather reasonable in the sense that they fare harmoniously with deeply rooted intuitions of our being in the world as humans. This means that none of them *in itself* displays a red flag that invites to its dismissal. Yet, the now diffused interpretations of the theorem are at times quite vocal with respect to which is the faulty conjunct. Their verdicts are reached via arguments that necessarily involve additional considerations justifying why that particular assumption has to give.

As each interpretation is forcefully biased in view of the additional claims on which it rests, in what follows I am not going to start from any of these in my exposition of the theorem. Rather, my task is to present it in its overall formal features, and to how different interpretations are composed. In order to do that, I rely conspicuously on Cavalcanti and Wiseman (2021), Wiseman and Cavalcanti (2017) and on other works that present the argument in a similar fashion. My aim for doing this is to provide a fairly complete overview on the theorem's structure and meaning, which enables the philosophical reflections I advance at the end of the chapter.

These pages are structured as follows. The chapter opens with the proof demonstrated by John Bell in 1976 (sect. 3.1.1), which is interesting also because it opens a window on Bell's own philosophical ideas with respect to quantum mechanics and the world. A derivation of the CHSH inequality is offered in section 3.1.2, followed by the violation of the CHSH inequality in section 3.1.3. In order to shed some light on how these terms are used in the present chapter and in the thesis, a note on *factorizability*, *locality* and *no-signalling* closes the first half of the chapter (section 3.1.4). The second half explores the theorem's interpretations (sections 3.2.2–3.2.5), organised according to which assumptions (section 3.2.1) they give up. Finally, a brief conclusion paves the way for chapter 4, devolved to the exploration of Bell tests and what we have learned from them. I have chosen to separate markedly the theory part of Bell's theorem from the experimental research on it, in order to facilitate an exposition that would stress the mathematical character of the theorem on one hand, and highlight and relate together the separate philosophical issues that surround theory and experiments on the other.

3.1.1 How it went down in history: Bell (1976)

In this paper, Bell shows that quantum mechanics is a theory that violates the relativistic notion of what he called *local causality* — see sect. 3.1.4 for how the related concepts are used in this thesis. In order to define it with precision, he introduces the notion of *local beable*. The term denotes referents of the theory that can be described classically, i.e. within the conceptual framework of classical physics. It was first introduced in Bell (1973), where he reflects about the empirical significance of quantum theory, and shows dissatisfaction with respect to Paul Dirac's idea that quantum mechanics would be limited to both provide a characterization of, and make predictions about, *measurement outcomes*. Beables make their appearance in this context, when Bell claims that the theory, ideally, should be able to say how things *are*, rather than just how they are *if measured* (p. 41). He specifies a clear relation between beables and observables, "observables are *made* out of beables" (ibid.), which he also reiterates later:

> The beables must include the settings of switches and knobs on experimental equipment, the currents in coils, and the readings of instruments. 'Observables' must be *made*, somehow, out of beables. The theory of local beables should contain, and give precise physical meaning to, the algebra of local observables. (Bell, 1976, p. 52)

Although he does not provide what could be considered a definition of the term, he clarifies further its scope when he indicates that beables help distinguishing between "'physical' and 'non-physical' quantities" (*ibid.*). Non-physical quantities in this context are understood as "convenient but inessential mathematical device[s] for formulating correlations between experimental procedures and experimental results" (Bell, 1976, p. 53). Among these devices appears, notably, the wave function. One important feature of beables, which follows from their being *physical*, is that they are supposed to have values independently of any measurement being carried out on them (Goldstein et al., 2011, section 6); in a sense, beables are conceived as the *objective* features of the world.

In the context of the exposition of the theorem, Bell is interested in *local* beables, i.e. "those which (unlike for example total energy) can be assigned to some bounded space-time region." (1976, p. 53), because it is in function of these that he defines local causality. It is important to specify that it is in relation with the *relativistic* notion of an event located in space-time that Bell phrases his notion of locality. In particular, the causes of an event in space-time can be located only within the backward light cone of the considered event, and not outside it. If two events are correlated, either one is in the backward light cone of the other, or they share a common cause located at the intersection of their backward light cones. This mimics precisely Reichenbach's principle of common cause, while relativistic space-time structure imposes a temporal order, and it can be used to postulate further restrictions for the spatial and temporal locations of causes with respect to their effects — causes cannot be located *outside* the past light-cone of their effects. This is exactly what Bell starts with in his exposition of local causality (1976, p. 54): he considers two space-like separated evens, that thus are located outside each other's light cones, even though these latter overlap in their respective pasts. In this scenario, local causality is defined in probabilistic terms. Let us follow the diagram drawn by Bell, reported in fig. 3.1.

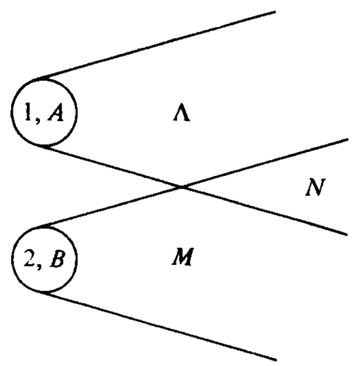

Figure 3.1: Representation of beables A and B in their respective space-time regions 1 and 2. In this diagram, the x-axis is the temporal axis where the past-to-future direction goes from right to left, whereas the y-axis represents the spatial dimension. Within the respective backward light cones of A and B, and within the intersection of these two, there is a complete specification of beables Λ, M and N respectively. (1976, p. 55)

He posits a set of beables Λ, to which it is possible to assign certain values which imply a unique probability distribution for the values of another beable A. The probability that A will assume certain values is conditional on the values assumed by Λ: $p(A \mid \Lambda)$. A is localized in space-time region 1; to explain how local causality works, Bell posits another beable, B, situated in space-time region 2. A and B are spacelike separated: N is a *complete specification* of the beables belonging to the intersection of the past light cones of regions 1 and 2, where A and B are located. Λ is a complete specification of the beables in the portion of A's light cone that does not overlap with the light cone of B. "In a *locally causal theory*

$$p(A \mid \Lambda, N, B) = p(A \mid \Lambda, N) \tag{3.1}$$

whenever both probabilities are given by the theory." (1976, p. 54) This means that what happens in the space-time region of A is probabilistically independent from what happens in the space-time region where the light cones of A and B do not overlap. In other words, the values that A assumes depend only on the values assigned to the beables in its past light cone.

This definition of local causality is crucial for understanding the proof of the theorem. Bell shows that quantum mechanics does not respect the assignment of probabilities represented in eq. (3.1), and that therefore it violates local causality. As far as quantum mechanics is concerned, the specification of the relevant beables in the past light cone of region 1 *is* complete. Thus, Bell asks,

> could it not be that quantum mechanics is a fragment of a more complete theory, in which there are other ways of using the given beables, or in which there are additional beables — hitherto 'hidden' beables? And could it not be that this more complete theory has local causality? Quantum mechanical predictions would then apply not to given values of all the beables, but to some probability distribution over them, in which the beables recognized as relevant by quantum mechanics are held fixed. (Bell, 1976, p. 55)

What the theorem powerfully shows is precisely a negative answer to these questions. Let us enter in the matter of the proof: I am now going to present the way in which Bell introduced his famous inequality.

He starts by assigning a precise bound to beables A and B, namely,

$$|A| \leq 1 \qquad |B| \leq 1$$

According to the probabilistic formulation of local causality (3.1), he then specifies the conditional dependence of the probabilities of A and B on the beables within their backward light cones:

$$p(A, B \mid \Lambda, M, N) = p(A \mid \Lambda, M, N, B)p(B \mid \Lambda, M, N) \tag{3.2}$$

which, because of eq. (3.1), is equal to:

$$p(A \mid \Lambda, N)p(B \mid M, N) \tag{3.3}$$

The inequality that Bell derives in this paper has the form of the famous CHSH inequality (Clauser et al., 1969). Importantly, the steps required to derive precisely that inequality involve some assumptions, now known as *parameter independence* and *outcome independence*, according to the terminology introduced by Abner Shimony (1990). The passage from parameter to outcome independence is left implicit by Bell in the paper we are considering now. The meaning of this expression is clear in relation with figure 3.1: the probabilities of what happens in space-time region 1 depend exclusively on the events located in the backward light cone of that region, and, *mutatis mutandis*, the same holds for space-time region 2. if there are correlations in the values of beables A and B, they must be due to some specific value-assignment of beable N, which would play the role of common cause.

Now, if A and B can assume only a finite number of values, the expectation value of their product can be expressed in form of a sum:

$$p(\Lambda, M, N) = \sum_{A,B} AB\, p(A \mid \Lambda, N) p(B \mid M, N) \tag{3.4}$$

This can be generalized to the case in which the possible number of values is infinite, by means of integration.[1] In order to derive an inequality that does not reach the maximal bound of the sample space, which would in principle prevent a potential violation, it is necessary to consider an alternative specification of beables Λ and M, indicated as Λ' and M'. These are the beables contained in the non-overlapping backward light cones of regions 1 and 2 respectively. The expectation values of A and B, indicated as $\overline{A}$ and $\overline{B}$, are bounded:

$$|\overline{A}| \leq 1 \qquad |\overline{B}| \leq 1 \tag{3.5}$$

The relation between p, $\overline{A}$, $\overline{B}$ and Λ', M', N is then the following:

$$\begin{cases} p(\Lambda, M, N) \pm p(\Lambda, M', N) = \overline{A}(\Lambda, N)[\overline{B}(M, N) \pm \overline{B}(M', N)] \\ p(\Lambda', M, N) \pm p(\Lambda', M', N) = \overline{A}(\Lambda', N)[\overline{B}(M, N) \pm \overline{B}(M', N)] \end{cases} \tag{3.6}$$

By applying (3.5) to the above, the following inequalities obtain:

$$\begin{cases} |p(\Lambda, M, N) \pm p(\Lambda, M', N)| \leq |\overline{B}(M, N) \pm \overline{B}(M', N)| \\ |p(\Lambda', M, N) \pm p(\Lambda', M', N)| \leq |\overline{B}(M, N) \pm \overline{B}(M', N)| \end{cases} \tag{3.7}$$

[1] In Bell's notation (*ibid.* p. 56), the integrated form of the expectation value is $p(\Lambda, M, N) = \overline{A}(\Lambda, N)\overline{B}(M, N)$, which in this case does not add particular property besides a generalization to the case of a continuous interval of values.

By applying (3.5) once again, one obtains

$$|p(\Lambda, M, N) \pm p(\Lambda, M', N)| + |p(\Lambda', M, N) \mp p(\Lambda', M', N)| \leq 2 \tag{3.8}$$

The inequality above already starts to resemble the familiar CHSH inequality, but the derivation involves a few more steps. Bell considers the case in which it is possible to further specify beables Λ, M and N as follows:

$$\Lambda \equiv (a, \lambda) \qquad M \equiv (b, \mu) \qquad N \equiv (c, \nu)$$

The interesting variables of this specification are a, b and c, which, in a quantum-mechanical context, represent "variables which specify the experimental set-up" (Bell, 1976, p. 57), while λ, μ and ν are hidden variables, thus it is possible to average over them. The expectation value of a, b and c becomes then

$$p(a, b, c) = \overline{p\left((a, \lambda), (b, \mu), (c, \nu)\right)} \tag{3.9}$$

As prescribed by the condition of local causality, the probability distribution over λ and ν is independent of b and μ; by considering an alternative specification also for a', b', λ' and μ', it is possible to write the following two inequalities

$$|p(a, b, c) \pm p(a, b', c)| \leq \overline{|p((a, \lambda), (b, \mu), (c, \nu)) \pm p((a, \lambda), (b', \mu'), (c, \nu))|} \tag{3.10}$$

and

$$|p(a', b, c) \mp p(a', b', c)| \leq \overline{|p((a', \lambda'), (b, \mu), (c, \nu) \mp p((a', \lambda'), (b', \mu'), (c, \nu))|} \tag{3.11}$$

from which, in connection with (3.8) it is possible to finally derive Bell's *locality inequality*:

$$|p(a, b, c) \mp p(a, b', c)| + |p(a', b, c) \pm p(a', b', c)| \leq 2 \tag{Bell inequality}$$

The derivation is now complete: any theory that satisfies the condition of local causality, expressed in probabilistic terms in eq. (3.1), will predict correlations that satisfy (Bell inequality).

3.1.2 CHSH

The version of the theorem we just considered has made history not only for the physical import of the proof, but also for the philosophical concepts that Bell employs in his derivation of the inequality. The notion of beable has received a great amount of attention in the literature, as well as other philosophical insights of Bell's understanding of quantum mechanics and its empirical implications. A faithful reproduction of how the theorem has been proven by its original author is therefore not only interesting, but perhaps insightful to understand it in the broader context of its author's thought.

Nevertheless, for the purposes of a broader overview of the theorem's meaning and interpretations, there is a less conceptually driven and formally more explicit derivation to which I am turning now. From a formal point of view the simplest derivation of the theorem is constituted by the CHSH inequality — named after Clauser, Horne, Shimony and Holt (1969), which applies to a bipartite system the two parts of which are measured separately with a choice among two measurement settings, and where each measurement can give one among two possible outcomes as a result. Importantly, the proof can be generalized to a scenario involving a multipartite system measurable via more than two choices, and outputting one among more than two outcomes. For reporting the derivation I am here referring to the proof of Shimony (1990). The notation I will adopt now and for the rest of the thesis is the following: A and B refer to measurement outcomes, x and y to their respective settings, λ to a complete specification of variables in space-time region constituted by the overlap among the light-cones of the events in which the outcomes are recorded (cf. fig. (3.2)). The CHSH inequality has an important property: not only it is violated by quantum mechanics, but it is possible to show the *maximum* quantum bound by which the inequality can be violated. In particular, in section (3.1.3), the quantum prediction is derived by choosing an entangled state that maximally violates the inequality. This is relevant as it hasn't been possible to derive analogous quantum bounds for more general cases.

The scenario considered by Shimony is the following. A source emits an ensemble of pairs of systems in a complete state λ in opposite directions, and each member of the pair enters in an analyser with certain parameters (x and y respectively) among which the experimenter can choose to perform the measurement. State λ is part of a space Λ of complete states, accompanied by the assumption that it allows to define probability measures over it. The possible measurement outcomes on each side are indicated by A and B, with values $\{-1, 1\}$. Knowledge of state λ, together with a

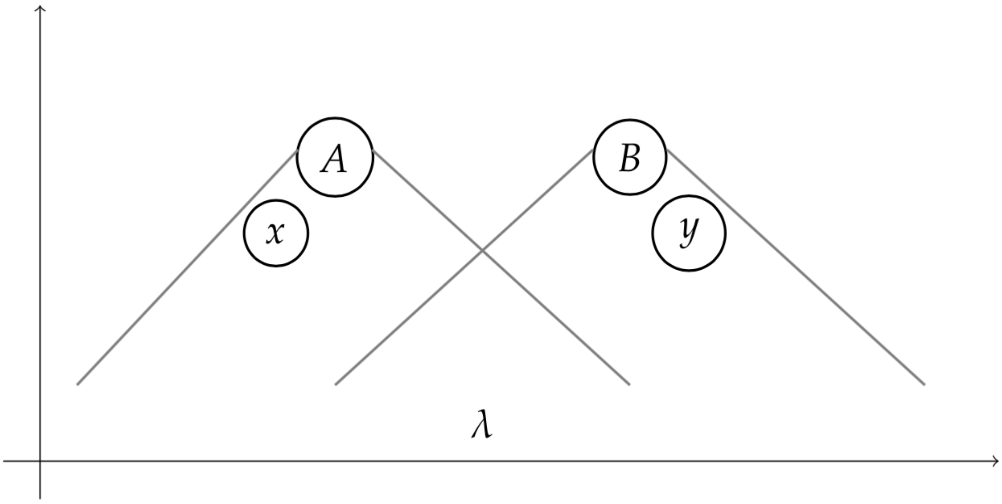

Figure 3.2: Space-time diagram of a bipartite Bell scenario. The vertical and the horizontal axes indicate the temporal and spatial coordinates respectively. A and B stand for the measurement outcomes located in space-time regions defined by the two circles enclosing the letters; x and y stand for the choices of measurement settings, while λ indicates a *complete* specification of the variables at a space-time region located in the common past of either outcomes and settings.

precise specification of the measurement settings, ensures that the probabilities of both the singular and joint outcomes are well-defined (p. 34). In order to reach the desired constraints on the probabilistic dependencies both among the outcome and the setting variables, Shimony starts by introducing the product rule for conditional probabilities:

$$\begin{aligned} p(A, B \mid x, y, \lambda) &= p(A \mid x, y, \lambda) p(B \mid A, x, y, \lambda) \\ &= p(B \mid x, y, \lambda) p(A \mid B, x, y, \lambda). \end{aligned} \tag{3.12}$$

At this point he spells out two conditions on how the measurement settings on each side of the experiment relate to the outcomes on the opposite side, and also how the outcomes on both sides relate to one another. In fact, the events (setting or outcome) located at either side of the experiment do not depend on the events located at the other side. He names these two *independencies* with the famous terms of parameter and outcome independence (Shimony, 1990, pp. 34-35), although the same conditions had been individuated also earlier by Jon Jarrett in (1984), with the names of *locality* and *completeness*.

Parameter independence ensures that the outcomes on one wing of the experiment are probabilistically independent of the experimental settings — the parameters —

on the other wing of the experiment:

$$\begin{aligned} p(A \mid x, y, \lambda) &= p(A \mid x, \lambda) \\ p(B \mid x, y, \lambda) &= p(B \mid y, \lambda). \end{aligned} \tag{3.13}$$

Outcome independence imposes that outcomes on one side of the experiment are probabilistically independent from the outcomes on the other side:

$$\begin{aligned} p(A \mid B, x, y, \lambda) &= p(A \mid x, y, \lambda) \\ p(B \mid A, x, y, \lambda) &= p(B \mid x, y, \lambda). \end{aligned} \tag{3.14}$$

The satisfaction of this constraint rules out the possibility that the outcome on one side may somehow influence the events on the other side of the experiment, again, at superluminal speed. In the reference frame in which space-time regions 1 and 2 are space-like separated and on the same space-like hypersurface, the two events take place simultaneously, thus, one would *instantaneously* exert its influence on the other. Parameter and outcome independence are relevant because it is from those that follows the conditions of *factorizability* (Myrvold et al., 2019, sect. 3.1.2) — which appears simply as "Bell's locality condition" in Shimony (1990, eq. (1), p. 35) — and is defined as follows: from equations (3.12), (3.13) and (3.14) we derive

$$p(A, B \mid x, y, \lambda) = p(A \mid x, \lambda)p(B \mid y, \lambda). \tag{Factorizability}$$

This condition appears also in Bell (1976, p. 56), in the form of equations (3.2) and (3.3) above, and it is the central core of Bell's theorem. This equation captures in fact a very powerful set of intuitions at the basis of the theorem's assumptions, which derive both from the conceptual framework of classical physics and from relativity, as we will see when we will consider the assumptions more thoroughly in sect. 3.2.1. The more one found natural and unquestionable the independence among the events as formulated in this condition, the more one was going to be astonished by the blatant negation of the Bell or CHSH inequalities by quantum mechanics. But to give up *local causality*, which is the manner in which Bell called the factorizability condition was not the only option open. There is in fact another assumption necessary in order to derive the inequalities, which is referred to as "freedom of choice", or "*no-superdeterminism*". According to this condition it is reasonable to assume that the choices of settings x and y are *not* correlated with the outcomes in any relevant manner. In other words, it is not possible that the event of choosing a particular measurement setting for the experiment carried out e.g. at time t is correlated with the outcomes in such a way that they share a common cause in the sense of Reichenbach's principle, which was recalled in the previous section. This principle ensures that the choice of measurement is as free as it is necessary for it to be uncorrelated with the outcomes of that measurement. As specified in Bell et al. (1985, p. 101), this assumption does not rule out the fact that the backward light cones of the events at which the choice of experiment is made overlap, as in fact they do. It does refuse the possibility that this needs to be counted as a cause that relevantly influences the selection of parameters, instantiating a correlation between such a selection of the parameters and the measurement outcomes. It is interesting to notice that this rejection is based on reasonableness, rather than on a physical state of affairs. Nevertheless, "Unless we proceed under the assumption that hidden conspiracies of this sort do not occur, we have abandoned in advance the whole enterprise of discovering the laws of nature by experimentation." (ibid.) This is discussed further at p. 94 below. As we will see in a moment, *the CHSH inequality is derived from* (Factorizability)*, and from "no-superdeterminism"*.

After having specified the product rules for conditional probabilities (eq. (3.12)) and the factorizability condition, Shimony defines the expectation values (E) attached to the measurement outcomes as follows:

$$\begin{aligned} E(x,\lambda) &= \sum_A p(A \mid x,\lambda)A \\ E(y,\lambda) &= \sum_B p(B \mid y,\lambda)B. \end{aligned} \tag{3.15}$$

The equations show the expectation values for a measurement carried on either particle, given complete state λ and the respective parameters. Importantly, the joint probability for the expectation values is given by:

$$E(x,y,\lambda) = \sum_{AB} p(A,B \mid x,y,\lambda)AB \tag{3.16}$$

which, by (Factorizability) is equal to:

$$E(x,y,\lambda) = E(x,\lambda)E(y,\lambda). \tag{3.17}$$

This means that the joint probabilities for the measurement settings factorize into a product of the singular probabilities, and can therefore be considered as independent of one another.

For the derivation of the CHSH inequality a crucial assumption is the following specification of the scenario. To recapitulate briefly, we are considering an ensemble of systems identically prepared in a bipartite entangled state that can be measured by choosing among two settings (denoted as x, x', y, y'). For each setting, the measurement can yield in turn two measurement outcomes (namely, $A = \pm 1; B = \pm 1$).

We consider now the following correlations among the outcomes given the choices of setting, which, for the premises and the scenario considered above, bound the joint expectation values as follows:

$$-2 \leq E(x, y, \lambda) + E(x, y', \lambda) + E(x', y, \lambda) - E(x', y', \lambda) \leq 2 \tag{3.18}$$

The possible combinations of results are sixteen. In the case in which all the signs of the inequality are positive, one obtains exactly twice a value of four, twice a value of minus four, and twelve times a value of zero. This would mean that in no scenario it would be possible to violate such inequality, or, more importantly, it would be impossible to distinguish between classical or quantum correlations.

It is now the time for the final step of the derivation: by integrating the inequality above over the state space Λ, a defined probability measure $d\lambda$ is used as a factor that weights the probabilities; probability distribution $d\lambda$ is justified in relation with the physical scenario considered: "Because of the uniform experimental control of emission, it is reasonable to suppose that there is a definite probability measure [$d\lambda$] defined over [Λ] which governs the ensemble of pairs; but the uniformity need not be such that [$d\lambda$] is a delta-function, i.e., that every pair of the ensemble is in the same complete state [λ]." (p. 34) For the integration, $d\lambda$ is normalized $\int_\Lambda d\lambda = 1$, and the ensemble expectation value is defined as: $E(x, y) = \int_\Lambda E(x, y, \lambda) d\lambda$. One then obtains CHSH inequality (Shimony, 1990, p. 36):

$$-2 \leq E(x, y) + E(x, y') + E(x', y) - E(x', y') \leq 2 \tag{CHSH inequality}$$

Factorizability and no-superdeterminism: a physical situation in which correlations obtain that violate the bound at either end of the inequality is a situation that cannot be described on the grounds of such assumptions. In particular, such a violation implies that at least one of the assumptions is refused. The correlations that obtain among entangled subsystems have precisely the property of violating the bounds derivable from such assumptions, and the demonstration of this fundamental fact about nature is exactly the seminal import of Bell's theorem. The manner in which Shimony describes the theorem is in function of the independence conditions:

> The theorem therefore asserts that *no physical theory satisfying the specified independence conditions can agree in all circumstances with the predictions of quantum mechanics*. The theorem becomes physically significant when the experimental arrangement is such that relativistic locality prima facie requires that the independence conditions be satisfied. Because such arrangements are in principle possible (and, in fact, actually realizable, if certain reasonable assumptions are made), one can restate Bell's Theorem more dramatically as follows: *no local physical theory can agree in all circumstances with the predictions of quantum mechanics*. (p. 32, emphasis in the original).

In order to complete the exposition of the proof, we need now to show how quantum mechanics violates the inequalities.

3.1.3 Violation

Bell's theorem is a powerful result because it inexorably uncovers a deep property of quantum mechanics. The inequalities derived in the previous two sections would not be of any particular interest on their own, if all our current physical theories shared the assumptions on which they rest, namely, factorizability and no-superdeterminism. As we will see in the next section, these two assumptions are not the whole story: in particular factorizability rests on the acceptance of further assumptions that depend on classical physics and on relativity. The reason why the derivation of the Bell inequality does not amount merely to an interesting exercise is precisely the fact that quantum theory systematically frees itself from the constraints that such an inequality incarnates.

The CHSH inequality displays the bounds on expectation values for each measurement outcome given the particular choice of measurement considered in the above scenario — namely, on each wing there are two measurement settings, each of which can yield two outcomes. In a quantum scenario, these measurement settings are represented by observable operators, mathematically represented by Hermitian operators that act on vectors (representing states) in Hilbert spaces. For instance, an observable operator A with eigenvalues a_n, has the following spectral decomposition:

$$A = \sum_n |n\rangle a_n \langle n|. \tag{3.19}$$

In this equation, $|n\rangle$ represents the operator's eigenstates: if a system is in state $|n\rangle$, then if operator A acts on that state, it will yield exactly eigenvalue a_n. Empirically, this means that if a measurement of the quantity represented by A is performed on a system in state $|n\rangle$, the outcome will be equal to value a_n with certainty.

In the scenario considered by CHSH, each of the operators a, a', b, and b' associated to the measurement settings have eigenvalues ± 1. The eigenstates can therefore be chosen to belong to a complex Hilbert (sub)space of two dimensions. With respect to standard basis vectors $|0\rangle$ and $|1\rangle$, a general unit vector has the following form:

$$
\begin{aligned}
&|\psi\rangle = |\vartheta, \varphi\rangle := \cos(\vartheta/2)|0\rangle + e^{i\varphi}\sin(\vartheta/2)|1\rangle \\
&\vartheta \in [0, \pi) \\
&\varphi \in [0, 2\pi).
\end{aligned}
\tag{3.20}
$$

Now, since the scenario we are considering involves only two parties with two choices of measurement each, it is sufficient to consider vectors on two axes of the Bloch sphere, and in such way, angle φ has value zero by convention (see Jaeger (2009, p. 14)) so eq. (3.25) above simplifies to: $|\psi\rangle = |\vartheta\rangle := |\vartheta, 0\rangle = \cos(\vartheta/2)|0\rangle + \sin(\vartheta/2)|1\rangle$. Moreover, the unique state orthogonal to $|\vartheta\rangle$ in the equatorial plane is $|-\vartheta\rangle := |\pi - \vartheta, \pi\rangle = \cos((\pi-\vartheta)/2)|0\rangle - \sin((\pi-\vartheta)/2)|1\rangle = \sin(\vartheta/2)|0\rangle - \cos(\vartheta/2)|1\rangle$.

The operators of the CHSH scenario can be represented as follows, by choosing two axes on the equatorial plane of the Bloch sphere that are in a computationally helpful relation with one another. In particular, we choose $|0\rangle$ and $|1\rangle$ for representing the two mutually orthogonal vectors associated with operator a, and vectors $|+\rangle$ and $|-\rangle$ for operator a', which can be expressed in terms of the angles just specified in the previous paragraph, as indicated below:

$$
\begin{aligned}
a &= |0\rangle\langle 0| - |1\rangle\langle 1| = |0\rangle\langle 0| - |-0\rangle\langle -0| \\
a' &= |+\rangle\langle +| - |-\rangle\langle -| = |\pi/2\rangle\langle \pi/2| - |-\pi/2\rangle\langle -\pi/2| \\
b &= -|\pi/4\rangle\langle \pi/4| + |-\pi/4\rangle\langle -\pi/4| \\
b' &= |3\pi/4\rangle\langle 3\pi/4| - |-3\pi/4\rangle\langle -3\pi/4|.
\end{aligned}
\tag{3.21}
$$

Finally, we choose an entangled state, in particular, the singlet state already discussed in chapter 2, eq. (2.3):

$$
|\Psi^-_{AB}\rangle = \frac{1}{\sqrt{2}}(|+\rangle_A|-\rangle_B - |-\rangle_A|+\rangle_B). \tag{Singlet state}
$$

The state represents a pair of systems, and A, B stand for the two wings of the experiment. Now, in order to calculate the expectation values indicated in the CHSH inequality, it is necessary to apply Born's rule, which, in the general form is $E_\rho(a) = \mathrm{tr}(a\rho)$, where a denotes the observable operator and ρ the density operator representing a general quantum state, either pure or mixed. Since we are considering a pure state, the application of Born's rule to the first case, namely, for operators a and b, looks as follows:

$$E_{\Psi^-_{AB}}(ab) = \langle \Psi^-_{AB} | a \otimes b | \Psi^-_{AB} \rangle. \tag{3.22}$$

We have to calculate the expectation values of each combination of operators, by using the values indicated in (3.21), and sum them as indicated in CHSH inequality. We obtain thus:

$$E_{\Psi^-_{AB}}(a, b) + E_{\Psi^-_{AB}}(a, b') + E_{\Psi^-_{AB}}(a', b) - E_{\Psi^-_{AB}}(a', b') = 2\sqrt{2}, \tag{Violation}$$

which violates in this case the upper bound of CHSH inequality.

For the derivation of this violation it hasn't been necessary to make reference to any particular physical system, nor to any particular property: the physical referents of the observable operators considered have remained open. This detail is not of secondary importance for the understanding of Bell's theorem's extent: the claim that *quantum theory does not abide by the constraints entailed by the conjunction of factorizability and no-superdeterminism* does not state an empirical fact, rather, it really describes a structural feature nestled in the very heart of the theory. In other words, the theory violates the mentioned constraints in virtue of its particular representation of nature. The exquisitely formal and theoretical nature both of the derivation of Bell's inequality and of its violation by quantum mechanics is a point upon which we will return again in chapters 4 and 5, where a closer reflection upon this fact in connection with observation will be advanced.

Before turning to the consideration of the interpretation of the theorem, I would like to add a brief note on the very much discussed notion(s) of locality in relation with Bell's theorem, in order to make explicit the referents at least of the words used in the present work.

3.1.4 Note on factorizability, locality, no-signalling

As it has emerged both in Bell's derivation of the theorem and in CHSH inequality, factorizability was referred to by Bell himself as *local causality*, thus, it is often claimed that Bell's theorem shows that quantum theory is *non-local* in the sense of *non-locally causal*. But, in the other hand, it has been mentioned that there are other assumptions in the backstage of the derivation, which make factorizabiliy richer than it appears.

Before starting an overview of the theorem's interpretations, I would like to clarify a point with respect to the use of locality associated with Bell's theorem and quantum mechanics. "Local" has a precise meaning in connection with relativistic space-time: namely, local are the interactions that do not exceed the speed of light, or, alternatively, a physical exchange of information among two events is local if it takes place within the light-cones associated with these events. More explicitly then, *"local" defines a physical interaction or a physical exchange of information that is bounded by the constraints derived by the structure of relativistic space-time*.

Of course, factorizability defines a type of probabilistic independence among events that is local in the sense just specified. But we will see that some accounts of the quantum correlations make them appear as *local* — in the sense just descibed — **and** *non-factorizable*. In fact, especially in the physics community, this sense of locality is much harder to give up, as opposed to the strong constraint of factorizability. The reason is that to dispense of locality in the sense we are discussing generates a contrast with relativity theory which is too high a price to pay given its enormous success as a physical theory.

Precisely for the importance attributed to the latter, already in the early days of the theorem's study yet another concept had been crafted to explain the quantum correlations without clashing with relativity, namely, the concept of *no-signalling*. This is perhaps the most permissive manner in which locality has been defined, and it goes as follows. Let us postulate that the quantum correlations may in fact arise because of an exchange of information between two events located at space-like separation; this exchange would, yes, exceed the relativistic constraints imposed by the relativistic space-time structure, but it would not violate relativity per se *if it was not possible to control such an exchange of information*. More explicitly, if it is not possible to exploit this exchange of information, e.g. to send a signal to a space-like separated location, this postulated exchange cannot be empirically proven, failing thus to demonstrate a consequent violation of relativity. Shimony himself (1990, pp. 40-41) referred to this fact as the grounds upon which a "peceful coexistence" between quantum mechanics and relativity could be maintained. The concept of no-signallig is an eminently operational constraint, which does not *really* rule out a possibly non-local explanation of the quantum correlations in the sense just described. For this reason, a theory which is merely able to achieve no-signalling is not as satisfactory from a theoretical point of view as a theory which is local in the sense specified above. Bell himself had expressed his dissatisfaction with such a notion, given its strong anthropocentric flavour. His expression of skepticism is quoted frequently: "More importantly, the "no signalling ..." notion rests on concepts which are desperately vague, or vaguely applicable. The assertion that "we cannot signal faster than light" immediately provokes the question: Who do we think *we* are? We who can make "measurements", we who can manipulate "external fields", we who can "signal" at all, even if not faster than light? Do we include chemists, or only physicists, plants, or only animals, pocket calculators, or only mainframe computers?" Bell (1990, p. 924). See also Goldstein et al. (2011, sect. 10.5 and endnote 71).

To summarize then: **factorizability** indicates the independence condition imposed on probability distributions specified for the derivation of the CHSH inequality above. Bell called it "local causality", but I will just use the term factorizability. **Locality** indicates the constraints on physical interactions or physical exchanges of information deriving by relativistic space-time. This is the sense in which I will use locality in this work. Finally, **no-signalling** indicates the impossibility to exploit an entangled state in order to send signals at a space-like separated location. It is an operational constraint which keeps the façade of a peaceful coexistence with relativity, although with obvious conceptual limitations. I will refer to this constraint with its name throughout the following pages.

3.2 Interpretations

As a matter of fact, to the question "what is the meaning of Bell's theorem?" it is not possible to find a unique answer able to tackle all the consequences of this proof, because the violation of the inequalities amounts to the violation of a constraint that at the end of the day depends on the conjunction of a multitude of assumptions. As pointed out before, the mere negation of a conjunction does not point *per se* at the "faulty" conjunct. Thus, since quantum mechanics is a physical theory and it does not have the privilege of remaining silent on a result that has physical implications, it calls for arguments that support an act of finger-pointing towards the assumption that needs to be given up for putting flesh on the result's claim. Admittedly, in an of itself the theory does in fact "merely" violate the constraints individuated by the theorem. Therefore, without complementary arguments, the theorem's significance remains only implicit. In this section I will briefly explore the main interpretations of Bell's theorem. A more general analysis of the logical structure of no-go theorems has been recently put forth by Radin Dardashti (2021), who provides a clear argument as to why it is common for no-go theorems to give rise to a plethora of different interpretations: this is in due in great part to the very logical structure of these results.

In section 3.1.3 I have shown in which manner quantum theory violates the CHSH inequality; it has emerged clearly that it is not necessary to make reference to any particular physical system, nor to any particular observable operator. By simply substituting the operators that appear in the CHSH inequality with quantum operators and considering a particular quantum state, the computation leads to a value that largely exceeds the inequality's bounds. It has emerged that such inequality is derived on the ground of the factorizability condition and of no-superdeterminism, and the demonstration shows that quantum theory evades the conjunction of these conditions, which means that at least one of them has to give.

Let us start with the option that quantum mechanics violates factorizability, but it maintains no-superdeterminism. In the derivation of the CHSH inequality we have seen that factorizability is deduced by the conjunction of parameter and outcome independence (equations (3.13) and (3.14) respectively). Its violation amounts thus to the violation of the conjunction of these two independences, which means in turn that at least one of them does not hold. Parameter independence states that it is not possible to affect the probabilities of outcome $A(B)$ by changing the settings $y(x)$. If this were possible, by looking at the outcome on one side one could obtain information about the choice of setting on the other side of the experiment — if also no-signalling was violated. Importantly, in the setting considered (see fig. (3.2)) the two outcomes are at space-like separation, as well as the choices of setting: the information about setting y would need to break the boundary of the light cone of B to reach A, which would result in a superluminal exchange of information, or, *non-local* in the sense just recalled in section 3.1.4.

On the other hand, outcome independence states that the probabilities attached to A cannot be influenced by the outcome B. Thus, it is not possible that once a measurement is performed on one of the systems the information about the outcome reaches and influences the system at the other end of the experiment in such a way as to reproduce the quantum correlations. This limit is again posited by relativity theory, in the sense that this exchange of information cannot take place for the same constraints which justify parameter independence. For understanding more sharply what are the implications of rejecting either of these independences or both, it is now time to say something more about what is playing in the background of factorizability, by making explicit the further assumptions on which the latter is grounded.

3.2.1 Key to the backstage: limelight on the premises

In order to make the exposition more precise, I am liberally relying especially on the work by Cavalcanti and Wiseman (2021), which recalls one previous work of theirs on Bell's theorem (Wiseman and Cavalcanti, 2017). Moreover, I also refer to Myrvold et al. (2019) and Tumulka (2016), who have tackled the conceptual foundations of Bell's theorem in a similar manner. What these works achieve, in different measures of precision, is a clear separation of all the premises that are involved in the derivation and that have been rejected over the decades by different interpretations of the theorem.

Let us start then by reporting what are the grounds on which factorizability stands. As we have already seen especially in Bell's own derivation, the independences that this condition enforces are directly connected to relativity theory. Some assumptions that derive from the latter are the following, taken from Cavalcanti and Wiseman (2021). Whereas in both this and the previous work the authors classify the assumptions as either axioms and postulates, or axioms and principles, I am going to simply refer to them as premises (P), and numbering them in the order in which I mention them, and, where only the page number is indicated, I am referring to the 2021's article:

(P_1) **Space-time** – *Every event can be located in a background relativistic space-time, where concepts like past and future light-cone, space-like separation, etc. can be made experimentally well-defined* (p. 2).

(P_2) **Temporal Causal Arrow** – *For any event A, there is a space-like hypersurface S containing A that separates events in the causal past of A (on the same side of S as A's past light-cone), from events that have A in their causal past.* (p. 8). The main import of this premise is to specify that there is a temporal order for events. No absolute past is implied, but only what can be involved in a partial ordering of events, for instance, in relation to the past light cone of the point in space-time where A is located. The temporal order between events at spacelike separation cannot be fixed uniquely.

(P_3) **Relativistic Causality** – *The causal past of an event cannot be outside the light-cones of that event* (p. 8). In connection with (P_3), this postulate fixes a causal structure for physical events.

We can already see how these constraints can be respected in principle by the situation graphically depicted in fig. 3.2, where the only events at space-like separation are the outcomes, their respective measurement settings are within their own light-cones, and the complete specification of the past state is in the area where the latter overlap. Moreover, the causes of the measurement events, namely, state λ and the respective measurement settings, are in the past light-cones of either outcome.

There are also some additional premises needed in order to derive factorizability, which are not strictly dependent on relativity theory. These are:

(P_4) **Absoluteness of Observed Events** – *Every observed event is an absolute single event not relative to anything or anyone* (p. 2). The name of this assumption has been recently introduced in the result by Bong et al. (2020). Alternatively, "(**R4**) Every experiment has an unambiguous outcome, and records and memories of that outcome agree with what the outcome was at the space-time location of the experiment." (Tumulka, 2016, p. 80); or, "*Unique Experimental Outcomes* — of the potential outcomes of a given experiment, one and only one occurs, and hence ... it makes sense to speak of the outcome of an experiment." (Myrvold et al., 2019, sect. 3.2.1). This is referred to as a minimalist realist assumptions by these authors, as it has the purpose of fixing the "uniqueness" of the physical world.

(P_5) **Principle of Common Cause** – *If two sets of events* **A** *and* **B** *are correlated, and no event in neither is a cause of any event in the other, then they have a set of common causes* **C** *that explains the correlations* (p. 5).

(P_6) **Decorrelating Explanations** – *A set of causes* **C**, *common to two sets of events* **A** *and* **B**, *explains a correlation between them only if conditioning on* **C** *eliminates the correlation* (ibid).

Reichenbach's principle of common cause is implied by the conjunction of (P_5) and (P_6), and it states that "if any two variables are dependent, then one is a cause of the other *or* there is a third variable causing both" (Pearl, 2000, p. 30), and a complete specification of the variables associated with the common cause factorizes the correlation among the two variables. What happens then when we reject factorizability, given this more precise specification of the premises? Let us start from the possibility that parameter independence has to go.

3.2.2 Rejection of factorizability

Parameter dependence One may explain the correlations among entangled subsystems as arising from the dependency of the outcomes on the distant measurement settings – outcome $A(B)$ would depend on setting $y(x)$. The influence that an event located in space-time can have received in its past and exerted in its future cannot exceed the limits of its past and future light cones, so the possibility that one parameter influences the space-like separated outcome, by interacting with or by sending information to it is a direct contravention of the relativistic limits to which physical interactions are subject. In particular, if this were possible, such information would be travelling from the event capturing the choice of setting to the distant outcome at superluminal speed, see fig. 3.3.

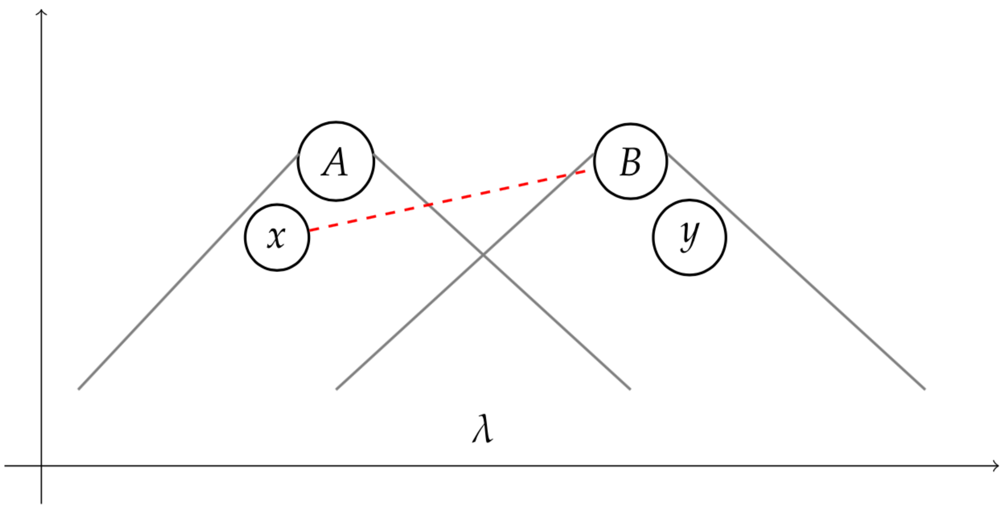

Figure 3.3: Space-time diagram of a bipartite Bell scenario, where the dashed, red line represents an exchange of information between measurement settings x and the space-like separated outcome B, in violation of parameter independence.

Where are we led then, by rejecting parameter independence? The immediate consequence seems to be an abandonment of the notion of relativistic causality, or, more precisely, of causality as constrained in the relativistic framework, according to premise (P_3), or, alternatively, also in the part of Reichenbach's principle of common cause that indicates the factorization of the conditional probabilities of two correlated events given their common cause, namely, (P_6). As we can clearly see in diagram 3.3 in fact, the causes of outcome B are not entirely in its past light-cone, as setting x is located at space-like separation both from the local setting y and from B itself. According to this scenario the factorizability condition would be violated as follows:

$$Pr(A, B \mid x, y, \lambda) = Pr(A \mid x, \lambda)Pr(B \mid x, y, \lambda)Pr(x)Pr(y)Pr(\lambda). \tag{3.23}$$

As we can see, the probability of outcome B is not independent of the distant outcome x — the same would happen for A with respect to y. The rejection of both (P_3) is implied by theories that clash with the relativistic spatio-temporal structure, such as for instance Bohmian mechanics, as specified in Myrvold et al. (2019, sect. 8.1.2). The latter theory outlines completely deterministic dynamics for quantum states, based on a specification of the positions of all systems relative to each other, requiring thus a privileged reference frame, which serves to specify a notion of absolute simultaneity. This view would be *non-local*, but *respecting no-signalling*.

On the other hand, one may give up the conditions according to which the causes of an event are in its past, namely, by postulating that causal effects could propagate in any direction within the light cones' boundaries. In this case, the assumption to be rejected would be (P_2), as the causes of an event may lie in its future, but not (P_3), since the notions of relativistic past and future in terms of light cones are not challenged within this interpretation. The latter view, commonly dubbed *retrocausation*, according to which causal influences can propagate in any direction in space-time, so it is possible that the choice of measurement influences retroactively the common state λ, making the two probability distributions dependent on each other, as represented in figure 3.4.

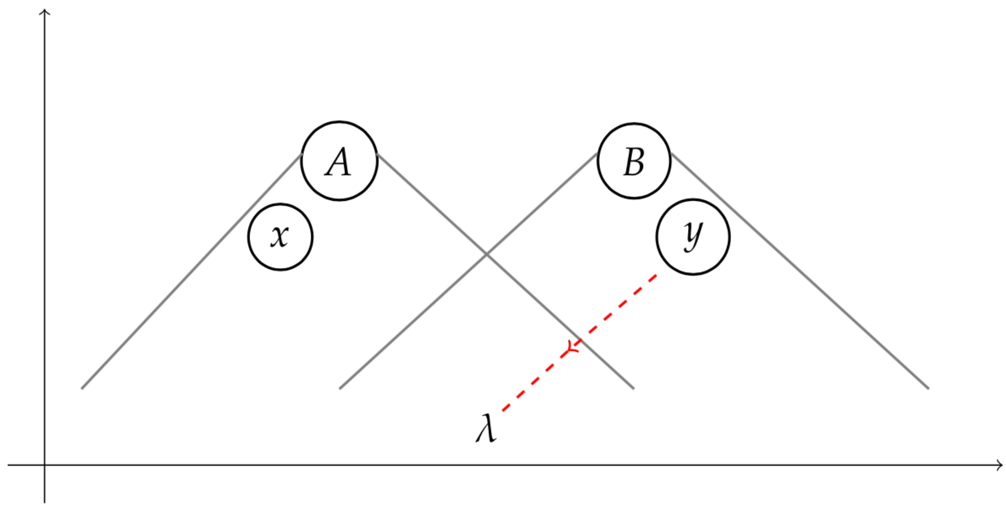

Figure 3.4: Space-time diagram of a bipartite Bell scenario, where the dashed red line represents an exchange of information between measurement setting y and common state λ, as in a retrocausal scenario. The arrow is in the middle of the dashed line to indicate the direction of this exchange, without generating a confusion between the notation of this graph and the DAG notation.

In this scenario, a more general independence between the measurement settings and the measured state *tout court* is broken. The assumption of retrocausation is consistent with a local hidden variables version of quantum theory, as explained by Corry (2015). Also interesting is Huw Price's clarification of the notion of retrocausality as opposed to simple reversed causation. Cf. Price (2008). As a consequence, the probabilistic dependences in the retrocausal scenario are the following:

$$Pr(A, B \mid x, y, \lambda) = Pr(A \mid x, \lambda)Pr(B \mid y, \lambda)Pr(x)Pr(y)Pr(\lambda \mid y), \tag{3.24}$$

where we can see in particular that the common state λ depends on the "future" choice of setting y — the same holds for x. This interpretation of the events is *local* and *respecting no-signalling*.

Finally, one may give up the assumptions that observed events are absolute (P_4), which would prevent us from formulating the condition of parameter independence in the first place. Interpretations that reject this, discussed in section 3.2.4, are relational quantum mechanics, QBism, and the pragmatist view. These views are also *local* and *respecting no-signalling*.

The import of the rejection of parameter independence has been discussed and challenged by some authors. In particular, it has been noted that the negation of parameter independence *per se* does not strictly entail a specific connection between parameter and outcome. For instance, Maudlin (2011, pp. 85-90), shows that the probabilities to which the parameter independence equation refers may have different interpretations, and that in general this terminology does not compel a univocal reading. Similarly, Norsen (2009) makes an analogous point, based on the interpretation of λ in the Bell scenario represented in figure (3.2). In particular, how one understands parameter and outcome independence derives also on one's understanding of whether λ is taken to represent a complete or only sufficient specification of all the relevant variables. Finally, a theory such as Bohmian mechanics, which gives up parameter independence, does not however allow for superluminal signalling, alleviating therefore the worry of a radical clash between this theory and relativity, besides the fact that Bohmian mechanics assumes a preferred foliation within relativistic space-time (Norsen, 2009, p. 285), (Myrvold, 2016, p. 240), (Myrvold et al., 2019, sect. 8.2).

To summarize, the rejection of parameter independence is a less desirable resolution for the rejection of factorizability, as it is harder to reconcile with relativity — as in the case of Bohmian mechanics, or because it is accompanied by the rejection also of no-superdeterminism, such as in the case of the retrocausal interpretation and of the superdeterministic account, which are endorsed by a minority in the philsophical community.

Outcome dependence Another manner to reject factorizability is to account for it via the rejection of outcome independence, respecting however parameter independence. This interpretation is the most frequent, as it seems to be naturally deriving from quantum mechanics. The fact that the probabilities of the outcomes of the subsystems involved depend on one another means that the outcomes on one wing are always correlated with what happens on the other. From a probabilistic point of view, this dependence is represented as follows:

$$Pr(A, B \mid x, y, \lambda) = Pr(A \mid x, \lambda) Pr(B \mid A, y, \lambda) Pr(x) Pr(y) Pr(\lambda). \tag{3.25}$$

Consistently with the exposition of parameter *dependence* in the previous paragraph, here as well the dependence of an outcome on the other outcome is represented for the case of B depending on A, but the same would hold for the symmetric case. Now, this mere probabilistic dependence leaves open more than one possibility as to how it may be justified physically. Let us consider which assumptions it may violate.

Outcome independence can be cast out as deriving from the conjunction of axiom (P_2), causal past and causal future of an event correspond with the event's past and future light-cones, (P_3) the causal past cannot be outside the past light-cone, and Reichenbach's principle of the common cause (P_5) + (P_6), as argued in Myrvold et al. (2019, 8.1.2). Now, if we regard the exchange of information that may arise between outcomes A and B as a cause of their correlations, what would need to go? If the correlations are instantiated *at the time of measurement*, as it would be the case in a stochastic theory such as standard quantum mechanics, it is not possible to write a conditional probability for the two events that allows to screen off the correlations on the account of a common cause in the past, so we could not use Reichenbach's principle to account for the relation between the two outcomes. Nevertheless, if the correlations were to arise due to an exchange at the time of measurement, the latter principle would not be called into account in the first place, as A and B *would* be causes of one another. How this can be viewed, in terms of a counterfactual analysis considering different relativistic scenarios is clearly spelled out by Bigaj (2020). The scenario we are considering is represented in diagram 3.5, which shows a situation that is *non-local* and *respecting no-signalling*.

If this were the case, the premises to be violated are, as in the case of parameter dependence, (P_3) and (P_3) or (P_4), again preventing the formulation of a joint probability distribution to which outcome independence could apply in the first place.

But this is not the only part of the story: the rejection of (P_2) and (P_3) in fact, could also result in a non-causal view of the correlations, which was defended some decades ago in the philosophical community, as a consequence of endorsing a (classical) account of interventionist causation, i.e. the account extensively articulated by

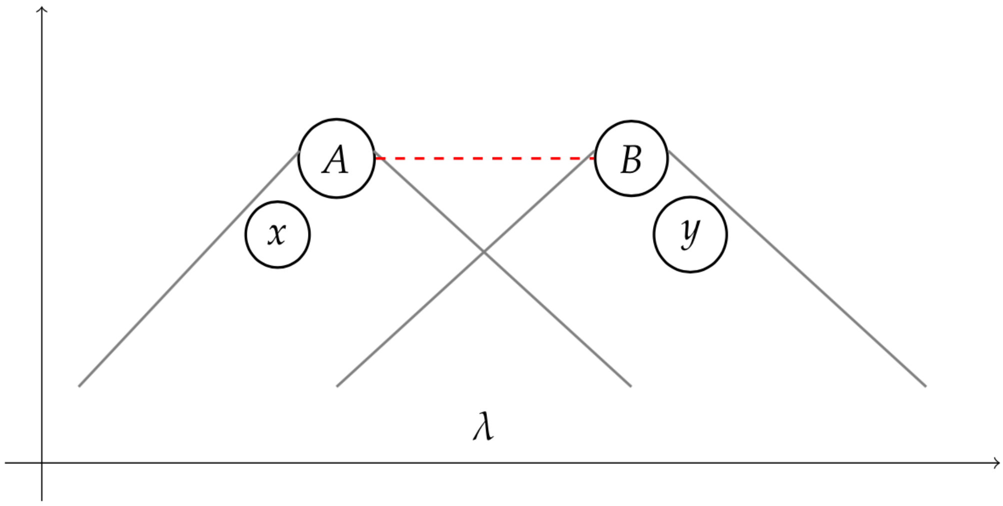

Figure 3.5: Space-time diagram of a bipartite Bell scenario, where the dashed red line represents an exchange of information between measurement outcomes A and B, in violation of outcome independence.

Pearl (2000), already quoted above. The non-causal view is supported as follows. On the one hand, the outcome variables cannot be controlled: the single values they assume upon every interaction with the detector do not depend on the intervention of the experimenter or on the device that carries out the measurement, thus, an interventionist view of causation *à la* Pearl cannot be applied in this case. A non-causal account of the correlations has been famously argued by Hausman and Woodward (1999). The non-controllability of the outcomes and the possibility that causation may not be involved have given rise to the view that outcome dependence is not in conflict with relativity. More precisely, even though the mutual influence of the subsystems involved would still have superluminal nature, since this phenomenon does not conflict with relativistic constraints in an obvious manner, this form of non-locality has been seen as compatible with relativity — on the proviso that no-signalling is respected.

The way in which outcome dependence emerges within standard quantum mechanics has been underlined by Shimony via the simple consideration of the entangled states that violate Bell's inequality (1990, p. 38). In virtue of the non-separability of the state, it is not possible to assign the two subsystems with conditional probabilities which do not depend on each other. Quantum mechanics *coexists peacefully* with relativity, given that parameter independence is left intact (pp. 40-41). As pointed out in (Myrvold et al., 2019, sect. 8.2) the notion of peaceful coexistence has evolved in the course of Shimony's works, this being only the first appearance of this concept. More generally, outcome independence is rejected by indeterministic versions of quantum theory, since information about the choice made on the distant setting is redundant, once information about the state is available.

However, one could also maintain (P_2) and (P_3), and account for the correlations in violation of outcome independence via a common cause, by modifying parts of Reichenbach's principle. One could in fact reject (P_6), opening multiple options. On the one hand, one may still maintain that the correlations are causal, but of a different kind than the notion of classical causation, as argued by Butterfield (1992), Chang and Cartwright (1993), and also Maudlin (2011, ch. 5). On the other hand, one may reformulate the principle of common cause as it has been done for instance by Hofer-Szabó and Vecsernyés (2013), Leifer and Spekkens (2013), (Egg and Esfeld, 2014), and discussed by Cavalcanti and Lal (2014). The last paper in particular constitutes a crucial step for the subsequent development of *quantum causal models* — see Shrapnel (2017) for a philosopher-friendly introduction —, which maintain a causal account of the entanglement correlations, but via a radical modification of the notion of classical interventionist causation. This latter family of interpretations is *local* and *respecting no-signalling*. The debate on the nature of causation connected to this interpretation of Bell's theorem shows how deeply this result touches other notions that are crucial for our understanding of the world, such as causation, and it unveils the rebellious nature of quantum phenomena, which call us to question classical intuitions and structures about the nature of the world, and to revise them profoundly, whichever way is preferred.

3.2.3 Rejection of "no-superdeterminism"

We have seen that the derivation of the CHSH inequatity depends on the assumptions of factorizability and of *no-superdeterminism*, i.e. that the choices of settings are *not* correlated with the outcomes in any relevant manner. The terminology leaves open the possibility that the choice of which experiment to perform may be carried out by an experimenter, or randomly by a device, so no assumption regarding free will or consciousness needs to be done. It was pointed out in section (3.2.1) that this assumption is based on the idea that even though the light cones of the outcomes events (fig. 3.2) may eventually overlap in the distant past, the variables related to the choices of measurement, i.e. the controllable parameters in this scenario, would still be "free": "For me this means that the values of such variables have implications only in their future light cones. They are in no sense a record of, and do not give information about, what has gone before. In particular they have no implications for the hidden variables [λ] in the overlap of the backward light cones" Bell (1977, p. 100).

The independence between choices of settings and outcomes invoked in this assumption is not "absolute", but it is rather based on an idea of reasonableness. In the relativistic framework that embeds the assumptions from which the theorem is derived, it is implied that the light cones of any event in space-time eventually overlap at some point in the past. Thus, by meticulously following the relativistic picture one could not formulate the no-superdeterminism condition: *strictly speaking*, the measurement settings and whatever process is involved in their selection *do* in fact depend on variables common to the outcomes events in their distant past. The point is however that acknowledging this fact would carry consequences that would undermine the scientific investigation of the world: the link between these variables in the distant past and the present measurement settings cannot arguably be pointed out in an informative way, but it remains a connection that is there *in principle*. For this reason, Shimony, Horne and Clauser (Bell et al., 1985, pp. 100-101) argue that such connection should be dismissed in the understanding of the Bell scenario, as its non-exploitable presence finally does not amount to much more than a conspiracy looming in the background. Similarly, in response to their comment, Bell proposes the illuminating example of a random generator which determines both variables x and x'. The difference between them is their dependence on the "oddness or evenness of the digit in the millionth decimal place of some input variable." (Bell, 1977, p. 103). Then,

> fixing [x] or [x'] indeed fixes something about the input — i.e., whether the millionth digit is odd or even. But this peculiar piece of information is unlikely to be the vital piece for any distinctively different purpose, i.e., it is otherwise rather useless. With a physical shuffling machine, we are unable to perform the analysis to the point of saying just what peculiar feature of the input is remembered in the output. But we can quite reasonably assume that it is not relevant for other purposes. In this sense the output of such a device is indeed a sufficiently free variable for the purpose at hand. For this purpose the assumption (1) [the settings of instruments are in some sense free variables] is then true enough, and the theorem follows. (*ibid.*)

This brief clarification on the freedom involved in "no-superdeterminism" serves to put the assumption in context. The rejection of this postulate does not generally hinge upon the possible complications arising from the mentioned relativistic view. A view that dispenses of this assumption is commonly known as superdeterminism, known for its unpopularity grounded on different reasons, one being that it "seem[s] to undercut the core assumptions necessary to undertake scientific experiments"

(Wiseman, 2006, p. 85), another being that it seems to resort to a *conspiratorial* view of events, as argued in two papers by Indrajit Sen and Anthony Valentini, (Sen and Valentini, 2020a) and (Sen and Valentini, 2020b).

In a superdeterministic universe, the correlations among entangled subsystems are *local*, because the values assumed by the variables of both measurement settings and outcomes arise entirely deterministically, see fig. 3.6. Although a fully-fledged superdeterministic theory does not exist yet, there is a set of tenets that are commonly held in the literature on this topic (see t Hooft (2016), Hossenfelder (2020), Hossenfelder and Palmer (2020), Baas and Le Bihan (2020)). In brief, superdeterminism commonly stands for a local hidden variables model that does not respect the principle of statistical independence (Hossenfelder, 2020), or of measurement independence (Baas and Le Bihan, 2020).

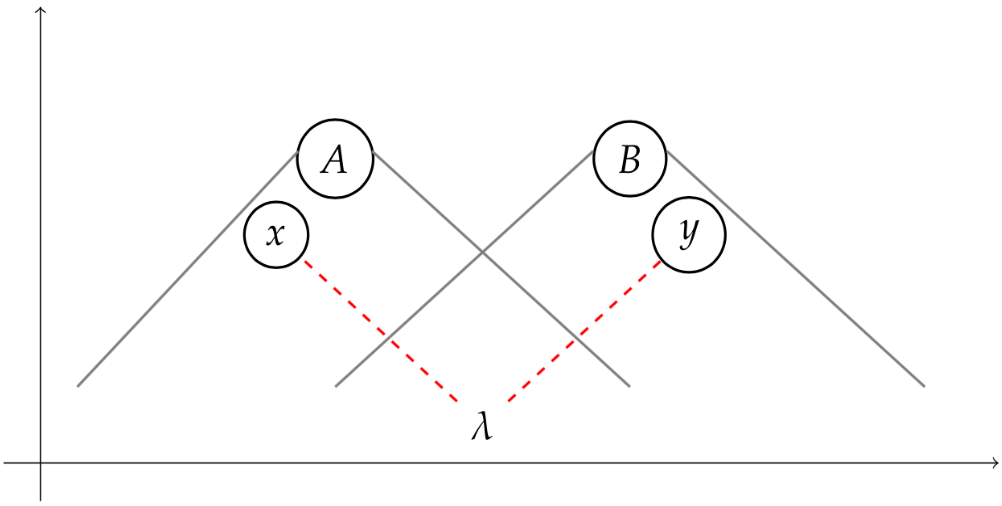

Figure 3.6: Space-time diagram of a bipartite Bell scenario, where the dashed red line represents an exchange of information between measurement outcomes A and B, in violation of outcome independence.

As shown in fig. 3.6, this is phrased in terms of a dependency of the choices of measurement x and y on a set of variables in the past of these choices. Importantly, in the superdeterministic picture λ does not indicate simply the entangled state, but it refers rather to *all* the variables that are relevant for bringing about the choices of measurement that were actually made. The probabilistic dependencies among the variables in a superdeterministic scenario would be the following:

$$Pr(A, B \mid x, y, \lambda) = Pr(A \mid x, \lambda)Pr(B \mid y, \lambda)Pr(x \mid \lambda)Pr(y \mid \lambda)Pr(\lambda). \tag{3.26}$$

This is precisely the view that Bell, and Shimony, Clauser and Horne dubbed as conspirational, and generally incompatible with the scientific enterprise.

3.2.4 Alternative interpretations

So far, we have considered views of the theorem that conflict with no-superdeterminism and with factorizability, in the latter case by considering which of the assumptions have to go on the basis of the rejection of either parameter or outcome independence. Nevertheless, there are further views of the theorem that account for the entanglement correlations by individuating a conflict with other assumptions. In this section I am going to consider some of these views.

I. Rejection of (P_4)–Absoluteness of observed events Premise (P_4) says that an observed event is real and single, or unique, and especially that it is not relative to anything nor to anyone. The interpretations mentioned in this section deny precisely these properties of observed events, although on grounds that differ from one interpretation to the other. Let us start from one of the most popular quantum interpretation, namely, Everettian quantum mechanics.

According to this view, what Bell's theorem really proves is that quantum theory teaches us something about the very structure of the universe. Rather than challenging locality, the theorem shows that the idea of the uniqueness of the world, where the observed experimental outcomes are the *only ones* which actually occur, is fundamentally a misconception. In other words, the theorem shows that the universe is actually a *multi*verse.[2] In the many worlds version of Everettian quantum mechanics, each measurement event causes the world to split into a number of worlds that corresponds to the number of eigenvalues involved in the quantum superposition considered. Measurement, in this case, does not collapse the state into one definite outcome while *discarding* all the others: it *actualizes* each outcome, each in a separate world. The connection between this interpretation and the correlations among entangled subsystems emerges from a rejection of their alleged non-local character. As stressed convincingly by Vaidman (2016, p. 197), the "mechanism" underlying the correlations is both unknown and unobservable, so, from a physicist's point of view, it may as well be considered as a miracle. The obvious dissatisfaction with respect to this conclusion can be healed by defending the hypothesis that non-locality as action at a distance may just be illusory, in the sense of emerging from an underlying

[2]For illustrative purposes, I am focussing on the many worlds version of Everettian quantum mechanics, although I am aware that there are alternatives to the many worlds interpretation, as outlined in (Conroy, Conroy, sect. 3).

structure of which we arguably do not have direct experience. In the many worlds interpretation, such a structure is given (a) from writing the global quantum wave function of the universe, which, in quantum mechanics is defined on abstract configuration space, and, to make sense of the classical-looking world of our experience, (b) by decomposing the universal wave function into a superposition of terms, "each corresponding to a different world", where "*A world is the totality of macroscopic objects: stars, cities, people, grains of sand, etc. in a definite classically described state*. The concept of a "world" in the MWI belongs to part (ii) of the theory, i.e., it is not a rigorously defined mathematical entity, but a term defined by us (sentient beings) in describing our experience. When we refer to the "definite classically described state" of, say, a cat, it means that the position and the state (alive, dead, smiling, etc.) of the cat is maximally specified according to our ability to distinguish between the alternatives, and that this specification corresponds to a classical picture, e.g., no superpositions of dead and alive cats are allowed in a single world." (Vaidman, 2018, sect. 2.1, emphasis in the original). In this manner, whereas a classical description would specify the trajectories of, e.g. all particles and the values of the fields, in the quantum case each term of the mentioned decomposition indicates the states of macroscopic objects as in a product state (ibid., p. 200).

To postulate the existence of a branching universe is not the only way to go to reject the uniqueness and absoluteness of observed events. A viable alternative, defended both within Relational Quantum Mechanics and by QBism, is to assign an entirely different role to the quantum state. Although these two interpretations differ in the manner in which they reach their conclusion about the quantum state, this latter ends up being the same object for both camps, reason why I am considering the latter two together in the present occasion.

In these interpretations, the quantum state is not an absolute notion to be endowed with metaphysical valence. Rather, in the relational interpretation it is to be understood as

> a mathematical device that refers to two systems, not a single one. It codes the values of the variables of the first that have been actualised in interacting with the second ... it therefore codes anything we can predict regarding the future values of these variables, relative to the second system. The state ψ, in other words, can be interpreted as *nothing more than a compendium of information assumed, known, or gathered through measurements, determined entirely by a specific history of interactions*: the interactions between the system and a second 'observing' system. (Laudisa and Rovelli, 2019, sect. 1.4, emphasis added).[3]

[3]Cf. also Rovelli (1996, 1663-1668) and Rovelli (2018, p. 6).

This manner of seeing a quantum state is analogous to the QBist notion of state as a *subjective epistemic state*, which contains information about the agent's possible future experiences (Healey, 2017a, sect. 1). In both the relational and the QBist cases therefore, there is not an observer-independent matter of fact described by the quantum state, as this latter captures the information that a particular observer obtains by interacting with a system via measurement. In the relational case, any system that interacts with another is an observer, whereas the QBist observer is a human agent. This difference does not however entail diverging notions of state. Conversely, an observed event is an event about which information is acquired by a particular other system, or that is experienced subjectively by an agent. In neither case it is regarded as observer-independent. In both interpretations, the rejection of (P_4) yields to a preservation of locality in the description of the correlations among entangled systems. Non-locality is not necessary to explain the correlations. Relationalists maintain locality as defined in sect. 3.1.4, according to which it is not possible to retrieve information from a system located outside the observer's light cone, and they endorse no-signalling. The limits posited by this structure are taken in an operationalist flavour: "nothing enables [observer] A to know the outcome of the measure carried out by B on [system] β, unless A measures the state of B. But A cannot measure the state of B instantaneously, precisely because of locality: B is far away." (Smerlak and Rovelli, 2007, p. 436).[4]

QBists do abide to relativity as well in their subjective notion of the state. Since no agent can travel faster than light, the experiences that they describe using quantum mechanics can never be space-like separated. Thus, the reasoning goes as follows:

> Quantum correlations, by their very nature, refer only to time-like separated events: the acquisition of experiences by any single agent. Quantum mechanics, in the QBist interpretation, cannot assign correlations, spooky or otherwise, to space-like separated events, since they cannot be experienced by any single agent. Quantum mechanics is thus explicitly local in the QBist interpretation. And that's all there is to it. Fuchs et al. (2014, pp. 750-751).

In both cases, quantum states describe a situation that needs to be specified relatively to two systems, or to an observer, and such description is embedded in the relativistic structure of space-time. Differently from the Many Worlds Interpretation, Relational Quantum Mechanics and QBism do not defend a realist view of the quantum state, and both explicitly question the tenability of Einstein's definition of an

[4] "Observer A can of course measure the state of B ... but only when A is back into causal contact with B. This, needless to say, is in the future light-cone of A, and therefore poses no challenge for locality." (ibid.).

element of physical reality, on the ground that it seems to be excessively demanding. Nevertheless, in all three cases, Bell's theorem is interpreted as compelling a rejection of the axiom that the world is unique and absolute.

3.2.5 An alternative approach: so long, classical physics

Despite it is not straightforward to phrase this approach along the lines of what has been done in the previous pages, namely, in terms of the set of assumptions recalled in section 3.2.1, there is yet another interesting manner to understand Bell's theorem's result. The latter can be seen as a powerful hint that not only some of our rooted intuitions about the world are inadequate in the face of quantum mechanics, but *the whole framework* within which these intuitions are crystallized in classical field theory and classical physics proves its overall inadequacy to account for all types of physical phenomena. Within this type of view, Bell's theorem is one theorem , among other results, proves something at the meta-level of theory structure. Some interpreters have argued that the assumption that generates a contradiction between the Bell's inequalities and the predictions of quantum mechanics is sort of a "framework" assumption of a classical structure for physical theories, which we may call *classicality*.

(F_1) **Classicality** – This is a formal assumption about the mathematical representation of systems in classical physics, which, in the context of Bell's theorem's scholarship, has been formulated clearly by Reinhard Werner (2014). This premise was not introduced by Werner in the first place, but its discussion draws back to the founding fathers of quantum mechanics (Birkhoff and Von Neumann, 1936). In physical theories, information about physical systems is encoded in the systems' states, which are represented as elements of a state space, i.e., an abstract mathematical space with certain properties: in the classical case, the state space is "a simplex, meaning that any state has a unique decomposition into extreme points, so can be understood as statistical mixture of dispersion free states: equivalently, any two measurements (POVMs, or decompositions of one into positive affine functionals) are the marginals of a joint measurement. We take these properties as a definition of *classicality*" (ibid., p. 3, italics added). A consequence of these spaces' properties is that the states defined on them have definite values regardless of a measurement being performed on the systems; they seem to capture "facts of the matter" (p. 3) about the systems.

Given this type of assumption, and Kolmogorov's axioms of probability — see

(Kolmogorov and Morrison, 2018), (Hájek, 2019) —, it is possbile to derive Bell's inequlities, which are then violated by the quantum predictions. Werner argues that it is precisely (F_1) the constraint broken by quantum mechanics. A complete discussion of this assumption is beyond the point of the present section, and it would require the right space to be fleshed out in all its technical complexity. However, it is possible to provide a minimal idea of what it amounts to, in order to clarify the meaning of its rejection. A thorough exposition and discussion is provided by Jeffrey Bub and William Demopoulos (1974), which I am following in these pages.

They start by considering both classical and quantum mechanics as "theories of logical structure", because they formulate general constraints for the "occurrence and non-occurrence of events" (p. 92). The division into two kinds of physical theories, namely, "principle" or "phase space" theories — i.e. theories of logical structure in our case —, and "constructive" theories follows a distinction traced initially by Einstein. Francisco Flores (1999) has introduced the useful nomenclature of "framework" vs. "interaction" theories, and explored this distinction within classical physics, while a thorough study of this classification for relativity theory, quantum mechanics, quantum field theory and the standard model has been carried out by Federico Benitez (2019). What *theories of logical structure*, or framework theories need then is a phase space structure, and a probability algorithm which represents the totality of *all possible events associated with a certain class of physical systems* (p. 95). The crucial point of this classification of theories is that in classical mechanics the algebra of all possible events related to a system is a Boolean algebra, which thus shares the same properties and structure of classical logic. On the other hand, the theoretical propositions[5] of quantum mechanics form only a *partial* Boolean algebra, which is not "embeddable" in a Boolean algebra, but it is rather emerging from a combination of Boolean sub-algebras. The fact that there is a fundamental incompatibility between the two structures is proven by Kochen-Specker theorem: "if $\mathcal{A}$ is a partial Boolean algebra of events corresponding to a Hilbert space of at least three dimensions, there is no imbedding of $\mathcal{A}$ into a Boolean algebra; hence, there is no imbedding of the partial Boolean algebra Q [i.e. for the quantum case] of theoretical propositions associated with $\mathcal{A}$ into a classical (i.e., Boolean) description. This follows from Kochen and Specker's Theorem 1, because the existence of two-valued homomorphisms is a necessary condition for the imbeddability of Q into a Boolean algebra." pp. 115-116 —

[5]"Consider the set of all intervals of the real line, R, half-open (on the right). The Borel subsets of R are the sets contained in the σ-ring generated by this set. A *theoretical proposition* about the system asserts that the value of a physical magnitude lies in one of these intervals." (ibid., p. 96, emphasis in the original).

the mathematical notions of embedding and homomorphism are defined in the paper. What is important about this is that it is not possible to reduce the representational apparatus of quantum theory to the structure that underlies classical mechanics, and this is not a mere mathematical convention. As the authors point out:

> If it is maintained that logical structure is conventional it must be possible to show that there is something which the question of logical structure does not share with other theoretical issues which would justify such an interpretation. For example, it is generally required that conventions be dispensible. Hence, if the choice of non-Boolean logical structure were conventional, it should be possible to reformulate the theory without this choice. But the logical structure of the quantum theory does not have this character. (p. 98)

A fundamental problem in the interpretation of quantum theory is that it does not contain "dispersion free" states, namely, states that assign a definite value to all the observables of a system at every time, which is the result proven precisely by Kochen-Specker's theorem. It is important to stress that this feature is ingrained in the very structure of the theory, which is the reason why the interpretative problems have foundational character. So, with respect to Bell's theorem, the way to understand it according to Werner's, Bub and Demopoulos' arguments would be the following: a set of probabilistic variables attached to a physical situation can be constrained as specified by the Bell's inequalities *if* we start from the framework of classical physics, which relies upon certain mathematical structures that enforce a certain treatment of these variables and their mutual relations — in particular, providing a single, joint probability distribution over a single state space for all the parts of the Bell scenario, where the relativistic space-time structure justifies factorizability. Of course, the use of these structure is optimal for encoding the way in which the world is according to classical physics. According to this view then, what the clash between the quantum predictions and the Bell's inequalities prove is the fact that *the whole framework* is defective for the representation of quantum phenomena such as entanglement and the correlations it yields. In a slogan, Bell's theorem is one result in the foundations of quantum mechanics which proves that the classical framework, here represented by (F_1), cannot be retained as a universal and general vehicle for representing the physical world.

This minimalist outline of the logical structure of classical mechanics and its incompatibility with the quantum counterpart serves the purpose of clarifying where the incompatibility is located for the interpreters who reject (F_1). The new structure introduced by quantum mechanics, and the testable accuracy of its predictions show us that the contradiction was at the heart of classical physics all along.

❊ ❊ ❊

In this second half of the chapter I have considered the main interpretations of Bell's theorem, and pointed out which assumptions each of them rejects. First, I have considered the views in direct rejection of factorizability, by dividing them in the views that give up either parameter or outcome independence. On different grounds, both seem to present some incompatibility with the relativistic understanding of space-time, this contrast being more explicit in the case of parameter dependence. I have then considered three other interpetative avenues, namely, superdeterminism, which, self-explanatorily gives up no-superdeterminism. Everettian quantum mechanics, relational quantum mechanics, and QBism have been taken collectively as a group of views that reject "Absoluteness of observed events". Finally, I have introduced briefly a separate way of approaching Bell's theorem and its interpretation, by pointing at the contrast between the logical structure of classical and quantum physics. This short exploration does not have any pretence of completeness, but it has rather the value of highlighting some important features of Bell's theorem in connection with quantum theory: in order to acquire physical meaning, or any significance that goes beyond the mere uncovering of an *impasse*, each interpretation constitutes a fully-fledged and coherent world-view that provides its claims with a particular perspective. In the next concluding section I put forth some philosophical considerations.

3.3 Any role for observation?

In this chapter I have focussed on one of the main results of contemporary physics, namely, Bell's theorem. Distinctive trait of the elegant proof of the theorem is that it can be derived without making strong commitments to a specific ontology: given the mentioned assumptions, what is required is a distinction of two different kinds of variables — measurement settings and outcomes, and a definition of the relations among the probability distributions attached to them. Interestingly, Bell did commit to the concept of *beable*, namely, a real aspect of the world that has definite properties irrespective of their being measured or not. Nevertheless, the proof withstands a more skeptical attitude towards such commitments, which is a crucial aspect of its strength. The theorem, and I am intentionally using a strong expression to convey the point, does not amount to anything more than a tautological implication from the premises *unless one considers the result in connection with quantum mechanics*. Fundamental part of this exposition has been in fact the *violation* of Bell's inequality, achieved by calculating the expectation values for the parameters defined in the inequality by using a maximally entangled state, and applying Born's rule. Once again, this derivation has not required to commit to any particular ontology either, as it has rather followed from the theory's bare formalism. For how satisfactory the mathematical beauty of these derivations may be to some however, the clash between Bell's inequality and quantum mechanics is no matter to be confined in the ivory tower of formalism. The gravity of this contrast explodes in all its difficulty when it is brought on the concrete level of its physical interpretations. Although to a limited extent, in this chapter I have also explored the main interpretations of the theorem, on the basis of the assumptions given up by the different camps. It has emerged that the claim that *Bell's theorem shows that the world is non-local* is uttered on the basis of the rejection of parameter independence or of outcome independence, and that the former is generally seen as the most problematic alternative, in light of its possible conflict with relativity. A brief exploration of the local alternatives of the theorem has been carried out too, subdividing them according to which of the other assumptions they reject.

What all this leaves us with is a complex theoretical apparatus with many different and mutually incompatible avenues to potentially pursue: is quantum mechanics a non-local theory or not? Its being in contrast with classical physics does not seem to be particularly alarming, since, if we disregard the human experience of the world, our physics ceased to be Newtonian a long time ago already. But we most certainly cannot disregard its clash with relativity. Is it true that the two can "peacefully" coexist, or is this coexistence an overtly war-like matter of fact? And, if we follow the local alternatives, are we really ready to give up the irreducible intuition that we are free in our choices? Or should we rather believe that the universe contains a plethora of parallel and inaccessible worlds, the existence of which is derived on the sole basis of the structure of the wave function? Perhaps the idea that the thousands-year old structure of our logic is completely misguided at the end of the day may be more palatable than such a problematic metaphysics...? These excessively provocative questions have the rhetorical function of putting the finger on the problems that Bell's theorem potentially raises, and to stress the fact that there simply is not a straightforward answer to the question of what this result exactly means: there is a reason if we found our melancholic expert explorer deep in the complications of this theoretical forest. As briefly noted in the conclusion of the previous section, each interpretation does not simply give up a premise, but it is rather based on a fully-fledged theory or interpretation which provides access to a coherent and consistent world-view.

Again, in this context we are dealing with *physical* theories, i.e. theories built to account for the (mind-independent) world. The fact that each provides a different view on the theorem based on the rejection of a different assumption does not provide much of a parameter as to how one may choose to endorse one view over another. But one can always test their theory, and it is here that the object of research of this thesis comes back: can observation come to the rescue, and make us overcome this apparent impasse in front of the mentioned interpretations? The fact that these latter are pretty much all still alive and well may give the hint that probably the answer to this question is "no". However, we are not faced with a monolithic negation. True, there are assumptions that cannot be empirically tested. It is interesting to notice that, *prima facie*, these are the framework assumption of classicality, the premise of absoluteness of observed events, and of no-superdeterminism. The first is an eminently formal assumption, thus it results quite unproblematic to leave it undisturbed by this concern. The other two however, aren't. Whether the universe *is* or *is not* a multiverse makes a big difference on the grand scheme of things; but it is constitutive part of the family of Everettian theories that the parallel dimensions do not interfere with one another, and cannot be accessed by an observer located in one of them. Thus, for the Everettian it is *as if* we lived in a unique universe, even though we do not. Do we have any more luck with the third? To believe that in fact the correct picture is a superdeterministic picture, when our intuitions and experiences of the world lead us to believe that we are free is as difficult as to believe that the world exists independently of us, although we are in fact brains in a vat. As in the latter case, however, this postulate is strictly non-falsifiable.

To answer to the question that titles this section then, for the interpretation of the theorem in all its richness, observation seems to not play a particularly important role. Nevertheless, there are properties unveiled by the dialectic between Bell's theorem and quantum mechanics that have been extensively tested, and that have increasingly and indisputably secured the theorem's success. These are the fact that entangled systems yield quantum correlations upon measurement, and that these correlations arise at space-like separation. With a tradition of experiments that started in the 70s and was established by the historical achievements of Alain Aspect and collaborators, the correctness of Bell's theorem reached empirical grounds as well. The details of these experiments will be outlined in the next chapter. The crucial point to be made already here is that, regardless of whether this is a mere illusion, these experiments prove that within the framework of how we both experience and scientifically measure the world the quantum predictions are correct. Before turning to the philosophical implications that this has for the role of observation in quantum theory in relation with the case study of Bell's theorem, which will be object of chapter 5, let us turn concretely to observation, by studying the experiments that have definitively corroborated the quantum predictions.

Chapter 4

Bell's theorem – Experiment

4.1 Introduction

In the previous chapter it was shown how the theorem does not rely upon the postulation of a particular ontology, as its extent is eminently mathematical. It is one of its interesting features the fact that, despite it considers a physical scenario, it remains independent from *particular* objects that such a scenario would involve. But the fact remains that it is central to this theorem to yield physical predictions. *What* can we apply them to then, if the formalism does not help with a particular reference on its own? Interestingly, it will be outlined in the following pages how it took a long time for the community to recognize the entanglement correlations in a concrete experiment. The present chapter provides an analysis of the Bell experiments from a historical perspective, and a brief philosophical reflection about the latter. It is at this point that we will encounter "legendary" characters of this history, because their seminal contributions to the physics of Bell's theorem regards the empirical corroboration of the quantum predictions, and therefore the verification of the proof.

The experimental confirmation of the theorem displays an important difference with the theoretical counterpart. If the theorem has given rise to a spectrum of different interpretations as to what it has actually proven, the same cannot be said for the experimental side. As we will see, one of the main concerns of the experiments is to rule out what the research groups designing and performing the Bell tests refer to as *local realistic theories*, namely, theories which would account for physical phenomena by means of classical variables, in harmony with the CHSH inequality and in disagreement with the quantum predictions that the latter should be violated. As we will see both in the exploration of the loopholes, and while surveying the main Bell tests, a main concern in this sort of experiments is to design a set up able to prevent a local realistic explanation of the observed evidence. This entails that the empirical study of the theorem is much more narrowly focussed and less open to interpretations than the theoretical study. One immediate reason for this difference is that other interpretations of the theorem refer to structures or to an ontology that are not open to empirical test. For instance, it is not possible to design an experiment to observe the presence of alternative worlds described by branches of the wave-function, as much as it is not possible to test the presence of an ensemble of point particles with definite positions at all times, let alone influences coming from the future. What it is possible to test, however, is that if these correlations were to violate the classical framework, they would need to break some of its constraints, such as for instance the spatio-temporal ones derived from relativity theory.

At the time in which I am writing, the first half of 2021, there is fairly no doubt that the predictions of quantum mechanics against Bell's theorem have been empirically corroborated. As it will be shown below, experiments that have closed all the relevant loopholes have already been successfully designed and performed. Thus, the issue of whether the predictions made on the basis of the theorem and of quantum mechanics are both testable and correct has been settled. Given the already extensive and rich literature on the experimental history of the theorem's testing, this chapter will not attempt a review. I will rather take the occasion to focus on aspects of this experimental history that are of philosophical interest for the present analysis of observation. We will come back in ch. 5 to Shapere's definition of observation to see how it fares with respect to the experiments presented in the current chapter. In what follows I focus on other aspects of this history of Bell tests which constitute fertile ground for the relevant philosophical discussion.

The structure of the chapter is as follows: I will first enter in the matter of what are known as loopholes, namely, explanations of the observational evidence that do not violate the theorem's constraints, and thus potentially undermine a quantum mechanical account of the evidence — section (4.2). In section (4.3.1) I address the question of how entanglement is produced and tested in the laboratory, and why photons have become the preferred systems for Bell tests. An overview of the main steps of the experimental history of the theorem is provided in sections (4.3.2) and (4.3.3). They are divided in two parts according to the main techniques used to produce pairs of entangled photons, namely, cascade atom decay and parametric down-conversion. Within these two sections I also explore the experiments of Alain Aspect and collaborators, and the first three loophole-free Bell tests respectively. After having provided a picture of the experimental tests, I advance a philosophical reflection (4.4) on the connection between theory and empirical phenomena that has been explored in the pages dedicated to the experiments.

4.2 Loopholes

Hypothesis H is ruled out via experiment when — amongst other things — the data produced by said experiment cannot be explained by H. In the case of a Bell test, the hypothesis to be tested (and to be ruled out) is a local realistic explanation of the quantum correlations. As eloquently argued by Marissa Giustina in (2017, p. 487): "It became clear that any attempts to recover the quantum mechanical theory on a local realist foundation were futile. The [Bell] inequality itself, a purely theoretical statement, sheds little light on the "nature of reality". However, it provides clear hints about the sort of experiment one should perform". This means that the correlations observed in the experimental outcomes must (a) violate the constraints derived from the theorem's assumptions, and (b) not be reproducible by means of a local mechanism. But even when the requirements to obtain an eminently quantum effect seem to be respected, there may still be ways to account for the data in a local realistic fashion. These ways, in the context of Bell tests, are referred to as loopholes. The negative connotation of the term is due to the fact that, if they were to hold, standard quantum mechanics could not be considered correct at face value. Jan-Åke Larsson (2014, pp. 1-2) makes reference to the meaning of the term in the legal context: a loophole in the legal system allows one to behave in disagreement with the law without however violating it. Analogously, the loopholes in Bell tests constitute the ground for explanations of the obtained evidence that do not resort to quantum mechanics and that do not violate the theorem's constraints.

An exploration of the loopholes affecting Bell tests provides fertile ground both for theoretical considerations and for understanding the impressive experimental achievements of the empirical tests that corroborated the theorem. The literature has grouped the various loopholes under two macro-categories. (See Larsson (2014), Giustina (2017), Genovese (2005), Myrvold et al. (2019, sect. 5), Goldstein et al. (2011, sect. 7)) These are the family of the locality loopholes on the one hand, and the family of the detection loopholes on the other. The list proposed below is not exhaustive, as the difficulties that characterize experimental science are by far a greater number than the schematic list considered here. The reason for focussing on the main loopholes grouped under these two families is that their resolution is based on a call for action that overlaps with theoretical physics as well. As a matter of fact, their *closure* has a very strong theoretical component as well. Let us delve now into the consideration of these fascinating problems, and of how they have been solved.

4.2.1 Locality loopholes

Although we have not entered in the matter of actual experiments yet, it has already been mentioned that a measurement on a bipartite entangled system which aims to experimentally violate local causality requires two measuring devices at two different locations at space-like separation from one another. Locality is the assumption according to which the physical influence that a system can exert on another at a separate and distant location does not exceed the relativistic limit of the speed of light; therefore, a conditional probability distribution for either outcome given the local setting and a shared original state is constrained, given additional classical assumptions, as indicated by the factorizability condition (see eq. Factorizability, section 3.2.1 in chapter 3). The hypothetical explanations of experimentally observed correlations among entangled systems that respect factorizability are part of the family of the locality loopholes. There are three eminent members to this family, namely, the *communication loophole*, the *memory loophole* and the *freedom of choice loophole* (Larsson, 2014, sect. 2). On the one hand, it could be possible for a signal to travel within the speed of light from one measurement setting to the other, influencing the measurement outcome on both wings by sharing information about the settings among the two sides. This explanation can be formulated by accounting for this communication via local hidden variables. On the other, information about the measurement settings may not be shared, but it may be predictable in such a way that the systems may somehow have access to it before reaching the measuring devices. Finally, it may be that the values of the measurement outcomes, as well as the particular choices of setting are deterministically fixed. In this case, at least the correlations among systems and measuring devices would have been produced in the distant past of these objects, again not requiring any non-local interaction.

As we will see in a dedicated section below, the series of experiments carried out by Alain Aspect and his collaborators in the early 80s constitute a landmark because they were the first that successfully addressed in a robust manner the locality loophole. However, the first time that any influence between the two separate wings of the experiment was successfully severed was in the test carried out by Weihs et al. (1998)[1]. In particular, this experiment closed both the *communication* and the *memory loopholes*. Three are the crucial arrangements that were made in this experiment:

[1] Similar to the experiment reported in Tittel et al. (1999).

first, the measuring devices were located at a distance that would prevent any signal travelling up to luminal speed to share information from one detector to the other. Second, the choice of settings were made in coordination with a physical random number generator, which made them completely unpredictable. Third, to ensure complete independence of the measurement outcomes on either wing, the results were recorded separately and merged only after completion of the measurement (p. 5039). The experiment was performed at the university of Innsbruck, and it involved pairs of entangled photons, produced via type-II parametric down-conversion and shot through glass fibres to two detectors at 400m of distance from one another. The state measured was a singlet state with an internal phase of π, used to test a version of the original CHSH inequalities. Despite the low detection efficiency — only of 5% — the experiment achieved a violation of the CHSH inequality by 30 standard deviations.

The third member of the family, the *freedom of choice loophole*, wasn't however closed in this test. It was already discussed in chapter 3, when introducing the assumptions from which Bell's theorem is derived. The premise of "no-superdeterminism" is precisely the assumption that the choices of the measurement settings are uncorrelated with the common state λ, and it was shown that a rejection of this assumption would lead to superdeterministic theories. This loophole is generally considered invincible on experimental grounds, as it is not possible for us to know whether it is in fact the case that every single object in the universe pedantically obeys a predetermined course of action which is specified up to the most minimal detail. Its power is therefore analogous to that of a skeptical argument that has in its nature not to allow for knock-down rebuttals. There are however powerful responses that challenge the strength of the superdeterministic hypothesis. Along the lines of the early arguments by Bell, Shimony, Horne and Clauser, the main argument usually advanced to dismiss this loophole is that a coherent superdeterministic portrait of any particular Bell test is characterised by strong conspirational tones; more importantly though, it undermines the very foundations of science, as in this framework the scientific method is devoid of any significance (Larsson, 2014, p. 15). By stressing the contrast between a superdeterministic world-view and scientific practice, one type of response is to "assumes away" superdeterminism. An equally effective approach is more tightly related to the experimental methodology, and it consists in undercutting practices that allow a superdeterministic explanation to thrive (Giustina, 2017, p. 498). Interestingly, a suggestion has been made that the freedom of choice loophole may be closed experimentally. Zeilinger (1999, pp. 492-493) says that if the two wings of the experiment were to be located at a distance of the order of "a few light seconds" from one another, and the choices of settings on each wing were made by a human such that they would be fruits of free will, there could not be joint events in the common past of the experimenters able to influence their respective local choices of measurement. Such an experiment would require a distance between the two wings that could be reached only in outer space. An experiment by Handsteiner et al. (2017) tries to emulate this principle by tying the choice of measurement settings to photons coming from distant stars in the Milky Way. More recent experiments that also claim to have closed the freedom of choice loophole are Scheidl et al. (2010) and Yin et al. (2017). Again, for the very nature of this particular loophole, these experiments can only close it up to a reasonable degree.

4.2.2 Detection loopholes

Differently from the previous group of loopholes, detection loopholes do not play a particular role at the theoretical level: whereas local causality and freedom of choice are also assumptions that need to be secured for the derivation of a Bell's inequality, the possible alternative explanations that we will consider in this section pertain the level of technological advance of the experimental apparatus. The members of this family are the *efficiency loophole* and the *coincidence-time loophole*.

The efficiency loophole, also known as detection loophole, is usually associated to Bell tests involving photons: by the very nature of these systems, the number of detected pairs of photons in the first experiments was very low with respect to the number of pairs actually produced. For this reason, although the correlations among the detected photons were still violating Bell's inequalities despite the low number of detected pairs, a hypothesis could be formulated that, *if the the detectors were more efficient, they could reveal that the inequalities are not actually violated*. This loophole is usually accompanied by the fair sampling assumption, which amounts to the claim that the measurement results, independently of the actual efficiency of the detectors, are at least a representative sample of a complete set of counts. This assumption potentially sanctions the reliability of the collected data, but at the price of opening another loophole: the undetected systems can be modelled as hidden variables, and it has been shown (e.g. see Pearle (1970), and Selleri and Zeilinger (1988)) that they can be used to build a local hidden variable model that agrees with the quantum predictions.

The detection loophole can be closed by improving the efficiency of the detectors, a solution that has been achieved with technological progress, and that also makes the fair sampling assumption less necessary. As we will see in section (4.3), over the years the experimental techniques related to the production of entangled systems and to the detection have increasingly improved the reliability of these experiments.

The loophole deriving from the fair sampling assumption on the other hand has also a separate theoretical solution. John Clauser and Michael Horne (1974), elaborated a type of testable inequality analogous to the original CHSH inequality, but which does not rely upon an ideal experimental apparatus with perfect efficiency: it considers in fact only the probability of detection, by calculating the expectation values only for two events, i.e. "detection" or "non-detection" of the system. In this manner, the minimal efficiency required to achieve a violation of the inequality is of 82.8%.[2] This type of inequality is more well known as Eberhard inequality, from the paper by Philippe Eberhard (1993), who produced a variation of the original inequality of Clauser and Horne that required only a 66.7% efficiency of the detectors in order to achieve a violation, in conjunction with a low background level. In particular photonic Bell tests use these rather than the original CHSH inequalities.

As mentioned above, there is also another issue that can be exploited to undermine the reliability of the experiment, known as *coincidence-time loophole*. This problem was individuated first by Hess and Philipp (2001) and discussed explicitly in Larsson and Gill (2004). In the experimental setting imagined by Bell and commonly referred to for deriving the inequalities, the variables considered are constituted by the experimental outcomes, the settings, and the past state of the entangled system. The experiment is embedded in relativistic space time, represented via the light-cone structure. What these arguments have shown is that there is another variable to be considered, which can both obstacle a derivation of the inequalities (Hess and Philipp, 2001), and that can be used to develop a local hidden variables model which agrees with the quantum predictions (Larsson and Gill, 2004). The variable is the time at which either experimental outcomes are recorded locally. For pairs of entangled systems, it is imagined that either detector records the results of measuring each subsystem separately at a certain time. Since the detectors are at space-like separation, the outcomes should be recorded at the same time, or within a reasonably small temporal interval. The way in which this temporal discrepancy can be represented is by assigning a local variable indicating the time of recording, which can then be related to that of the distant detector. The loophole discussed by Larsson and Gill (2004) is that it is possible to build a local hidden variable model in which the there is a local mechanism that produces either outcome at a specific time, without the outcomes

[2]This value is reported in (Eberhard, 1993, eq. 11), as the result of a calculation made by Mermin on the basis of the Clauser-Horne inequality.

being actually non-locally correlated, but still producing correlations equivalent to the quantum ones. Analogously to the detection loophole, this problem can be addressed via a fair-coincidence assumption, according to which the pairs of recorded subsystems are a sample of all the pairs that have been produced at the source. And, precisely as in the case of the detection loophole, this loophole can also be closed by deriving inequalities that do not depend on this assumption. This was first achieved by Larsson et al. (2014).

❊ ❊ ❊

The loopholes exposed in the previous pages affect Bell's theorem both from a theoretical and an empirical angle. Some have been individuated and discussed since the formulation of the theorem, such as the locality loopholes, while others have emerged as a consequence of the chosen systems to be tested, as for instance the detection loopholes. Whereas the quantum predictions were accepted as correct even before the performance of a loophole-free Bell test — the experiments of Aspects and collaborators (1982) are still considered the first reliable confirmation, although truly loophole-free tests were made only in 2015 — the existence of these possible obstacles to such an acceptance constituted sufficient ground for the formulation of arguments that needed to be addressed. With a clear idea of what these problems are under our belt, we have an interesting perspective from which to approach the following sections.

4.3 Testing Bell

The description of Bell's theorem from a theoretical point of view, as seen in chapter 3, is broad and cannot be easily confined — nor should it be — to a unique and reductive perspective. In particular, it was stressed that the derivation of the theorem is grounded on a large and layered set of assumptions, and that the predictions of quantum theory violate the inequalities derived from these assumptions. Importantly, it was stressed how this violation per se does not individuate which of these assumptions is to be rejected, leaving the problem of which part "has to give" underdetermined. The consequence of this fact is a large array of mutually conflicting interpretations, each of which provides a coherent world-view that emphasizes some aspects over others (see sect. 3.2). As a consequence of the list of theoretical interpretations of the theorem, one may legitimately expect that the empirical implications would be at least as many. But this is not the case: what a survey of experimental

Bell tests reveals is a consistent and widespread main practice. This latter involves the production of an entangled state with respect to a particular observable, for the vastest majority, pairs of photons entangled with respect to polarization, and an experimental setting that aims to test salient features of the correlations among the polarizations of the entangled pairs. Although at a theoretical level it may not be possible to claim whole-heartedly that Bell's theorem shows this or that particular fact of the world, the history of experimental tests is largely directed at producing correlations that are maintained and revealed at space-like separation. A hint that this is a central feature of experimental tests was provided already by the presentation of the locality loopholes above.

As already mentioned, given the existence of rich reviews of the experimental history surrounding Bell's theorem (such as Genovese (2005)), the present chapter has the freedom to approach the topic from different angles. The following pages are thus structured as follows. First, I will focus on how entanglement — the protagonist of this story — is experimentally produced, and how its features are tested. The history of how the formal prediction of entanglement was connected to a concrete, physical phenomenon is highly fascinating, and I will reflect further on its implications also in the final section of the chapter. I will then consider some representative Bell tests, performed on entangled photons via two different techniques, namely, cascade atom decay, and parametric down-conversion. In connection with the first technique I will focus on the renowned Aspect's experiments, while I will report the salient features of the three loophole-free Bell tests in connection with the second technique. At the end of this section, we will have all the instruments to proceed towards some philosophical considerations.

4.3.1 Entanglement in vitro – why photons?

In section 3.1.3 it was reported how it is possible to derive predictions from quantum mechanics that violate the Bell inequalities. In particular, it was emphasized that these predictions are formulated without referring to any particular physical system. For deriving the famous quantum expectation value of $2\sqrt{2}$ displayed in eq. (Violation) it sufficed to use Hermitian operators acting on state vectors living on Hilbert spaces, associated with the entangled state used to maximally violate the inequalities, namely, in that case, a singlet state displaying perfect anti-correlations. At the same time, this result leaves many possibilities open as to which physical system can be chosen to experimentally achieve this violation.

From an experimental point of view, pairs of entangled photons have historically been highly preferred to massive systems, and they are still the favoured choice in current Bell tests. But why is this the case? The mathematical expression of the entangled state does not "name" any system in particular, as the reference to state vectors is completely sufficient to make the correct predictions. This means that *any* system that behaves quantum mechanically can in principle be used to test the predictions of the theory. Therefore, the question of *how*, in the past century, it was understood that it was possible to induce entanglement among pairs of photons is not trivial by any means.

The first observable phenomenon to be concretely connected to entanglement was the spontaneous production of entangled photons in the process of particle annihilation.[3] This latter was predicted by Paul Dirac in his paper "On the annihilation of electron and protons" (1930), where he described the effect of the interaction between an electron and a positron (erroneously thought to be the proton in the article). The interesting properties of the electron for this argument are that they have positive kinetic energy and negative charge. At the time it was known that in consequence of the quantum character of energy, electrons would occupy discrete energy levels in the atom, and that, for Pauli exclusion principle, each level would contain only one electron at a time. In this picture however, the theory allows for one energy level to not be occupied by an electron, but rather by a system of equal mass but opposite charge, the positron. In the paper, Dirac shows that if an electron happens to interact with a positron, both systems annihilate, and, for the principle of the conservation of energy and momentum, their joint energy would be converted in electromagnetic radiation. This mechanism, together with the probability of occurrence of the spontaneous annihilation, is rigorously derived by the principles of quantum mechanics. The presence of the positron, previously doubted, was more robustly theorized in a paper the following year (Dirac, 1931), where the effects of the electron-positron annihilation were confirmed.[4] In the subsequent years, the correlation among the

[3]Since it is of philosophical interest for the present thesis, I will discuss the connection between entanglement in its mathematical form and the physical phenomena in more detail in section (4.4). Francisco Duarte, in his book Duarte (2019) provides a historical introduction on the development of the concept of entanglement in physics, and he individuates two avenues, which he names "philosophical" and "physical". While these two paths eventually converge, he claims that they were initially not obviously connected. The physical path, which he traces via Dirac, Wheeler, Pryce-Ward, Snyder-Pasternack-Hornbostel and Wu-Shaknov, is at the basis of the experimental realization of what became the canonical way to test entanglement, namely, the production of pairs of entangled photons with opposite polarizations. The philosophical path starts with Einstein-Podolsky-Rosen, followed by Schrödinger, Bohm-Aharonov, and Bell, and it underlies the theoretical understanding of the implications of entanglement.

[4]"An encounter between two hard γ-rays (of energy at least half a million volts) could lead to the creation simultaneously of an electron and anti-electron, the probability of occurrence of this process

polarization of the two γ-rays emitted as a consequence of the electron-positron annihilation gained increasing interest. In an article that studied the features of Dirac's predictions concerning electron-positron annihilations, John Wheeler (1946) proposed an experiment that could test such predictions by measuring the polarizations of the two photons emitted in the annihilation process. The premise of the proposed experiment is a clear acknowledgement of the entanglement correlations among the photons:

> We have already remarked that by far the dominating type of annihilation is that in which the positron combines with an electron whose spin forms a singlet state with respect to the spin of the positron. Associated with this selection of pairs which have zero relative angular momentum, before the annihilation process, is an analogous polarization phenomenon in the two quanta which are left at the end of the process. According to the pair theory, if one of these photons is linearly polarized in one plane, then the photon which goes off in the opposite direction with equal momentum is linearly polarized in the perpendicular plane. (Wheeler, 1946, pp. 234-235)

He proposes an experiment in the following paragraphs, and suggests the values of the angles at which two scatterers at two opposite ends of the source of electron-protons pairs should be set for the correlation among the photons to be the strongest. In a subsequent paper, Maurice Pryce and John Ward (1947) proposed an experiment to perform this test, with a corrected calculation of the angles, on the basis of Wheeler's suggestion; interestingly, the diagram reporting their proposed experiment appears to be the first representing a photonic experiment testing entanglement, at least according to Duarte (2019, p. 5-2). In a paper published the following year, Hartland Snyder, Simon Pasternack and John Horbnostel (1948) reflect upon the experiment proposed by Wheeler, and stress the fact that the analysis of the polarization of one member of the emitted photons "gives information about the initial state of polarization of *the other photon*." (p. 441, italic added).

The first experiment that decisively confirmed[5] all these predictions was performed by Chien-Shiung Wu and Irving Shaknov in the middle of the last century (Wu and Shaknov, 1950). The experimental set up, reported in fig. 4.1, can be briefly summarised as follows. A source of positrons (S) composed by the isotope copper-64

being of the same order of magnitude as that of the collision of the two γ-rays on the assumption that they are spheres of the same size as classical electrons." (pp. 61-62). Since this process is reversible, it is equivalent to state that the encounter of an electron and a positron would cause the creation simultaneously of two γ-rays. The existence of the prositron was experimentally confirmed soon after, see Anderson (1933).

[5]Two other experiments had been performed before, namely, Bleuler and Bradt (1948) and Hanna (1948), but the margin of error affecting their results were considered too high to allow a comparison with the theoretical predictions.

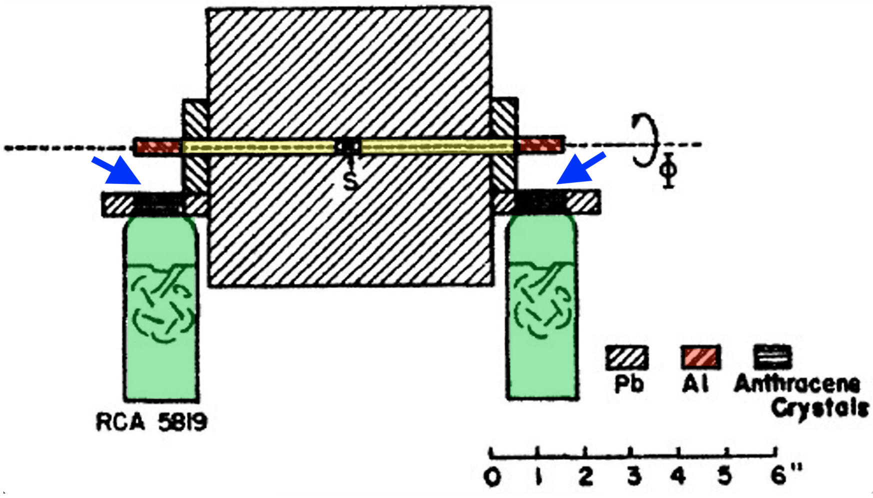

Figure 4.1: Diagrammatic representation of Wu-Shaknov experiment, from Wu and Shaknov (1950), colors added. It is considered the first robust experimental reproduction of entanglement, according to the predictions of Dirac's *pair theory* (Dirac, 1930, 1931), and the thought experiment elaborated by Wheeler (Wheeler, 1946).

was isolated at the center of a block of lead (the bigger shaded square in the picture) with two narrow channels drilled at two opposite ends of the isotope, highlighted in yellow. The positron source was radiated via deuteron bombardment to stimulate the spontaneous beta-decay to which copper-64 is subject. The electron-positron annihilation generated at the source resulted in the electromagnetic radiation constituted by two γ-rays which were collimated through the two narrow channels of the lead block. Once the photons exited the channels, they would hit two aluminium scatterers (highlighted in red) which would be oriented at a perpendicular angle with respect to one-another in order to reach the highest value of coincidence counts of the photons (the rotation angle is indicated by the letter Φ in the diagram). Two detectors counting the photons would be located at either end of the experimental set up. Thanks to the efficiency of the newly introduced scintillation counters — composed by two anthracene crystals, indicated by the blue arrows, and two photomultipliers highlighted in green — the data obtained were in very high agreement with the predictions of the theory in contrast with the previous two experiments. The need for scatterers in the analysis of the γ-rays is due to the fact that it is not possible to directly analyse the polarization of these highly energetic photons. The interpolation of the two scatterers allowed to subtract some of the original energy from the photons, and to deflect them towards the detectors only at a second stage in the experiment. As Bohm and Aharonov (1957, pp. 1074-1075) specified, the scattering angles of the two photons were calculated via the Klein-Nishina formula, which allowed to derive the probabilities for the coincidence counts at selected angles.[6] The indirectness of this process was one of the problems of the first Bell tests as well, and it was solved by choosing different decay processes for the production of entangled photon pairs — see Genovese (2005, p. 339).

Almost a decade after the Wu-Shaknov experiment, David Bohm and Yakir Aharonov wrote two papers (Bohm and Aharonov, 1957, 1960) in which they regard this experiment as a concrete test of the situation described by EPR. In these works they focus on the correlations among entangled systems predicted by the theory, formalised by Dirac and Schrödinger. Whereas such a prediction was undisputed at the time, the fact that the correlations may disappear if the systems were separated in space was still an open possibility. Crucially, Bohm and Aharonov reported Wu and Shaknov's experiment as the empirical confirmation that the predictions of quantum theory are correct, and that the entanglement correlations are not affected by spatial distance.[7]

In the light of all this, the question as to why the majority of Bell tests involves the production of entangled photons can be answered as follows. On the theoretical side, entanglement was recognized early on as a prediction of quantum mechanics, which follows from the properties of the mathematical objects representing systems and their interactions within quantum mechanics. The connection between the prediction of these correlations and a particular physical phenomenon originated in Dirac's pair theory, but was then recognized and explored by Wheeler, followed by Pryce and Ward, and by Snyder, Pasternack and Hornbostel, who all accompanied their theoretical studies with suggestions for a concrete experiment in which the polarizations of the electromagnetic radiation could be experimentally tested. On the experimental side, this test was already technologically feasible, as the electron-positron annihilation can be triggered at low energies and the study of the electromagnetic radiation was already well established at the time. The improvement on the detectors in the Wu-Shaknov experiment was the factor that determined the collection of low-error data, which clearly corroborated the quantum predictions. By the time the first Bell test was made, twenty-five years after Wu and Shaknov's experiment, the existence and the features of photonic correlations was well-established in the experimental practice, and relatively easy to reproduce in a laboratory. Given the success of these experiments, the use of photons remained the main practice, although Bell tests

[6]The connection between the polarization of the γ-rays and the scattered photons is explained very clearly in the introduction of Kasday et al. (1975), which reports a positronium decay experiment aimed at testing local hidden variable theories.

[7]These works by Bohm and Aharonov also sanction the encounter between the *philosophical* and *physical* way to entanglement according to Duarte (2019, 1-4), as stated in footnote 3 above.

involving massive particles also exist, but they are generally experimentally more demanding. Let us enter in the matter now of the first Bell tests, which led to Aspect's experiments.

4.3.2 Bell tests I

The first experiment to claim a successful violation of Bell's inequalities in agreement with quantum mechanics was performed by Stuart Freedman and John Clauser (Freedman and Clauser, 1972). It is also a photonic experiment that analyses the polarization correlations, but it exploits another type of phenomenon, namely, cascade calcium decay. The advantage of this technique over positronium decay is that the photons emitted in this case have a frequency that is within the visible spectrum, thus their polarization can be analysed directly, without the need for an intermediate scattering process. On the other hand, these photons are of more difficult detection, to the extent that the first experiments were less than one percent efficient: in this experiment, only 0.4% of the emitted photons could be detected. The setting was the following: a source of calcium stimulated by a beam triggered a continuous emission of entangled photon pairs, each member of which interacted separately with rotatable and removable polarizers, placed symmetrically with respect to one another, and were then recombined into a single detector. The essential datum was constituted by the detected coincidence rate, which can be calculated as a function of the angles of orientation of the two polarizers, by also considering the case where one polarizer was removed. As a result, they calculated the coincidence rate predicted by quantum theory on one hand, and as it would be restricted by the same constraints imposed on the CHSH inequality on the other.[8] The evidence they obtained was in striking agreement with the quantum predictions, and in clear contradiction with the classical ones, reason why they conclude by stating "We consider these results to be strong evidence against local hidden-variable theories." (p. 940).

Despite the success of the experiment, this setting had some limitations, which can be individuated in terms of the loopholes discussed above. By not detecting each photon at space-like separation, and for using fixed choices of settings, this experiment left open both the communication and the memory loopholes. Moreover, due to the low detection efficiency on one hand, and to the use of an inequality that is based on the fair-coincidence assumption — as the CH inequality would have been

[8] In fact, the inequality encoding the "classical" prediction was derived explicitly as a form of the CHSH inequality. Cf. equations (2) and (3) for the classical prediction, and eq. (1) for the quantum prediction, p. 939.

h

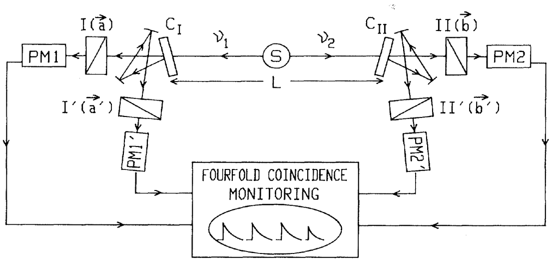

Figure 4.2: Set up of the experiment, diagram taken from Aspect et al. (1982, p. 1802).

derived only two years later — the results were subject also to the efficiency and the coincidence-time loophole. However, as we have seen in section 4.2, the more poignant and severe family of loopholes is constituted by the locality loopholes, which have theoretical pregnancy, and do not depend exclusively on technological limitations.

This is why the results achieved by Alain Aspect and his collaborators in the early 80s constituted a break-through in the history of Bell test. The main result is reported in Aspect et al. (1982), and it is the experiment Aspect made with Jean Dalibard and Gérard Roger. The main achievement of the experiment is that for the first time it reasonably closed the locality loophole, and it achieved also the highest violation of the Bell's inequalities. The set up is similar to two experiments Aspect carried out the year before with Philippe Grangier and Gérard Roger (Aspect et al., 1981, 1982). By exploiting cascade atomic decay with calcium, these experiments tried to actualize the situation described in the original CHSH paper, where entangled systems correlated according to the CHSH inequality were analysed separately, and the statistics of the measurement outcomes could reveal the contradiction with Bell's theorem's assumptions. The experimental setting of Aspect et al. (1982) is the following.

The source of entangled photons was located in between two optical devices, each composed of an optical switch (C_I, C_{II}) which could re-direct the photons towards two different polarizers (I(a), I′(a′), II(b), II′(b′)), followed by photomultipliers (PM1, PM1′, PM2, PM2′), which finally directed the detected systems towards a computer monitoring the coincidence counts (fig. (4.3.2). The distance L between the two optical devices was of 12m. The innovation of this experiment was the addition of the optical switches, which would rotate *while the systems were in flight*,[9] combined with a distance between source and detectors which allowed to create a space-like sepa-

[9]Reference to the possibility of changing the measurement settings before the particles reached the detectors was made already in Bohm and Aharonov (1957, p. 1070).

rated interval between the entangled photons at the time of detection. The switches would rotate at a rate related to a standing wave propagated in water connected to the switching device (for details of this mechanism see p. 1806 of the paper), and each switch would be uncoordinated with the other. Although either switching rate was periodical, and thus not truly random, the space-like separation and the rapid change of measurement setting made this the very first experiment to approach the requirements of a real test of local hidden variable theories. The predictions were formulated in form of an inequality analogous to the original CHSH inequality, which was confronted with the predictions of quantum theory for the same relation. The evidence obtained by Aspect and collaborators agreed almost exactly with the quantum predictions, and achieved a violation of the Bell inequality of five standard deviations. Despite the great achievement of this experiment, in particular, the closure of the communication loophole, some loopholes were still open. On the one hand, given the periodicity of the switching rate of the measurement settings, the memory loophole was not completely closed. On the other hand, the detection loopholes were left all open: the photons produced in the cascade atom decay are very difficult to detect,[10] and the inequality used in the prediction were not of the form of the CH inequality, leaving thus open the coincidence-time loophole as well. Part of the limitation to which this experiment was subject was precisely the source of entangled photons. As we will see in the following section, the development of parametric down-conversion allowed to make considerable progress in the development of Bell tests, leading eventually to three successful loophole-free experiments.

4.3.3 Bell tests II

A significant technological improvement in the history of Bell tests is represented by the use of a new technique for producing entangled photon pairs, namely, the process of *parametric down-conversion*.[11] The name indicates an eminently quantum effect that had been studied, both theoretically and experimentally between the 1960s and 1970s, although it started to be consistently implemented in Bell tests in the 90s. It consists of a type of photon decay that takes place when high-energy photons, usually

[10] The coincidences in this experiment were recorded in an interval between 0 and 40 hertz, against the typical rate of production of entangled photon pairs for this atom decay, which is of $5x10^7$ hertz (Genovese, 2005, p. 341).

[11] The exploration of this part of the history of Bell tests will be even less thorough in reporting all the relevant experiments that were made after those of Aspect, as the number of these is rather large. I focus instead on the salient features of this experimental technique, and especially on the three loophole-free tests. For a complete history see, again, the review from Marco Genovese (2005) and the rich list of references therein.

carried in laser beams ("pumps"), interact with materials such as nonlinear crystals that do not preserve the polarization density of the electric field. In particular, the product of this decay consists in a pair of entangled photons that conserve the energy and momentum of the original photon, and have correlated polarizations (they are parallel if produced in parametric down-conversion of type I, and orthogonal if in type II). Interestingly, the entangled photons are produced within a cone that has a very narrow angle due to the way in which energy-momentum conservation acts in this particular decay, which makes the photons also *spatially correlated*. For this reason, photons produced in this manner are subject to a significantly higher detection rate with respect to those produced in cascade atom decay.

The first Bell test involving parametric down-conversion was proposed by James Franson, and it exploited the photon phase-momentum entanglement in connection with the phenomenon of optical interference (Franson, 1989). By designing a cirquit for either entangled photon in such a way that the photon pairs are detected only if both are transmitted by a beam-splitter, it is possible to calculate the rate of coincidence counts in such a way as to derive a CHSH-type inequality that is violated by quantum mechanics. The relevance of this proposal is that what became the "Franson scheme" was then realized in a series of experiments testing the Bell inequalities. The first experiments that closed the communication loophole[12] used this particular scheme. Experiments analysing the polarizations of entangled photons produced via parametric down-conversion were also realized (see in particular Ou and Mandel (1988) and Kiess et al. (1993)). What is relevant of this particular technique however, is the fact that it was employed in two of the three loophole-free Bell tests, to which we are turning now.

Annus mirabilis 2015: Loophole-free Bell tests

It is worth focussing on these experiments because they close the loopholes we discussed in the previous pages *at the same time*. All three were reported in 2015, two of which were submitted to *Physical Review Letters* on the same date.

Let us start from the first, "*Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres*", performed by a research group at Delft University of Technology (Hensen et al., 2015). The technique used to produce the entangled state violating the inequality in this case is *entanglement swapping* (see Żukowski et al. (1993)), which consists in creating two separate and independent sources of bipartite

[12]Namely, Weihs et al. (1998) and Tittel et al. (1999).

entangled states: in this case, the sources were constituted by electrons in a particular spin state, which would emit pairs of entangled photons upon suitable excitation. Importantly, the photon pairs are entangled also with the spin of the electron that emitted them. One photon of either pair eventually interacts with one member of the other pair in such a way that they become entangled — for instance, by interacting with a beam-splitter. As a consequence, the two independent electrons are projected in an entangled spin state, thus a spin measurement on the electrons at this point would display quantum correlations. Entanglement swapping is particularly helpful for designing an experiment that intends to securely close the locality loophole: not only the two independent sources can in principle be separated at an arbitrary distance, but the systems that eventually become entangled do not share a common memory of what had happened individually to either of them — as also stressed in Hensen et al. (2015, p. 686).

In more detail, the set-up of the Delft group was the following. Two different laboratories (A, B) containing the independent sources of entangled pairs were located at a distance of 1.280km from one another, while a third laboratory (C) was located at an intermediate but inequivalent distance from A and B. The entangled photon pairs were prepared by stimulating with a particular beam a nitrogen vacancy in synthetic diamonds.[13] The excitation of the outermost electrons in the nitrogen ions triggered a decay that releases as a product pairs of photons that are entangled with the spin of the electrons. These latter are highly controllable in this particular set-up, feature that makes it possible to clearly distinguish among two choices of measurement, and to prepare a particular spin state on either source. After two members of either photon pairs have interacted at location C, given entanglement-swapping the spins of the electrons located at A and B are projected into a maximally entangled state, which can then be measured at space-like separation to observe violations of the chosen Bell's inequality. The choices of settings on each side were made via random number generators connected to fast switches, such that the choices would be unpredictable and random on either side. In this particular measurement setting, the frequency of emission of entangled photons on both sides in such a way as to achieve entanglement swapping was approximately of one per hour. Thus, to record the 245 Bell tests made in this experiment, the apparatus ran for a total of 220 hours over 18 days.

From a theoretical point of view, the obtained spin state after swapping is exactly the singlet state $|\Psi^-\rangle = (\uparrow\downarrow) - (\downarrow\uparrow)/\sqrt{2}$, which was measured in two different bases, named Z and X, in such a way as to have two choices of measurement settings

[13]This technique was implemented in an experiment reported in Robledo et al. (2011), which already contains the suggestion for a Bell test.

on each side (indicated with 0 and 1), which could yield in turn two measurement outcomes (labelled 0 and -1). The inequality tested was a CHSH inequality, with the canonical bound of $S \leq 2$; the predicted quantum violation for this experiment was of $S = 2.30 \pm 0.07$, whereas they actually achieved a value of $S = 2.42$.

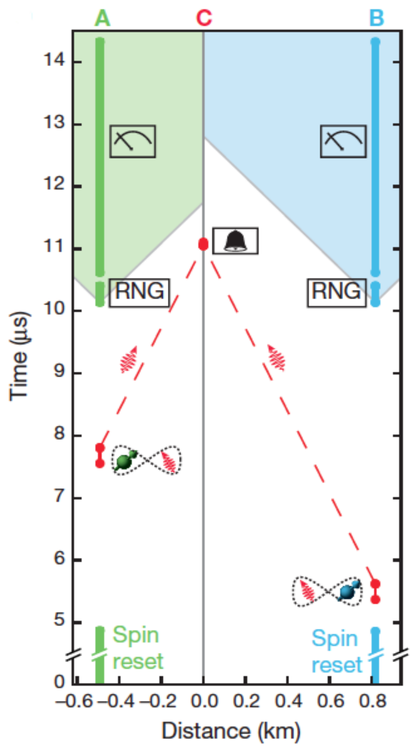

Figure 4.3: Space-time diagram of the Bell test of Hensen et. al. (2015, p. 684). The dashed red lines indicate the space-time trajectories of two photons, eventually converging in a common detection event at C (bell symbol). The choices of settings are made independently of the common detection event (at RNG) and the recordings of the two spin measurements happen at space-like separation (pointer symbol).

The locality loopholes, with the exception of the freedom of choice, have been closed, as the measurement were performed at space-like separation, and the measurement settings were chosen randomly; moreover, the detection loopholes were closed via the implementation of a highly efficient system for detecting the electrons. This is what allowed the use of the original CHSH inequality, as opposed to the adapted Eberhard inequality.

The second experiment "*Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons*" (Giustina et al., 2015a) was performed by a research group at the Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences in Vienna. The Bell tests in this case consisted in polarization measurement of entangled photons produced via parametric down-conversion of type II, measured at two space-like separated locations, a setting by now more familiar than the previous one.

In more detail, the source of entangled photons was composed by a periodically poled crystal excited by means of a polarized pulsed diode laser, which was directed at both ends of the crystal. Each entangled photon pair was split by a beam-splitter, which collected each member of the pair in two different optical fibres leading to two measurement apparatuses located at 58m from one another. Each wing of the experiment was equipped with a switch for the angle of polarization at which each photon's polarizations would be measured, correlated with a random number generator for the change in the choice of measurement. The out-coming photon, after being transmitted by the measuring

h

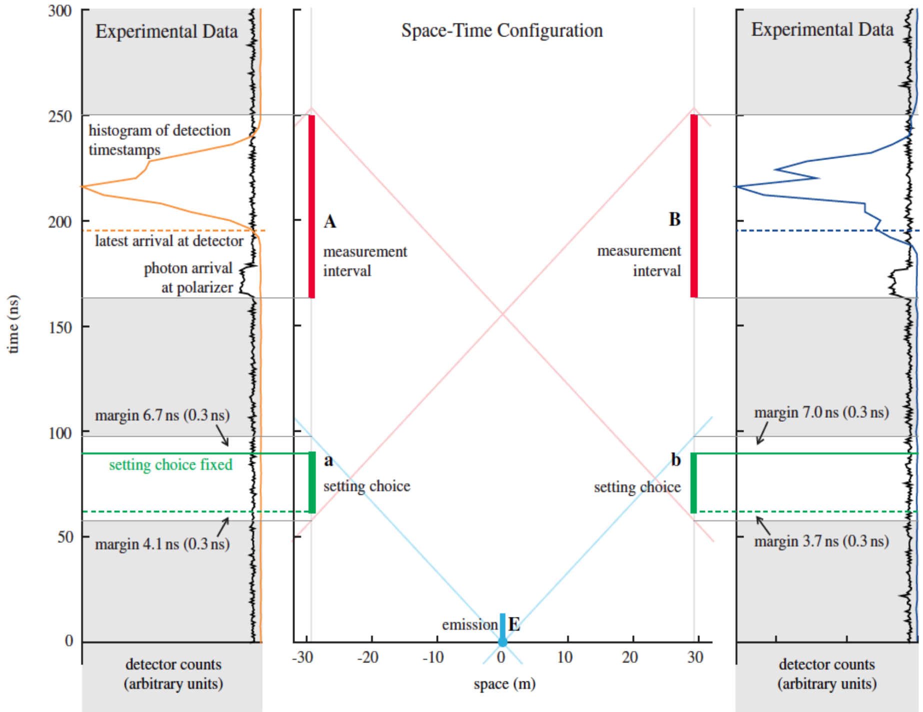

Figure 4.4: Space-time diagram of the Bell test of Giustina et. al. (2015a, p. 4).

polarizer would reach a transition-edge sensor, which is a single-photon detector. The measured polarization of either photon, together with the associated choice of measurement, were independently recorded by a computer. The number of photon pairs produced in the non-linear crystal was of 3500 per second. The experiment was run for a total of 4.8 hours, subdivided in two blocks of 1 hour each, plus a third block of 2.8 hours.[14] The number of valid trials recorded in this block was of over 3.5 billions.

The theoretical framework was different from the previous experiment, as photonic experiments are subject to a low detection efficiency. The inequality tested in this case was thus a CH-Eberhard inequality of this form (p. 5):

$$J \equiv p_{++}(a_1 b_1) - p_{+0}(a_1 b_2) - p_{0+}(a_2 b_1) - p_{++}(a_2 b_2) \leq 0. \tag{4.1}$$

The terms of the equations are probabilities, and are to be read as follows: the lovercase indexes in the probabilities indicate the two possible measurement outcomes, "0" for no-detection events, and "+" for detections, whereas a and b indicate the two

[14]For the details of the experiment, cf. Giustina et al. (2015b).

possible choices of settings, 1 or 2, on either side of the experiment. If Bell's theorem's assumptions hold, the correlations should be bounded by the value zero. The state used to violate the inequality was of this form, where "H" and "V" stand for the orientations of the polarizers, namely, horizontal and vertical:

$$|\Psi\rangle = \frac{1}{\sqrt{1+r^2}}(|V\rangle_A|H\rangle_B + r|H\rangle_A|H\rangle_B). \tag{4.2}$$

The value "r" is related to the actual measurement setting, and it was set to zero when the state was not entangled, i.e. when it was a product state, and to -1 when it was instead maximally entangled. The predicted value for the violation of eq. (4.1) was of $J = 4\times10^{-5}$, and the measured value was $J = 7.27\times10^{-6}$: the inequality was clearly violated, in accordance with the quantum predictions, and in contradiction with Bell theorem's assumptions. Thanks to the space-like separation of the detectors, the choice of setting based on fast-switching polarizers correlated with random number generators, and the use of a CH-Eberhard type inequality, all the locality loopholes are closed in this experiment. The conclusion even contains a remark addressed to the freedom of choice loophole.

> By closing the freedom-of-choice loophole to one natural stopping point — the first moment at which the particles come into existence — we reduce the possible local-realist explanations to truly exotic hypotheses. Any theory seeking to explain our result by exploiting this loophole would require λ to originate before the emission event and to influence setting choices derived from spontaneous emission. It has been suggested that setting choices determined by events from distant cosmological sources could push this limit back by billions of years." (Giustina et al., 2015a, p. 5).

The detection loopholes were instead addressed both with the use of an inequality that does not rely upon assumptions as fair-sampling — which relates both to detection and coincidence-time loopholes — and with an adequate experimental setting producing a very large amount of entangled photon-pairs.

The final loophole-free experiment we are going to consider in this section is also a photonic experiment, carried out by a research group at the National Institute of Standards and Technology in Boulder, Colorado: "*Strong Loophole-Free Test of Local Realism*" Shalm et al. (2015a). The setting closely resembles the experiment just considered, but there are some differences.

The entangled photons are produced almost in the same manner, via parametric down-conversion, after which either member of the photon pairs are sent to the two measurement locations, at almost the same distance from the source, and at a distance of 184.9m from one another. The random number generators correlated

h

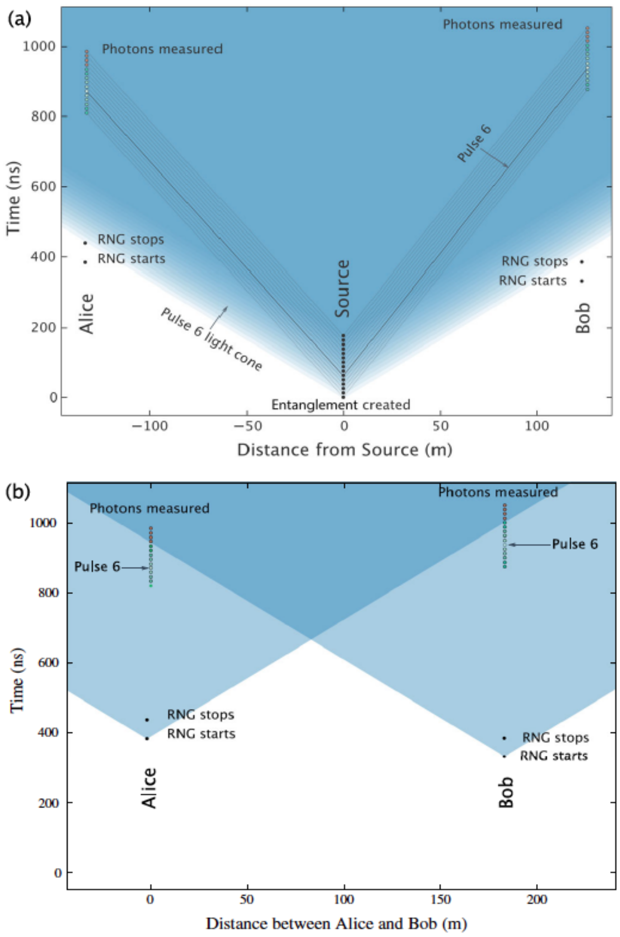

Figure 4.5: Space-time diagram of the Bell test of Shalm et. al. (2015a, p. 5), showing (a) the light-cone of the source with respect to the two measurement locations, and (b) the light-cones of the separate measurement events.

with a fast-switch for the choice of polarization angle at which the photons will be measured are located at the two measurement stations. The fast switches are constituted by a device called Pockels cell, which were activated only when the random number generator would output a "1" choice of measurement. After the incoming photons were measured separately, they were re-directed towards a superconducting nanowire single-photon detector, which transmitted the signal to a computer that recorded also the measurement setting, the time of synchronization with the detection from the other wing, a signal from a GPS.[15] This experiment was carried out over two days, and the selected data for the result were those belonging to the last block of measurement, which lasted 30 minutes. The number of entangled photons was of 99.100 per second, at a rate of 12.6 nanoseconds; however, the Pockels cells are activated at a rate of around every 200 nanoseconds, more or less fifteen times slower than the rate of production of the entangled photons. For this reason, each experimental trial was divided in 15 photon counts, as indicated in figure (4.3.3), where it is possible to see that for the alignment of each detector with respect to the source, the event count that was the most space-like separated was the 6^{th}.

The theoretical framework of this experiment resembles again that of Giustina et. al. (2015). The inequality tested is in fact exactly the same as eq. (4.1), whereas the state prepared for a maximal violation of this inequality in this set up was the non-maximally entangled state:

$$|\Psi\rangle = 0.961|H_A H_B\rangle + 0.276|V_A V_B\rangle, \tag{4.3}$$

where "H" and "V" stand for horizontal and vertical axis of polarization respectively. The observed value of violation of the inequality was subdivided by number of pulses, i.e. of produced entangled pairs, and all achieved a clear violation of the CH-Eberhard inequality and were in good agreement with the predictions of quantum theory.[16] The loopholes were closed as in the previous experiment: the detectors were at space-like separation, the choices of measurement were made randomly and also outside the past light-cone of the opposite measurement event, the detectors employed were highly efficient and the inequality tested dispensed of the fair-sampling assumption. Also this paper makes reference to the freedom of choice assumption, making the observation that the superdeterministic hidden variables in this experiment would need to be correlated with elements involving cultural references, such

[15]For more details concerning the experiment, cf. Shalm et al. (2015b).

[16]In detail, the values were of 2.5×10^{-3} for a single pulse, 2.4×10^{-6} for three pulses, 5.8×10^{-9} for five pulses and for seven pulses 2.0×10^{-7}. (p. 6)

as the titles of famous TV series and movies that were connected attached to the random number generators (p. 5).

4.4 About the theory's unseizable link with the world

In this chapter we have seen how entanglement has become the experimental resource to show that Bell's theorem is correct, and that the world is quantum. Regardless of which interpretation of the theorem one may endorse, there is no escape from this fact, neither theoretically, nor experimentally. What I am proposing in this section is a philosophical reflection on the stock we can take from what has emerged so far. The starting point is thus entanglement: in particular, the appreciation of how sophisticated is the passage from reading a state that cannot be decomposed into a product, to pointing at a particular phenomenon that this state represents. The main role of section 4.3.1 was precisely to reply to the question of how this bridge was discovered, or built in the first place. My depiction of such a bridge was guided specifically by two factors, namely, the experimental path to testing entanglement on one hand, and the use of photons on the other. The arguments I present in what follows move our attention to a more general level, as they constitute a reflection on the metaphysical implications of the story we have seen in sect. 4.3.1.

Let us start from the facts: the historical sequence of relevant events unveils some peculiarities. As we will see in a moment, an entangled state emerges naturally in consequence of the formal structure of the theory, which had been built already in the middle of the 1920s (Born et al., 1926), (Schrödinger, 1926). Recognition of the meaning of entanglement — as well as the coinage of its name — and of what this particular state entails, begun a decade later,[17] and fully hit the physical community with the inexorability of Bell's theorem. But the discovery of a phenomenon that could be fully described by an entangled state (e.g. the production of pairs of entangled γ-rays in an electron-positron annihilation) was made in 1930, *before* this entanglement awareness really started to take off. The connection between the properties of entanglement recognized explicitly first by EPR and Schrödinger on one hand and experiments revealing them on the other became established only after the brilliant intuition of Bohm and Aharonov in 1957, and the first publication of Bell's theorem in 1964. In particular, it should be noticed that in the experimental literature, correlations among pairs of photons produced by atomic decay started to be explicitly phrased in terms of the EPR paradox only in the 1970s. With this I

[17] Naturally, I am referring to EPR (1935) and Schrödinger (1935; 1936).

refer to the fact that the experiments we took into account as the first examples of production of entanglement (representatively, Wu-Shaknov (1950)), do not cite the Einstein-Podolski-Rosen paper, despite being carried out an abundant fifteen years after the publication of this latter.

So, if it is true that (a) the formal structures for deriving an entangled state had been present all along from the early 1920s, (b) the prediction of a physical phenomenon that would then have been identified as an instance of entanglement was formulated already by Dirac in 1930, (c) recognition of the problematicity of entanglement was sanctioned by EPR and Schrödinger in the mid-1930s, and (d) a photonic experiment that corroborated Dirac's prediction was concluded already in 1950, why is it the case that a clear identification between entanglement as a theoretical representational structure on one hand and a testable physical phenomenon on the other was established clearly only in 1957 by Bohm and Aharonov, and then definitively with the formulation of Bell's theorem?

Let me specify this question further. Despite the generality of the mathematical formulation of quantum mechanics, the theory was born in reaction to the discovery of anomalies in radiation phenomena that could not be described by classical physics — thus, in clear connection with a whole group of empirical phenomena; of course, my clear focus on entanglement notwithstanding, I am not suggesting that more generally the connections between the theory and physical phenomena were established only half a century after its birth. More precisely, by enquiring about how the experimental history of Bell tests unfolded, it has emerged that the link between entanglement as a mathematical prediction and a physical phenomenon that would empirically corroborate such a prediction *was by no means a trivial connection*: the chronological order of the salient events in this story, reduced to a minimum in points (a)-(d), seems to support this intuition.

In the previous pages we have examined the works that have paved the way for photonic experiments that would have eventually constituted the basis for definitive Bell tests.[18] Interestingly, all of them refer to the particular correlations preserved by the spatially separated photons, *without* making reference to either EPR or to Schrödinger's paper. Similarly, neither EPR nor Schrödinger take into account Dirac's paper on electron-positron annihilation, published five years earlier, as an instance where an entangled state is produced. Moreover, also the experimental articles only refer to the predictions made in particular by Wheeler on the grounds of Dirac's pair

[18]These were Dirac (1930), Wheeler (1946), Pryce and Ward (1947), and Snyder et al. (1948) on the theoretical side, and Bleuler and Bradt (1948), Hanna (1948) and Wu and Shaknov (1950) on the experimental side.

theory, but without mentioning EPR or Schrödinger's works. It seems unlikely that these two sides were unaware of each other's works, thus the explanation for this apparently missing ring must be elsewhere. The minimal hypothesis I am going to focus on now is precisely the claim that the connection between these two sides is non trivial, in the sense that such a connection *needs to be actively made*.

To understand what this means, let us look more closely at how entangled states are formulated. As specified in section 2.3.1 in chapter 1, quantum theory has among its very pillars that it represents physical systems by means of Hilbert spaces rather than of a classical phase space. The structure of Hilbert spaces displays a crucial feature for defining an example of how entanglement differs from classical correlations, namely, it is not the case that the state of a multi-partite system can be written as the Cartesian product of the states of its subsystems: the Hilbert space of the multi-partite system is the tensor product of the Hilbert spaces of its subsystems (Horodecki et al., 2009, p. 873). For a bi-partite system composed of systems a and b, with respective Hilbert spaces $\mathcal{H}_a, \mathcal{H}_b$, the joint space can be written in form of a tensor product: $\mathcal{H}_{ab} = \mathcal{H}_a \otimes \mathcal{H}_b$, which means that the joint state $|\Psi\rangle_{ab}$ of subsystems a and b can be written as:

$$|\Psi\rangle_{ab} = \sum_{i_a, i_b} c_{i_a i_b} |i_a\rangle \otimes |i_b\rangle, \tag{4.4}$$

where the $|i\rangle$ are basis vectors of the respective sub-Hilbert spaces, and c are the complex coefficients associated to the vectors. The state described in the general form of eq. (4.4) cannot be rewritten by assigning a particular state to subsystems a and b, thus the following condition holds:

$$|\Psi\rangle_{ab} \neq |\psi\rangle_a \otimes |\psi\rangle_b. \tag{4.5}$$

The structure of entanglement is not a feature of the Hilbert space formulation of quantum mechanics, but a central characteristic of the theory itself — see (Earman, 2015) for an exposition of the algebraic formulation of entanglement. For understanding what is entanglement within quantum theory, one does not need to know more than what is stated by these definitions, and the key to understand the meaning of these definitions is located in turn within the mathematics necessary to set up these structures. Rather than pointing our gaze towards the world, to understand what is entanglement we seem to be requested to look at the sophistication of constructs that are eminently mathematical.

However, quantum mechanics is a physical theory, and in particular Bell tests demonstrate that the correlations that arise among the relevant observables measured in experiments on entangled systems do have the features that are encoded in the mathematical representation of such states. The claim I want to put forth in light of the historical evidence leading to Bell tests on one hand, and the definitions of entanglement just considered on the other is that *the connection between the quantum mechanical formulation of the entangled state and a natural phenomenon can be made only on the basis of extra assumptions that exceed the domain of quantum theory*.

To unpack what I mean with this claim, let me proceed with some specifications. First and foremost, this does not mean that this connection has an approximate nature: what I intend is not that the theory provides imprecise predictions. Instead, the evolution of Bell tests has shown that more sophisticated technology and improved experimental techniques have provided data that have increasingly converged with the theory's predictions. Second, although the axioms of quantum theory — minus the measurement postulate — are exclusively mathematical, I am not claiming that the theory *could be*, or *is to be* considered as pure mathematics. In other words, I am not challenging the physical status of quantum mechanics as a physical theory. On the other hand, what I am arguing for is that the theory per se is not committed a priori to the description of any particular thing. The definition of entanglement and the time it actually took to connect it to already existing photonic experiments is only one instance of this fact.

But what is really interesting about this apparent gap between the mathematical description and its physical significance? Since the entangled state is the general form of a quantum state, my claim about the arbitrariness of its connection with particular physical phenomena have more general consequences. The second claim I am advancing is that *quantum theory is not a theory about any object in particular, despite being a physical theory*.

Perhaps differently from the first claim I put forth, this second statement should not be surprising: there are in fact existing positions in the philosophy of physics that seem to align quite naturally with this idea. For example, a type of classification of physical theories that started already with Einstein (1919) and has been explored by Flores (1999) and Benitez (2019), makes precisely a distinction between framework and interaction theories, the first providing structural foundations, while the second apply more specifically to a particular ontology. In this taxonomy, quantum mechanics falls into the group of framework theories, as it provides general principles — such as the Born rule and the use of Hilbert spaces to model systems and states (Benitez, 2019, sect. 5) — that constrain a plurality of interaction theories. A crucial feature of framework theories is that they are purely structural in character, not treating a particular ontology on their own. They shape how the ontology of the world is to be modelled by interaction theories, such as electromagnetism, but they do not postulate such an ontology to start with. These arguments resonate particularly well with the claims I have advanced, as they seem to support the envision of a certain distance between quantum theory and the objects it may eventually model.

The second example is constituted by a position that seems to have amongst its motivations the apparent "ontological emptiness" of quantum theory. The *primitive ontology program* (Allori, 2013), (Allori et al., 2008), (Esfeld and Deckert, 2017) argues for the need of a specific ontology — objects in three-dimensional space — to which the theory needs to refer. As made precise in the mentioned papers, examples of quantum theories that follow this idea are Bohmian Mechanics and the Ghirardi-Rimini-Weber theory. Unifying trait among the two is the postulation of a particular set of objects, point particles in the first case, and a matter density field or a flash ontology in the second, that are described by a modified version of quantum mechanics. What is interesting for the present purposes is the fact that among the motivations for this program there seems to be also a recognition of the unsatisfactory ontological commitment of standard quantum mechanics; the starting point is a criticism of what is known as wave-function realism, as this takes the wave function, a field defined over configuration space, to constitute the basic ontological building block of the natural world.

Despite both these positions have different aims than mine here, they seem to gravitate around the claim I have risen above, namely, that the connection between the structures described in quantum theory and the physical world requires further assumptions, in those cases, reference to an interaction theory or the postulation of a primitive ontology for quantum mechanics respectively, which exceed the domain of quantum theory. But if at least the second claim I advanced is part of the conceptual kernel of other philosophical arguments, where does my argument intend to lead? In the previous pages I have written that there is a certain *distance* between the structures of the theory (e.g. entanglement) and natural phenomena (e.g. atom decay, parametric down-conversion). If this distance can be accounted for in terms of the eminently structural character of quantum theory, or of its lack of a specific ontology, this is because the theory is valid in *general*, and not only within a delimited domain. For instance, unless a counterexample is found that falsifies the theory's predictions, entanglement can be considered as a general feature of the natural world, that holds at possibly any* scale, and for any object that can be suitably represented by the mentioned mathematical structures — the star* refers to the limitations of quantum mechanics below the Planck scale.

Let us then circle back to the constitutive matter of this chapter, namely, Bell tests. The exploration of the first and second generation of experiments (sections 4.3.2 and 4.3.3), has shown that a common feature of the experimental results is a strong agreement with the predictions of quantum mechanics. The Bell inequalities (CHSH or Eberhard) yield a numerical bound that needs to be violated in order to rule out the conjunction of the assumptions of Bell's theorem. But quantum theory allows to derive another numerical value that describes precisely the *quantum* bound for the correlations that are effectively obtained in nature.[19] Well, if the very first experiments were already able to achieve the violation of the inequalities, further experiments have matched with near exactness the predictions of the theory — an astonishing agreement between predictions and data is visible already in Alain Aspect's experiment (Aspect et al., 1982, fig. 4 p. 1806). And the exactness of the quantum predictions is not merely a feature of entanglement, but of quantum mechanics in general. This means that the theory is also *highly precise, whenever it finds suitable empirical application*.

By unifying the two horns of the reflection carried out so far, one can see that the theory has two co-existing sides. On the one hand, it is correct to the extent that its predictions are confirmed by the experimental data with near exactness. On the other, it is general and holds universally*, to the extent that it is not tied to the description of any particular thing. Both sides have emerged in the exploration of the experimental facet of the case study of this thesis, i.e. Bell's theorem. In the face of this duality, both sides of which are well defined, two points are particularly interesting. The first is that quantum mechanics works successfully as a physical theory in virtue of the interplay of these two sides. The fact that it applies to such a broad range of phenomena is due both to its generality, and to the fact that its predictions have been empirically confirmed with an almost unparalleled level of precision: the theory is characterized *both* by an incredible unificatory power, and by a high level of empirical adequacy at the same time. The second is that, whereas these two features are well defined, the joint that unifies them, the bridge that makes this theory so powerful for the understanding of the world does not seem to be so readily definable. How does one pass from a general and abstract description to the prediction for example that these photons will display correlations in their polarizations? Looking at the experiments presented in the previous pages provides

[19] Importantly, the quantum bound has been proven to be *unique* by means of a purely mathematical proof by Boris Tsirel'son (1980).

all the confirmations for the correctness of the predictions, but not an answer to this particular question. Carving at this particular joint will be among the objects of the next chapter, where this discussion will be re-connected with the problem of observation. As argued in chapter 2 of this thesis, this latter is in fact a scientific practice that requires the connection between the theory and the world. For this reason, it constitutes an ideal ground for the discussion of the connection between quantum theory and the world.

Part III

Analysis

Chapter 5

The teachings of observation

5.1 Introduction

The last two chapters have explored Bell's theorem as the thesis' case study, both on the theoretical and experimental sides. In chapter 3 we highlighted that the assumptions of the theorem are various and have different nature. The violation of Bell's inequalities by quantum mechanics falsifies the conjunction of all the assumptions, but it is not sufficient, on its own, to single out which premises are rejected. We have seen that different interpretations of quantum mechanics and different quantum theories provide a specific interpretation of the theorem. Some of these derive from theories that have non-testable features. We have shown that there are two elements that can be exploited to formulate a testable prediction, namely, the reproduction of entanglement correlations, and distancing subsystems at space-like separation, which have been used to design experiments that have definitively proven that quantum mechanics violates the inequalities. Nevertheless, the confirmation of these testable predictions is able to corroborate only some aspects within the broader extent of Bell's theorem's significance.

What the study of the Bell tests highlighted in ch. 4 is that there seems to be a measure of distance between the empirical phenomena studied as evidence for the violation of Bell's inequalities on one hand, and the theoretical account of these phenomena on the other. More precisely, it was remarked that there is a distinction between the technical notion of entanglement as a non-separable state and the particular phenomena that incarnate its features in experiment. It was noticed that this separation is not a symptom of vagueness of the theory, as the data matched the quantum predictions with almost exactness. It was also stressed that if there is a gap between the general predictions of the theory and the particular experiments confirming the latter, this gap is also bridged. The open question remained as to *how* this bridge is constructed.

The present chapter constitutes the occasion to take a distance from the case study just analysed, and to address the poignant questions of this thesis. In the following sections I will take stock of what has been explored so far, by connecting some of these points to the problem of observation. This will allow me to reconnect in particular with the considerations made in chapter 2, and to put forth arguments that address the core question of this research on the relation between theory and the world. In this process, two points in particular will emerge: first, in section 5.3.3 I will expand on the metaphysical considerations advanced at the end of ch. 4. Second, in sect. 5.4.3 I will advance a point on the problem of the underdetermination of the quantum theories and interpretations. Having assessed that the empirical data do not rule out any of the interpretations, I argue that positions in favour or against some of those are based on normative grounds (sect. 5.4.4).

This chapter is organized as follows. In section 5.2 I individuate the "minimal observational" units of a Bell test, and assess whether they represent cases of direct observability or not. Shapere's definition of observation is recalled, and I enter in the matter of what the main quantum interpretations reply to the observation question in this case. Section 5.3 moves to the consideration of the two testable elements in a Bell test, correlations and space-like separation. First I individuate the important characteristics of these two elements, and then ask the observation question also at this more general level. It is at the end of this section that I expand on the argument presented at the end of ch. 4. Section 5.4 opens a reflection on Bell's theorem in the light of the path traced in the previous pages. Namely, it starts with the question of whether the experimental evidence leads to a particular interpretation of the theorem. The main theories and interpretations are considered, and the analysis closes with the second argument mentioned above, on a solution of underdetermination by evidence based on normative grounds.

5.2 On observation

If one asks for a general definition of observation by taking stock of the procedures carried out in science in general, and that captures the *essence* of what it is to observe something scientifically, Dudley Shapere's definition is a precise, clear and sufficiently general answer — reason why we endorsed it for the work in this thesis. It does not target any object of observation in particular, nor any specific technique one may implement to collect the observational data. By stripping the definition down to the minimum, Shapere individuates a *process* that qualifies as observational by respecting two conditions: the information about the object to be observed can be received by an appropriate receptor, and such information can be transmitted directly, which is, without interference, from the object to the receptor. It was pointed out that this definition has the advantage of adhering particularly well to the actual research practices, which rely upon complex techniques for the collection of evidence for empirical testing of hypotheses. It was also recalled that this definition allows to free the concept of observation from the chains of human perception, consequently broadening the spectrum of the observable, in particular on the side of the smaller scales of magnitude. The fact that the definition provides a criterion for direct observations is relevant, as, whereas there is a precise way to define what counts as directly observing something, the indirectness of observation seems to come in a spectrum of degrees. Of course, since Shapere's observation is just a type of interaction, a lot depends on how physical interactions are characterized in our scientific theories. We recall that each single instance of observation is further specified by a theory of the source, a theory of the transmission and a theory of the receiver that allow to classify an experiment either as one that achieves direct observation of something or not. Well, one important aspect of the experiments we are considering is that physical interactions are constrained, within relativity theory, to take place within spatio-temporal limits as shaped by the constant of the speed of light — more on this below, in sect. 5.3.2. Moreover, to this day there is no empirical record that we may be able to exploit hypothetical non-local effects to interact with distant objects, as stated by the no-signalling principle. For these reasons, it was tempting to further constrain Shapere's definition of observation by adding that the information should be obtained as a result of a local interaction, as all interactions are local in relativity theory, our best theory of space-time. However, this turn out as problematic for theories that do not endorse this principle, and for which interactions could be non-local, such as Bohmian Mechanics. Therefore, the definition is more general in its current formulation, and it will be applied as such.

In the light of the framework elaborated in chapters 1 and 2, the question to ask now thus is: *how does this definition of observation fare with the experiments considered in chapter 4*? Let us recall that the scope of the experiments was, in a sentence, to rule out local realistic explanations of the entanglement correlations, namely, to exclude explanations of the data in terms of local hidden variables. Fulfilling this objective has required the satisfaction of several experimental steps, which we will consider now in order to individuate the *observational core* of the experiments.

Let us thus proceed by breaking down the general account of the experiments in its smaller constituents. First, we have acknowledged that quantum mechanics in principle can apply to any system, as its validity is not constrained to a particular ontology. The phenomena that display these features with clarity, however, famously involve entities at the smallest scales of magnitude, such as the objects of the standard model. From an experimental point of view then, to design tests directly on these entities has the advantage of producing clear evidence of these properties. We have already seen that a particular chain of historical events has led to the use of photons for the controlled reproduction of entanglement in the laboratory, independently of all the types of systems that have been used in the subsequent decades for performing Bell tests. The loophole-free tests (section 4.3.3 ch. 4) have exploited photons and electrons for the violation of the inequalities they tested. To rule out local realistic theories, the experiments have also focussed on positioning the two measuring stations at space-like separation.

As it was acknowledged both at a theoretical level and by looking at concrete tests, there are two main empirical ingredients to set up a Bell test: the reproduction of phenomena that are known to lead to entanglement among measurable systems, and the possibility to preserve this entanglement for testing at space-like separation. Despite important differences in other respects, the three loophole-free experiments have essentially the same schematic structure: a mechanism is designed to produce pairs of entangled systems, which are set up to reach respectively the locations of two mutually independent and largely separated detectors. At these locations, each subsystem in the pair interacts with a detector that records the value of a particular quantity, i.e. photon polarization (Giustina et al., 2015a), (Shalm et al., 2015a), and electrons' spin value (Hensen et al., 2015). The overall analysis of the recorded values provides empirical evidence of the correlations encoded in the mathematical form of the entangled state. How the correlations are empirically detected is discussed in section 5.3. What is of interest in this section is really to try and carve at the joint of what is the most "fundamental" observational block of these experiments, namely, we can walk the additional step of considering what happens at each measurement station independently.

At this point then, the situation we have in front of us is that of a system, a photon or an electron, of which a detector is going to measure a particular property. Without loss of generality, we can focus on the photonic experiments: each detector records the polarization value of one photon at a time, for a great amount of times within a time interval. Each photon in the entangled pair of these experiments is also in a superposition of two possible polarization values, one of which is recorded in each measurement. Thus, a series of measurements at one location will display an alternation of these two values. Now, what is the observational significance of this series of measurements which record a specific piece of information (polarization value) about the entity (the photons)? The important condition Shapere spells out for describing a *direct* observation of an entity is the absence of interference in the transmission of the information from the entity to the receiver. The key element then is precisely what qualifies as the relevant information here, which can be specified further: we are interested in the value of the polarization of the individual photons, as these will then allow to understand how the pairs in the ensemble are correlated.

At each measurement station then, are we observing the polarization of the photons? We can clearly see that this question cannot be answered seamlessly on the spot. Soposed, it is incomplete. In the following, I argue that the same set of data is explained differently by different quantum interpretations. Although there is a commonsensical way to reply to the question "what is observed?" in terms of recorded values, there is also another sense that directly draws from the theories' explanations of the events, and these two are tightly connected. To attempt to reply to the observation question just asked, below I explore briefly how the main interpretations of quantum mechanics answer the question of what we are observing in these individual polarization measurements: while they all agree that the photon's polarization is not directly observed, the explanation they adduce as to what happens in the experiments, i.e. how the data is produced, leads to the production of arguments that differ, in hindsight, on what the data reports. After that, I will consider the record of the correlations, and the notion of space-like separation, both in connection with observation.

5.2.1 Tell me what you believe, and I'll tell you what you observe

Zooming in on the measurement event at a single measurement station provides a well-defined case to consider what it really means that what is observed, what is the relevant information to be retrieved from the object, and what counts as a detector *are all elements that depend, or, are even determined by, the theory considered*. In particular, the case at hand stresses this fact with respect to the first point: what is observed is described differently depending on the interpretations of quantum mechanics or quantum theories one is considering, despite each interpretation agrees on a reply to our *observation question*, of whether the photon's polarization is directly observed at each measurement station.

Von Neumann collapse postulate. The bare quantum formalism contains a clear specification of the kinematics and dynamics, plus an algorithm following which it is possible to make fully empirically adequate predictions on measurement outcomes. The simplest reply that has been provided to our question has been to postulate a physical process accounting for the interruption of the unitary evolution of the state (Von Neumann, 1955, p. 214):

> "...if the system is initially found in a state in which the values of [quantity] $\mathfrak{R}$ cannot be predicted with certainty, then this state is transformed by a measurement M of $\mathfrak{R}$ ... into another state: namely, into one in which the value of $\mathfrak{R}$ is uniquely determined."

This process turns the superposition into one of the two possible eigenstates. To the question "what have we observed?" one could thus reply that the recorded value arises in consequence of the collapse of the original state, induced by the interaction with the measuring apparatus. According to Shapere's definition then, has the photon's polarization been directly observed? As his main requirement of directness is spelled out in terms of a transfer of information without further interference between source and detector, the answer is negative. The measurement interaction perturbs the system in such a measure as to radically change its original state, causing a loss of the information encoded in the original state of the photon. If what is directly observed is the result of this perturbation, the original state is only indirectly observed at best, via an a posteriori reconstruction from an ensemble of data.

Dynamical collapse theories. There are also more recent theories that have maintained the idea of a collapse as a physical process, independently of the act of someone measuring the systems. Both Ghirardi-Rimini-Weber theory (Ghirardi et al., 1986; Bassi and Ghirardi, 2003; Ghirardi and Bassi, 2020) and Diósi-Penrose model (Diósi, 1987, 1989; Pearle, 1990; Penrose, 1996) envisage the collapse of the quantum state as a natural physical phenomenon. GRW postulates a recurrent interruption of the unitary evolution of the state by modifying the original Schrödinger's equation with two parameters indicating the rate at which the collapse occurs. The Diósi-Penrose model makes the analogous case, but by designating gravitational effects as responsible for the non-survival of states such as quantum superpositions. Once again, the details and the viability of these theories is not the focus of this thesis, as we are pressed by the comparison with our definition of observation: the process which brings about the definite values effectively registered by detectors occurs independently of someone measuring. Thus, according to these theories, we directly observe the effects of the spontaneous collapse of the quantum states. For the case at hand however, the state is isolated before it reaches the measuring station, and it is the interaction with the latter that prompts the whole system to collapse into one of the possible eigenstates. Despite the mechanism is different from the previous case, the result is analogous: in both cases the measured system passes from a superposition to an eigenstate during the measurement event. The reply that these theories give to the observation question is thus the same as in the previous case, although for different reasons.

Hidden variables: Bohmian mechanics. However, evidently collapse models are not the only accounts to solve the measurement problem, but the interpretations of quantum mechanics and quantum theories, as mentioned several times in the thesis, are many and diverse. I am following the classification offered by Myrvold (2018, sect. 4.2), which is based on how they understand the quantum state. Within this taxonomy, the collapse models just recalled constitute the family of theories that modify the original dynamics of the theory in order to allow for the occurrence of a physical interruption of the state's uitary evolution, being this dependent on the act of measurement itself, or recurring spontaneously. There are then two more manners in which one may proceed: either by holding the canonical quantum state as an incomplete representation, or by explaining in which sense the formalism accounts for physical reality as it is, and *without* the collapse postulate.

Let us start from *realist* theories of the first type, which proceed by complementing the original theory with additional variables. Main representatives of this family are the modal interpretations and Bohmian mechanics. Both types of theory dispense with the collapse postulate, by accounting for how it is the case that only a particular range of values among those specified by the quantum state are detected by measurements via reference to other structures. The most well-known hidden variables theory is Bohmian mechanics, a quantum theory the fundamental tenets of which consist in the postulation of a discrete particle ontology arranged in an exact configuration and of the construction of an additional dynamical equation, namely, the guiding equation, which describes the particles' motion (Bohm, 1952; Holland, 1993; Goldstein, 2017). The evolution in time of a configuration of particles is represented in Bohmian mechanics by a single wave-function that describes the motion of all the particles in the configuration. Within this theory the only observable that has a well-defined value at all times is position, and it is claimed that all the other observables can be reduced to position.[1] The example we are considering displays also the additional complication that the status of photons is not clear in the theory, as the latter seems compatible rather with the modelling of fermions. Nevertheless, if, for the purposes of the observation argument we consider a fermionic system like an electron and, e.g. a spin measurement, the situation would be the following. The system would be described like a point moving along a well-defined trajectory, until when it becomes entangled with the systems composing the measuring apparatus. The latter event would influence the system's trajectory in such a way as to prompt the apparatus to output a precise measurement outcome.

The answer to the observation question is thus problematic: photon polarization in this theory can be accounted for by means of measurement counts, rather than by point-particles "landing" on a particular edge of a detector. But even in the unproblematic example of a spin measurement, all what is observed according to Bohmian mechanics is the system's position. We can see then already how radically a different theory diverges in its response from the previous accounts of the same situation.

A bare theory without collapse: Everettian interpretations. According to the classification above, there is also a family of realist interpretations that represents the physical world through the lenses of the original quantum formalism, without endorsing the collapse postulate, namely, the group of the Everettian interpretations.

Both in the original Everettian formulation of the "relative state" (Everett, 1957) and in the most prominent Everettian interpretation, namely, the Many Worlds Interpretation (DeWitt, 1970; Wallace, 2012; Vaidman, 2018), the quantum state occupies a fundamental ontological role, as the structure of the world is described in function of the structure of the quantum state. For the present purposes, the metaphysics of the many-worlds interpretation does not play an important role, in particular because the structure of the world described in this theory is not empirically accessible. The key idea to understand the relation between the quantum state and the effective detection of particular measurement outcomes is that the state describes different situations all of which are effectively realized. Rather than "collapsing", the alternative descriptions of the world encoded in the quantum state are all actualized in all

[1] For the philosophical arguments and an explanation of how this could be done cf. the already cited Lazarovici et al. (2018). The reason why this choice of a preferred observable with non-contextual values is a *necessity* in a hidden variable theory as Bohmian mechanics is that this choice allows it to withstand the result of Kochen-Specker theorem. Whether the reduction of all observables to a unique non-contextual one is preferable to the reduction to a group of observables can be debated, but this is not the object of the present chapter.

their features. In the case of the superposition we are considering, according to the Everettian interpretations it is not the case that the system was in a superposed state that collapsed at some point, nor that its state was always well-defined, deflating the ontological significance of the superposition. Rather, all the terms involved in the superposition are physically realized. This means that there is no unique measurement that, alone, is able to reveal the complete quantum state of the system. The answer to the observation question of the photon polarization would be that we directly access only *part* of the information encoded by the state.

Non-realist theories: relational quantum mechanics, QBism, and the pragmatist view. Relational Quantum Mechanics (Rovelli, 1996) does not modify the original quantum formalism, but it does not endorse a reification of the quantum state either. The world is not described by branches of the wave function, but as a collection of relational events. The values of physical magnitudes are relational, in the sense that they are determinate always relative to another system. For instance, the polarization of our superposed photon assumes a certain value relative both to the detector and to the observer that records the outcome. The physical significance of quantum mechanics in this view is not to account for the ontology of the world, but rather to provide a consistent manner of *codifying the information* that it is possible to acquire from the interaction among physical systems.

The core tenet of this interpretation is that there is no such a thing as a God's eye perspective on the world, but only a network of facts, accounted for differently by different observers. This is not a realist interpretation *of quantum mechanics*, as it does not consider the quantum state as an ontological entity, but it is not an anti-realist interpretation *tout court*: there is a matter of fact about events in relativistic space-time, which science is able to account for. A supporter of relational quantum mechanics, in the light of the experiment we considered throughout this section, would say that what they have observed is the photon's polarization relative to their own measurement. The latter thus reveals a pre-existing value but only relative to their own measuring apparatus, and nothing more can be said about the photon's polarization "in itself", because this is a meaningless concept in this account of the world. If we consider the situation more generally, we would need to specify that the state of that system is also well-defined relative to other systems in its proximity, and not only to the apparatus that records the outcome. But, for the scenario we are considering, the interaction between photon and detector is the only relevant one, as we are interested in understanding whether the "minimal observational units" of a bipartite Bell test do carry out direct observations of photon's polarizations. The answer to whether we directly observe the polarization value is then affirmative, but only within the limits of the particular network of relations that needs to be taken into account for the description of the experiment itself. As Shapere's definition does not make additional comments on the absoluteness of observation, its application to this type of reasoning needs further specifications besides a simple affirmative or negative answer. This makes the relational position on observation hard to compare to the ones considered so far, as it relies upon additional constraints not equally taken into account by the other interpretations.

If the view just recalled eliminates reference to the world as absolutely and uniquely describable, and accounts for the quantum state as a means to encode information about a system relative to another, according to QBism (Fuchs and Schack, 2014), the significance of the quantum state loses completely its agent-independence. According to relational quantum mechanics the absence of a distinction between observer and observed is a crucial tenet that is given up in its entirety in the QBist view. According to this latter, the significance of the quantum state is to allow rational agents to calibrate their degrees of belief about a physical state of affairs, playing the role of a development of canonical probability theory. Probabilities in this interpretation are understood as completely subjective, as they play the sole role of guiding an agent's beliefs about the world. For this reason, the explanatory power of quantum mechanics so understood is limited to the sole subjective assignment of probabilities to a certain state of affairs that will be experienced by the agent. While the theory explains why the probability of each measurement outcome is of a certain form, it does not account for the single outcome actually obtained. If we ask our observation question, then such a theory sabotages any attempt of asking whether we have observed the polarization of the photon, because the boundaries of what can be described are constituted by what pertains subjective experience of the world. In this sense, questions about anything *outside* these boundaries would not only be left unanswered, but be devoid of meaning to start with. For this reason, we ought to conclude that our question can neither be asked nor answered within the QBist framework.

The last member we are considering in this concise gallery is represented by the pragmatist interpretation of the quantum state (Healey, 2012, 2017b). While also taking the latter not as a representation of physical reality, this view disagrees with the previous on the understanding of probabilities. These are not in fact phrased as subjective chances assigned to personal experiences of an agent, but as objective measures of accuracy with which *situated agents* ascribe degrees of belief to claims about the value of physical quantities. In this view, there is an objective matter of fact about the probabilities assigned by any identically situated agent with respect to a state of affairs. However, these states of affairs are neither *represented* by quantum states as e.g. in the Everettian interpretations, nor *known* via quantum states, as in a ψ-epistemic understanding of the state. Rather, a rational agent finds in the quantum state a *prescription* about what it is rational to believe with respect to those states of affairs. In relation to our question of observation then, this interpretation does not fare much better than the QBist reply. Although there is a matter of fact about what a measurement reveals, there is no underlying explanation with respect to such a revelation, as it is not the scope of quantum mechanics to describe reality: this theory is rather a prescriptive tool for the formation of rational beliefs of its users. Whether we have observed the polarization of the photon according to Shapere's definition is a question that cannot be answered in this interpretation, as there is no story as of what happens to the photon independently of the agent.

Since the last two non-realist interpretations block the arising of the observation question at the origin, they are not informative with respect to the present investigation.

❊ ❊ ❊

This survey of what the main interpretations of quantum mechanics and quantum theories reply to the observation question has shown clearly that the same conundrum can be solved in radically different manners, depending on which additional commitments one endorses in their understanding of the theory. On the one hand, the problem is that the theory alone is incapacitated to answer this question, as even the interpretations that seem to adhere the most to the original formalism need to justify its empirical significance via additional claims. On the other hand, even with these additional claims, the interpretations diverge strongly on what is the answer to the question, some of which are even unable to address it in the first place. This array of mutually conflicting answers is generated by the same original theory, which is held as empirically adequate by all its interpreters. The common trait among all answers seems to be that no account of the events would fit what Shapere defines as direct observation, as according to no interpretation does the measurement event convey the information carried by the system before measurement. Therefore, the data are not an example of direct observation according to any view. This is not a problem embedded in Shapere's definition of observation, but rather of how quantum mechanics represents the world. The fact that there can be real superpositions that evolve unitarily, but that do not correspond to the seemingly non-unitary evolution of our measuring devices is a problem that depends on the theory, not on the account of observation one adopts.

What is this fact a symptom of, if of anything? The rhetorical undertone of this question is the drum roll for a to-be-expected remark, as I take it to be a confirmation of the idea advanced at the end of chapter 4. First, this seems to be an explicit illustration of what it means that quantum mechanics is a theory about no object in particular, as it calls for a conceptual completion to even start to say anything about *any thing*. Second, I suggest that the array of contradictory answers we are presented with when this conceptual completion is attempted with respect to quantum mechanics *itself* may be the problem in the first place. Rather than trying to fill the apparent void left by the incapacity of the bare formalism to account for empirical reality, one may acknowledge this void, and take it as the symptom that the physical significance of quantum mechanics may not be that of accounting for an ontology *directly*, but rather to provide a set of categories for how physical theories should model the world in the first place. More specialized theories built by following the categorical blueprint provided by quantum mechanics would be the adequate ones to model the natural world's ontology directly. This comment will be fleshed out in section 5.3.3.

Let us now proceed to an analysis of the features of the experiments that uncontroversially reveal entanglement and its non-classical properties; namely, correlations and space-like separation.

5.3 On entanglement

If every interpretation provides a different explanation of the same events, and if the measurements at each station are the only places where direct observation could take place, what are we left with, once we consider the data that are significant for entanglement? It would seem that if the grounds are uncertain, they cannot be solid foundations for a structure raising from them. However, this is a misleading premise: although different theories may disagree on what one observes, they all agree on the data sets obtained from these experiments, which constitute the relevant basis for the empirical study of entanglement. Moreover, entanglement is more than a way to describe the correlations empirically recorded, as it is primarily a structure that derives from the abstract mathematics of the theory, which is a substrate common to all the interpretations. We should now continue our analysis of the data by considering the minimal observational units of the previous section *in their* *ensemble*. As we have called them, the two "ingredients" to perform a Bell test are the reproduction of the entanglement correlations among specific physical systems, and the maintenance of these correlations at space-like separation. Let us go through each of them individually.

5.3.1 Correlations

The study of the correlations among entangled systems and their properties is an extended research area in quantum physics (Horodecki et al., 2009, sect. IV). What interests us here is precisely the fact that these correlations cannot be reproduced *classically*, as this is the essence of what Bell's theorem demonstrates. But what does that mean? Crucially, they cannot be reproduced by the mathematics of classical physics, which lacks the structures that allow their arising in the first place. Entanglement in quantum mechanics derives naturally from the fact that the formalism yields *non separable* states. By referring to the common axiomatization of quantum mechanics via Hilbert spaces, I am going to describe how such states depend on the tensor product structure. This constitutes a rather clear example, as it allows to make a direct comparison with classical physics, where states compose by the Cartesian product, which notably does not yield correlations as strong as the quantum ones. The Cartesian product of classical state spaces gives rise only to a subset of the set of states that can be written via the tensor product. What exceeds the limits of this subset cannot be reproduced within the subset.

This comparison can be seen within a minimal formal clarification. In the Hamiltonian formalism of classical physics — here I am borrowing from Mallesh et al. (2012), the state space is represented by the real-valued phase space Γ of dimension $2n$, with the canonical coordinate system constituted by positions $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ and momenta $\mathbf{p} = (p_1, p_2, \ldots, p_n)$. Importantly, the state space is a set with a specific structure. A pure state is a point on the phase space: $(\mathbf{q}_0, \mathbf{p}_0) \in \Gamma$. Let us consider how the state of a composite system is built in this formalism. Let S_1, S_2 denote the phase spaces of two physical systems, s_1, s_2. For the case of pure states, a multipartite state is given by the composition of the elements of each sate space, following the rules of set theory. The structure that yields this composition is the Cartesian product, for which the general definition is the following (Cf. Suppes (1972, ch. 2, sect. 2.8) and Goldrei (2017, ch. 4, sect. 4.4)):

Let us consider sets X and Y, containing respectively elements x_i and y_j. The Cartesian product ($\times$) of X and Y is a new set, the elements of which come from the union set of X and Y, and they are *pairs* formed by an element of X and one of Y

respectively:

$$X \times Y := \{\langle x_i, y_j \rangle : x_i \in X \wedge y_j \in Y\}. \qquad \text{(Cartesian product)}$$

A pure composite state in classical physics is therefore specified in the phase space formed by the Cartesian product $S_1 \times S_2$, and it has the form of a pair: (s_1, s_2).[2]

For the quantum case, the starting point is entirely different, as the minimal units to represent physical states are vectors. The state spaces of quantum mechanics are vector spaces over a complex-valued field, and they have a *linear* structure. Let us start from the mathematical definition of the tensor product. The vector space K^n, where K is a field and $n \in \mathbb{N}$, as a set, is given by $K^n = \{(v_1, \dots, v_n) \mid v_i \in K\}$. The two operations possible on this set are addition:

$$\begin{aligned} &K^n \times K^n \to K^n \\ &(v, w) \to v + w = (v_1 + w_1, \dots, v_n + w_n), \end{aligned} \tag{5.1}$$

where both v and w denote vectors, and scalar multiplication:

$$\begin{aligned} &K \times K^n \to K^n \\ &(\lambda, v) \to \lambda \cdot v = (\lambda v_1, \dots, \lambda v_n), \end{aligned} \tag{5.2}$$

where λ indicates a scalar. The linear structure proper of vector spaces allows a greater number of possible combinations among the elements of different spaces than those yielded by the Cartesian product. For this reason, the latter is insufficient to capture these combinations, and it is substituted by the *tensor product* ($\otimes$), which is defined as follows. Given vector spaces V and W:

$$V \otimes W := \left\{ \sum_i \lambda_i (v_i, w_i) \mid \lambda_i \in K, (v_i, w_i) \in V \times W \right\}. \qquad \text{(Tensor product)}$$

A quantum composite state is written in analogous manner as for the classical case, i.e. via the composition of the state spaces of the two systems by means of the tensor product. The general form of a quantum pure state is built as follows (Cf. Le Bellac (2011, ch. 6)). Let us start from two systems living on Hilbert spaces $\mathcal{H}_1^N$ and $\mathcal{H}_2^M$ with dimensions N and M respectively. The states of the two systems are

[2] The Cartesian product composition of states works also for the general case of mixed states, which is here omitted for reasons of simplicity. We will focus on pure states only also when defining composite quantum states below.

represented by vectors $|\varphi\rangle \in \mathcal{H}_1^N$ and $|\chi\rangle \in \mathcal{H}_2^M$, and their composite state is also a vector, specified over the space $\mathcal{H}_1^N \otimes \mathcal{H}_2^M$. Vectors $|\varphi\rangle$ and $|\chi\rangle$ can be decomposed into the vectors belonging to the orthonormal basis of their Hilbert spaces, namely: $|\varphi\rangle = \sum_{n=1}^{N} c_n|n\rangle$, $|\chi\rangle = \sum_{m=1}^{M} d_m|m\rangle$. Extending this decomposition to $\mathcal{H}_1^N \otimes \mathcal{H}_2^M$, the elements of the composite state space are expressed as pairs of basis vectors belonging respectively to each space: $|n \otimes m\rangle$. The general form of a vector on state space $\mathcal{H}_1^N \otimes \mathcal{H}_2^M$ is:

$$|\Psi\rangle_{1,2} = \sum_{n,m} b_{nm}|n\rangle \otimes |m\rangle, \tag{5.3}$$

where, generally, the coefficient b_{nm} *cannot* be factorized in terms of the coefficients of the individual vectors, i.e. c_n and d_m, which is precisely the defining condition of quantum entanglement. This means that this state cannot be written as the mere tensor product of two vectors belonging respectively to each of the two state spaces:

$$|\Psi\rangle_{1,2} \neq \left(|\varphi \otimes \chi\rangle = \sum_{n,m} c_n d_m |n\rangle \otimes |m\rangle \right). \tag{5.4}$$

There is a direct comparison between the *correlations* yielded by the separable state in the parentheses of eq. 5.4 and the classical composite states, as the two state vectors can be written in the form of a pair $((|\varphi\rangle, |\chi\rangle) \in \mathcal{H}_1^N \otimes \mathcal{H}_2^M)$, and they *can* yield the same correlations.[3] However, as already mentioned in the beginning of this section, this type of states are only a subset of the states normally living on the composite Hilbert space. This analogy fails in fact for entanglement: the correlations yielded by a state such as $|\Psi\rangle_{1,2}$ are opaque to the classical formalism alone, where they could be mimicked only by carefully crafting a scenario suitable to reproduce correlations that look like those yielded by entangled states, but within classical field theoretic constraints. The most fundamental manner in which the entanglement correlations differ from the classical case is therefore rooted in the very mathematical ingredients with which physical states are specified.

But this is not necessarily the only perspective through which this difference becomes visible: the non-classicality of the correlations manifests also on a logical level. Another face of the particularity of the entanglement correlations is constituted by the fact that classical reasoning encounters paradoxes when it tries to account for the entanglement correlations. The most important result witnessing this fact is represented by the GHZ theorem (Greenberger et al., 1989). The authors start

[3] Nevertheless, separable quantum states are not classical states; for instance, quantum discord lead to correlations among quantum separable states that cannot be reproduced classically.

from the fact that Bell's theorem rules out a local hidden variables account for the entanglement correlations for the general case. They apply the EPR argument to the CHSH scenario, where measurement on one side determines the outcome on the other side with certainty. For each of these measurements it is possible to build a classical model that accounts for what happens. The doubt arises then on whether the general case could always be rephrased as classical in particular scenarios as these. By considering a slightly more complex entangled system, the theorem proves exactly that this is not the case.[4] In trying to represent the constellations of outcomes at each instance of measurement by applying the categories of classical reasoning, one encounters paradoxical dead ends. For example, Maudlin (2011, appendix A, p. 25) contains a simple representation of the inadequacy of classical reasoning for encompassing quantum correlations, in an analogous way as previously done by David Mermin (Mermin, 1993).

With respect to these formal results, it is interesting to notice that there is also research going in the other direction: on one extreme, it has been proven the impossibility of building a classical model for the quantum correlations. On the other side though, questions have been asked with respect to the quantum model independently of its relation with classical physics. Starting for the bi-partite case considered by CHSH, where two subsystems can be measured via two possible choices of settings, and they can give only two possible values as outputs, it is well-known that the quantum value for the violation of the inequality is *exactly* of $2\sqrt{2}$ — see derivation in section 3.1.3 in chapter 3. An important result by Boris Tsirel'son (Tsirel'son, 1980) has shown that this is also the maximal amount of correlation that can be achieved by an entangled state under the conditions specified, although in the CHSH scenario one could tweak the inequality and reach, in principle, the maximal value of 4. It has been shown by Sandu Popescu and Daniel Rohrlich (Popescu and Rohrlich, 1998) that it is possible to design an ad hoc state able to reach the latter amount of correlations, and without violating the no-signalling constraint.

This result is not relevant within quantum mechanics *per se*, but precisely for a series of questions regarding quantum mechanics as a physical theory, which ask whether it is possible to find a more general theory able to reproduce correlations that are stronger than the quantum ones, and that would include those in a subset. In other words, as quantum mechanics yields correlations that are stronger than the classical ones, there could be a more general theory related in the same way to the quantum. This is the project of general probabilistic theories (Janotta and Hinrichsen, 2014). The

[4] Soon after, Lucien Hardy proved that the same result can be obtained also for a bipartite entangled system: Hardy (1992, 1993).

answers found to these questions are important: quantum mechanics emerges among other possible general probabilistic theories for the minimality of its assumptions, for the generality of its predictions, and for its empirical adequacy. These results are also interestingly matched on the empirical side: if the first Bell tests mainly confirmed the violation of the classical predictions, the later, more sophisticated experiments have increasingly approached the quantum prediction, without violating it. These results, both formal and empirical, seem to show that not only the classical framework is violated, but the quantum is singled out among alternatives.

From these different angles from which one can look at the correlations, they clearly display their "quantumness" in contrast with the type of correlations allowed by classical physics. Mathematically, their form is neat and precise, and it has emerged several times in the previous pages that the statistical data on repeated measurements on entangled systems constitute evidence for their accuracy. How do they fare with observability then?

Let us consider again the type of experiment of the loophole-free Bell tests. After the entanglement source, the experimental design and the measuring stations have been set up, the experiment runs for a time that allows to collect a sufficiently great number of data for an adequate statistical analysis. It is at this point that one is put in front of the empirical manifestation of the entanglement correlations. The outcomes on each side of the experiment are in fact correlated as described by the prepared state. The question of observability at the level of the correlations can be asked with respect to the statistical data: can we conclude that by analysing the statistics, we are directly observing entanglement correlations? The answer is two-fold, and, in brief, it is negative.

First of all, the correlations cannot be inferred *unless* one has access to the data in their completeness. Let us imagine to ask Alice and Bob, sitting at each measuring station and not communicating with each other, to keep track of their measurement separately, and report independently what they have detected. If the source was emitting entangled systems in the singlet state, they would say that they have measured a superposition. To the question of whether they have measured part of an entangled system, they would reply that it is possible, but they do not have enough elements to know whether that is the case or not. Unless one has access to measurements performed on all the parts of an entangled system, they cannot make any claim on it being an entangled state.

This is however not a definitive no-go for the observability question, as it is also important to specify that the availability of appropriate statistical data and knowledge of quantum mechanics in some cases is *everything* one needs to know in order to recognize an entangled state, regardless thus of which systems have been measured, which experimental settings one has chosen, or how the experiment was designed. The term "self-testing" (Mayers and Yao, 1998; Scarani, 2019, ch. 7), refers to entangled states that are recognizable *with certainty* via knowledge of the statistical data ensemble and of quantum mechanics alone. For example, if by looking at a given statistics of data output from a black box, one can conclude that the only possible state that could give rise to them is a GHZ tri-partite state, then this state has been self-tested. The latter is in fact among the states suitable for self-testing.

It is the second part of the two-folded answer that posits a bigger obstacle to observability. Namely, one cannot interact with both horns of the experiment *at the same time*. We have seen how all the loophole-free Bell tests have prepared the experiment in such a way as to definitely close the locality loophole (see sect. 4.2.1). The detectors at each location record a particular measurement outcome independently, and it is not possible to witness the "appearance" of both outcomes in the moment in which the recording takes place, at the moment in which each subsystem reaches its detector. As we are going to see in the section below, we can see the setting in its completeness only by being located at a space-time location that encompasses both detectors either in its future or in its past light-cone — which is, either before, or after the detection event. For this reason, entanglement in terms of a correlation among a number of subsystems is not directly observable, as non-local interactions are not achievable in practice.

Let us thus explore further the notion of space-like separation, which is the second fundamental ingredient for the realization of Bell tests.

5.3.2 Space-like separation

In section 4.2.1 of the previous chapter we have explored three types of locality loopholes, namely, communication loophole, memory loophole, and freedom of choice. In this section we are interested especially in the first two, as they have been definitely closed in the loophole-free experiment, whereas we have seen that the closure of the third loophole can be tackled experimentally reasonably enough, but its closure is open to conceptual debates. Let us recall briefly what they amount to: the communication loophole refers to the possibility that information may travel from one measuring station to the other, allowing in this manner a coordination of the outcomes of the measurements on the two subsystems. If this were the case, there would be no difference between the behaviour of the photons and the behaviour of two people coordinating their behaviours by communicating with one another on the phone. The memory loophole refers instead to the fact that each subsystem may be separately influenced in a deterministic manner by the choice of measurement on each side, for instance if this choice had been set before the pair was released from the source. In this manner, along the lines of the previous analogy, the two people would not be on the phone, but they would know already the precise input they would be requested to react to each time. In both cases, it is possible to model the behaviour of the physical subsystems by adding extra variables, not contained in the quantum description, which support a classical account for the outcomes actually obtained upon measurement.

But, as we have already seen, this option has been experimentally ruled out. Although the bare quantum formalism alone does not refer to any particular theory of space-time, in the very instant in which we want to perform an experiment we need to come to terms with the fact that we, as well as the objects we experience and any empirical test we perform, are located *somewhere* in space and in time. Currently, the scientific manner in which we conceive space and time is prescribed by the theory of relativity, born twenty years before quantum mechanics, and currently also not falsified. Consequently, the form of locality which plays a crucial role within the Bell tests is the relativistic, precisely formalized notion that we are going to recall below on the basis of the first chapters of Schutz (2009) and Carroll (2004) respectively. It is however important to stress that locality is not a relativistic notion *per se*, but a deeper, common assumption in the manner in which we think of the world physically. A philosophical paper by Einstein (1948, sect. II) conveys this idea particularly sharply. For what concerns the current analysis, what we need to focus on from the theory of relativity are two facts: on the one hand it provides a unified description of space and time, which are uniformly part of the same structure. On the other hand, it postulates a fundamental constant of nature, the speed of light in the void, which remains invariant and prescribes a limit, amongst other things, to the manner in which physical interactions can take place. We are going to consider especially the second element, and to refer to the space-time diagrams for a theoretical representation of space-time, without however entering in the matter of the precise geometry of space-time and of the effect of gravity.

A point in relativistic space-time has four coordinates (x, y, z, t), and it is called an *event*. There are two fundamental postulates from which special relativity can be derived. The *principle of relativity* is the Galilean postulate that the results of an experiment do not depend on the observer's speed relative to other observers external to the experiment, because no experiment can measure the absolute velocity of an observer; the second institutes a new absolute, a constant of nature, the "northern star" in a completely relativistic universe. It is the principle of the *universality of the speed of light*:

> The speed of light relative to any unaccelerated observer is $c = 3\times10^8 \text{ms}^{-1}$, regardless of the motion of the light's source relative to the observer. ... two different unaccelerated observers measuring the speed of the *same photon* will each find it to be moving at $3\times10^8 \text{ms}^{-1}$ relative to themselves, regardless of their state of motion relative to each other. (Schutz, 2009, p. 1).

In light of these postulates, the distances among separate events are not thought of any more in terms of portions of an absolute background space, but they always involve the temporal dimension as well, and they have meaning only relative to two or more events. For this reason, a basic relativistic element is the notion of space-time interval $(\Delta s)^2$:

$$(\Delta s)^2 = -(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 . \tag{5.5}$$

The notion of space-like separation is defined very simply: two events are space-like separated when the space-time interval between the two is positive ($(\Delta s)^2 > 0$). It is easy to see that for this to be the case it is necessary that the sum of the three spatial intervals needs to exceed the temporal interval parametrized by the constant of c. The manner in which we have visualised this notion so far is via reference to the light-cone structure, defined as "A set of points that are all connected to a single event by straight lines moving at the speed of light" (Carroll, 2004, p. 9). As we already knew then, two events are space-like separated if they lie outside each other's light-cone.

We have already seen both in Bell's formulation of the theorem (cf. section 3.1.1) and in the exploration of the theorem's premises (section 3.2.1) that, with the assumption of special relativity as a theory of space-time, the notions of locality and of local causality are grounded on the notion of space-like separation, in the sense that local interactions or local causal chains are those which do not violate the limit of the speed of light, and thus that do not directly connect space-like separated events.

There is a measure of disagreement with respect to whether entanglement unveils a form of non-locality in the world, and about exactly *what* would be subject to such a non-local behaviour. But one element generally agreed upon is the fact that it is not possible to controllably exchange information above the speed of light. The principle of *no-signalling* in relation to entanglement is understood as the independence between the statistics of the outcomes on one side and the choices of setting on the other side — for example, cf. Scarani (2019, ch. 2, sect. 2.2). This entails that, although the correlations among entangled systems cannot be accounted for by means of local hidden variables, this result cannot be taken as a straightforward violation of relativity.

These clarifications serve to address the experimental closure of the two locality loopholes upon which we are focusing. The crucial point here is that space-like separation is experimentally realizable, and it is in this sense the fundamental second ingredient for the realization of a Bell test. Since this is the part of the experiment that empirically corroborates the *quantumness* of the entanglement correlations, as we already claimed it is also with respect to this point that we ought to address the observation question of entanglement. We have already seen that asking whether we observe the correlations does not have a clearly negative answer; it is time now to clarify why we have claimed that the space-like separation does indeed put a boundary to what is observable.

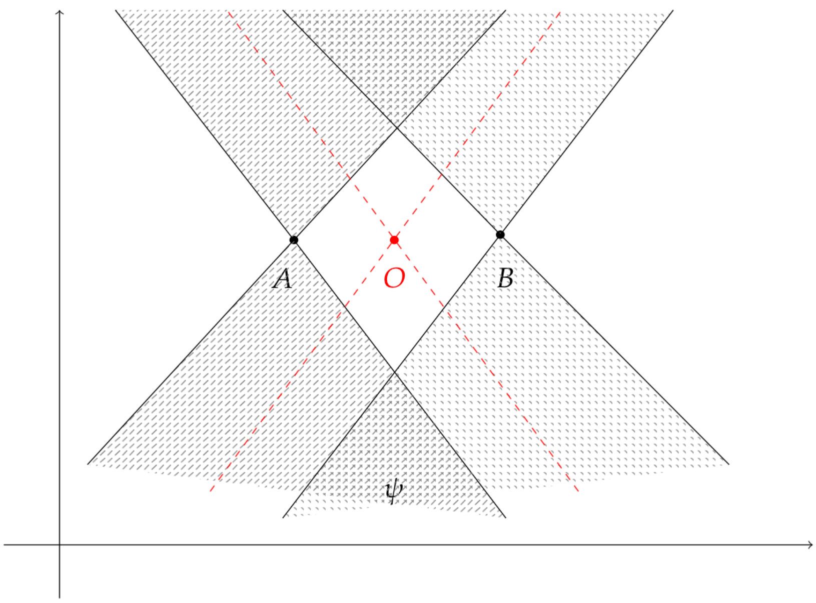

Figure 5.1: This two-dimensional diagram represents two spatially separated measurement events and an observation event located on the same hypersurface. The vertical and horizontal axes stand for the temporal and spatial dimensions respectively. "A" and "B" label the measurement events — indicated by the black dots, from which depart each event's past and future light-cones. At the intersection of the measurement events' past light-cones lies the source of entangled bi-partite systems "ψ", while the red dot labelled by "O" indicates an observation event outside either measurement event's light-cones.

We concluded the previous section with the remark that a unique observer can have access to both measurement events only if they are located at the intersection of the latter's past or future light-cones. Figure 5.1 provides a basic visual representation of the fact that, even if the observation event is equidistant from both measurement events, the observer wouldn't be able to observe the measurement events *at the time in which they occur*. The diagram shows that at that instant the observer does not observe either of them. One will access both measurement records only once the observation event will take place in the common future of the measurement events, dashed with tiny arrows in the image. The point is that the interactions through which the observer sees both measurement outcomes are completely local, so they will not have had a direct experience of the occurrence of the outcomes at space-like separation.

Earlier in this chapter we have remarked that one possibility was to further constrain Shapere's definition of observation to local interactions, but that such a temptation had to be resisted, given the fact that some theories may postulate the possibility of non-local interactions in the world. The fact remains however that, *within a relativistic universe*, it is a predicament of relativity theory that the speed of light posits a limit to the possible physical interactions that can take place in space-time. As observation is a physical interaction, according to this theory it is subject to the same constraints as any other, and this is particularly relevant for the case we are considering. The "surfaces" of the light-cones provide a clear demarcation between directly observable and unobservable.

❊ ❊ ❊

The current section has explored the observation question asked in section 5.2, but at the level of correlations and space-like separation, i.e. the two testable elements which allow to prove the non-classicality of the entanglement correlations, and the correctness of Bell's theorem. In particular, first we have asked whether we observe the entanglement correlations, focusing on these latter exclusively: after a brief overview of their crucial properties, we have remarked that the correlations can be accessed empirically only via the statistics on a large ensemble of data points. On the one hand, we made the example that if one receives the data only of one part of a bi-partite entangled system, then they would have no elements to decide whether they are measuring part of an entangled system rather than a single system in a superposed state. On the other, we acknowledged that the statistical data set is a very powerful tool for the recognition of entanglement, with the limit case of a self-testing state, where knowledge of quantum mechanics and an adequate data-set are completely sufficient not only for noticing entanglement, but for singling out the precise state of the system. For this reason, the question of whether we are directly observing the entanglement correlations by considering an appropriate data set tends towards the positive. The hesitation in this lukewarm reply derives in part from a more general problem, not tackled in this chapter, about whether correlations in general are subject to observation. As this question does not seem to find either an obviously positive or obviously negative answer, we have left it intentionally open.

The negative verdict on whether we directly observe the entanglement correlations derives from the second horn of the matter, namely, from the space-like separation at which the measurements on the entangled subsystems are performed. First, we have reported a precise definition of space-like separation, and clarified the relativistic origin of this concept. Second, we have recalled that the speed of light represents a physical limit to interactions in space-time. As observation is a physical interaction, it is also subject to such a limit. We have thus drawn a simple space-time diagram where we have located the observation event at the same time of the occurrence of the measurement events, and at spatial equidistance from them. By highlighting that the former would lie outside both the future and the past light-cones of either measurement event, we have visually represented what it means that it is not possible to interact locally with both measuring stations at the time in which they record the outcomes. The data are accessible only if the "access-to-data" event is located at the intersection of the future light-cones of the measurement events, which means that such an "access" is *local*. For this reason, the verdict on the direct observation of the entanglement correlations is negative. In a Bell test we observe entanglement only indirectly, by inferring its properties and non-classical features in hindsight, from a study of the data. Let us take stock now of the analysis of entanglement made in these pages.

5.3.3 Corroboration by evidence and realism

Alone, the verdict of unobservability just issued does not lead towards any particular direction. Actually, it may come out as puzzling: what about all the remarks about quantum mechanics being one of the most empirically adequate physical theories we have? And what about previous considerations about the experimental confirmation that Bell tests have provided for Bell's theorem? How does this verdict fare with the seemingly opposite, positive considerations about the theory's empirical adequacy? One assumption lingering behind considerations of this sort seems to be that empirical corroboration of a hypothesis about an object or a state of affairs goes hand in hand with the observability of such an object or state of affairs. Thus, to say that entanglement is not directly observable seems to raise some sort of doubt regarding whether we have empirical evidence about entanglement. For the fact that to gather empirical evidence about something is independent of its observability,[5] this impression needs however to be undercut at the origin: as far as the current experimental knowledge goes, there is no doubt that entanglement is a real feature of the natural world.

The argument we proposed at the end of chapter 4 can be summarized as follows:

- The connection between entanglement as a theoretical structure and a particular phenomenon that is correctly represented by such a structure is non-trivial, in the sense that it requires one to make extra assumptions that exceed the domain of quantum theory.
- Since entanglement is a general feature of quantum states, a more general claim can be made that quantum theory per se is not a theory about any particular thing.
- This apparent distance between the theory and the phenomena it accounts for can be seen as a symptom of the fact that the theory applies in general and universally (excluding the level beyond the Planck scale).
- The experiments discussed are a set of examples that the theory is also highly empirically adequate, as its predictions are confirmed with near exactness by the data.
- Points (iii) and (iv) are two sides that co-exist as crucial features of the theory.

  – Arguably, it is because of the interplay of these two features that the theory is such a powerful tool for the understanding of the natural world.

  – The connecting link between the general level of description e.g. of an entangled state and the correlations among these particular polarized photons is not well defined, in terms of what it is, and of how it is established.

I want now to elaborate briefly on the claims just recalled. The manner in which entanglement is formalized in the theory can be phrased in terms of particular non-classical properties of physical phenomena, which can be tested empirically. Tests to

[5] We have already seen in section 1.2.2 of chapter 1 what are the general features that a set of empirical data must have in order to be considered scientific evidence. That overview made clear the fact that observation is only one of the processes that may lead to the collection of evidence, but not the only one. For this reason, I am not going to linger further here on the relation between evidence and observability.

confirm the existence of these properties and to rule out that they could be traced back to a classical account have been successfully performed. From an empirical point of view, we have tested the "reality" of entanglement. How is this to be understood from a metaphysical or ontological perspective? The first meaning to the distance claim is that entanglement as a formal structure does not directly describe any phenomenon or any particular physical structure. As we have seen, there is a variety of physical systems that can be used to perform Bell tests, and, as we haven't explored, there is a variety of experiments also other than Bell tests that can measure entanglement — for an overiew, cf. van Enk et al. (2007) and Gühne and Tóth (2009).

A minimal claim that can be made on the basis of the strength of the experimental evidence of entanglement is that, within the realist account of scientific theories endorsed as an assumption for this thesis, entanglement *yields true propositions* about the natural world. For instance, it is true that a system in an entangled state is subject to correlations that cannot be accounted for classically, it is true that these correlations arise at space-like separation, that entanglement is not broken by spatial distance, and so on. However, it is part of a realist attitude towards scientific theories to believe that the objects they model are *real*. The question then would be whether it makes sense to think of "entanglement" as a real thing in the world, and in which sense this ought to be understood. As we have argued, entanglement is a formal structure that subsumes the behaviour of a variety of systems, and a range of properties of this behaviour.

Hence, the following are viable options: on the one hand, we could take it to be a real structure that can be thought of as a universal, which correctly accounts for a variety of particular instances; I am not going to explore this option, as it seems to be rather far from the approach generally carried out in this thesis, and as it seems to clash with the conceptual framework set out in chapter 1. Moreover, my resistance to the idea of committing to the existence of abstract entities derives from a general resistance to Platonism as a metaphysical position. On the other hand, entanglement could be considered a structural property of the world, but not characterized as an abstract entity. More precisely, the second option would amount to claim that in our current scientific understanding of the world, it is a property of physical objects to be subject to quantum entanglement. The second position has been argued for by authoritative metaphysicians in detail. Representative examples are the notion of entanglement as relational holism put forth by Paul Teller (1986), Michael Esfeld's proposal of a metaphysics of relations not supervenient on the intrinsic properties of their relata (Esfeld, 2004), or the recent paper from David Glick and George Darby

(2018), who argue for a form of entanglement realism.

One further option however is to deflate the ontological significance of entanglement, and to consider it as part of a conceptual set of tools that should be used to think of the world as quantum. This is the view I endorse. This particular position with respect to quantum mechanics plays a marginal role in the overall architecture of this thesis, thus I am not going to explore it in full depth, but it is subject of further work. Nevertheless, the main features of this position are the following. This conception of entanglement would be part of a more general consideration of quantum mechanics as a physical theory, which, rather than postulating or describing an ontology *directly*, would provide the conceptual categories to model physical objects in nature. Its role would be that of a source of the categories we need in order to think of natural objects as quantum. Entanglement being a fundamental and general property of physical systems as formalized within the theory would constitute one of these conceptual tools. The manner in which this view conceives the relation between a theory as quantum mechanics and empirical evidence then is that the systems that are used in experiments, for instance in the Bell tests we have considered, and that have confirmed the predictions yielded by entanglement, are successfully understood and accounted for through the lenses of the conceptual tool kit quantum mechanics provides. Matters concerning a more precise definition of what is the ontology of the fundamental objects in particular would be deferred to more specific theories that model such an ontology and its properties, *on the basis of* the conceptual categories provided by quantum mechanics.

The notion of "category" at the sketched stage of this is idea is of Kantian inspiration, as I imagine this categorical toolbox as something that *enables* us to think matter in contemporary science. However, a more precise account of this idea is in progress. It clearly resonates quite closely with the taxonomy of physical theories into framework and interaction theories mentioned already at the end of ch. 4. Since framework theories put constraints on how the world can be modelled — notably, by more specific interaction theories —, to speak of quantum mechanics in terms of a categorical toolbox instead of as framework theory would seem redundant. Nevertheless, in the recent formulation of Benitez (2019) framework theories too model a particular ontology of structures. For the fact that in my view a categorical toolbox would not do this, I think that my proposal displays important differences with the latter. The possibility that the two views may reduce to one another is not excluded, but to flesh this out more thoroughly is subject of separate work.

In the following section I am going to reach a yet a higher level of generality, by reflecting about Bell's theorem continuing with the bottom-up approach we have pursued in these pages.

5.4 On Bell's theorem

5.4.1 Under the guidance of evidence

The question I want to ask now is: is it legitimate to favour a particular understanding of Bell's theorem on the basis of the analysis of observation provided in the previous pages, and what would this be? First, we need to individuate what exactly is the starting point from which to attempt an answer to these questions. Throughout the development of the case study in chapters 3 and 4, we have tried to understand both what role observation plays in an analysis of Bell's theorem, and which claims of observability it is possible to make given the available evidence. The former point has been stressed several times already, namely, observation may play a role with respect to the testable predictions of quantum mechanics that violate Bell's inequalities. These are the entanglement correlations and their preservation at space-like separation. The second point has been addressed at three stages in this chapter. Starting from the minimal observational unit of a Bell test, the polarization measurements on each branch of the experiment, we have shown how (i) it is necessary to make additional ontological claims in order to answer the question of whether photon polarization is observable; (ii) there is a range of different additional claims, belonging to the different interpretations of the theory, and they yield mutually contradictory explanations of the data. Nevertheless, they all agree upon the fact that direct observation, in Shapere's sense, of the state pre-measurement is not possible. We then have focussed on the observability of the correlations, and on the space-like separation of the measurement events, which constitutes the crucial limit to a direct observability of the correlations. We have concluded with the important remark that, independently of the question of direct observability, the empirical evidence confirming the violation of Bell's inequalities is abundant, and solid. Despite the heated debates on the theorem's meaning, no one disagrees on the reliability of the empirical corroboration of the quantum predictions. The *perspective* that we are adopting here is precisely this latter: we are going to consider what one would say about the meaning of Bell's theorem starting from the evidence.

Let us outline what could be four premises of this perspective. As the present thesis has been framed within a realist account of scientific theories, realism is going to be part of these premises. It would be interesting to make an analysis of how the argument would change with anti-realist assumptions, but this will be subject of separate work outside the thesis. Apart from (a), the following are not postulates or hypotheses, but facts not usually debated. For this reason, to ask whether together they favour a particular interpretation of the theorem is a legitimate question.

(a) Scientific theories yield true claims about the world; or, they refer to real entities/properties/structures of the world.

(b) Our experiences of objects in the world, and the experimental evidence we can produce about objects in the world and their properties is restricted to three spatial dimensions.

(c) In order to make any experiment in the world, one needs to rely upon a theory of space-time. Currently, this is special relativity, which restricts the possible physical interactions to local interactions.

(d) Testable predictions with respect to Bell's theorem are the entanglement correlations and their occurrence at space-like separation.

Premise (b) can also be formulated more mathematically, by stating that the dimensionality of the empirical world as we experience and test it is given by $\mathbb{R}^3$. The latter point is debated in relation with metaphysical interpretations of the quantum formalism. In particular with respect to forms of wave-function realism (cf. Ney and Albert (2013)), which sometimes argue for configuration space realism, it is discussed how to account for the three-dimensional world we seem to live in, according to our experience. Again, behind premise (b) there is no particular metaphysical assumption with respect to this debate.

So, on the basis of these premises, the outlook we have about the evidence is the following: there is a conspicuous set of measurement outcomes, which, analysed statistically and given the design of the experiment, suggest that the instantaneous events that took place at space-like separation cannot be explained by means of local hidden variables. This is the way in which the experimental papers on Bell tests qualify the results. Keeping assumptions (a)-(d) fixed, the evidence brings to the following possible conclusions.

1 — This is evidence of non-local phenomena: if we are realist about what the theory states about entanglement, we believe in special relativity, and we are modelling objects in three-dimensional space, then there seems to be no choice but to see the occurrence of the correlations at space-like separation as a symptom of a non-local behaviour. This manner of interpreting the situation gives ontological prominence to quantum mechanics, entering in contrast with special relativity. Although one may preserve no-signalling and declare a "peaceful coexistence" between the two theories, this solution is not fully satisfactory. *Really*, the no-signalling principle is based more on an empirical impossibility than on a theoretical basis, so the peace it maintains has a merely instrumentalist flavour. In the sections at pages 88 and 91 we have explored two possible ways in which to understand the violation of the factorizability condition which the CHSH inequality obeys. Importantly, the latter is tightly dependent on (b), as, to hold and to make sense it requires the possibility to model physical objects in a three-dimensional space. It is interesting to see that the most prominent interpretations that hold that there is non-locality in the world are Bohmian Mechanics and GRW theory, which are both postulating a primitive ontology for quantum mechanics in three dimensional space; this view was held also from some members of the Copenhagen interpretation, for instance cf. Heisenberg (1949, p. 39), but we are not entering in the details of this view. To the fact that the entanglement correlations cannot be accounted for by means of local hidden variables, they have replied by postulating *non-local hidden variables*, and by developing fully-fledged new theories. Apart from GRW theory with an ontology of flashes (Maudlin (2011, ch. 10)), GRW theory with an ontology of matter density field and Bohmian Mechanics are in open conflict with special relativity — there is literature on how Bohmian mechanics still does not allow superluminal signalling, and this limit is grounded on further postulations which fine-tune the parameters of the theory.[6] Consider the CHSH inequality. With a particular specification of variables λ, it so happens that choices of settings are correlated with the distant outcomes, but not in an exploitable way to send superluminal signals. Bohmian mechanics accounts for this specification of λ in terms of the *equilibrium hypothesis* (Dürr et al., 1992), which is a particular probability distribution (ρ) over the configuration space where the wave-function (ψ) of the configuration of particles is specified, which is equivalent to the distribution given by the Born's rule ($\rho = |\psi|^2$). The important point is that according to these theories, the meaning of Bell's theorem is to prove that the world as quantum is fundamentally non-local.

2 — Given premises (a)-(d), one may also make another conceptual operation, namely, to deflate the ontological significance of quantum mechanics, and to rather consider single events and their relations to one another as primary. In this view, one would think of the world as a set of events that are structured and regulated

[6]For an exploration of the relation and (in)compatibility between GRW and special relativity see Albert (2000).

by the relativistic postulates. Quantum mechanics would still be fundamental, but it would play the role of a non-classical probabilistic theory and not of a theory modelling an ontology. Two representative interpretations going in this direction are relational quantum mechanics, and the pragmatist interpretation. The first explicitly takes quantum mechanics to be a theory about the exchange of information between physical systems, the quantum state being the object in the theory that encodes such information. The caveat in this understanding is that the state needs to be specified in relation between at least two systems. What the violation of Bell's inequality reveals in this case is that it is not possible to provide a unified, view-from-nowhere account of the experimental evidence (Goldstein et al., 2011, sect. 10.7), (Smerlak and Rovelli, 2007). More concretely, the structure of the world as conceived within the relational view does not allow to provide a unique specification for all the variables belonging to separate light-cones. There is no such a unified description of the world in a relational universe, but only discrete accounts relative to the systems one is considering. Quantum mechanics provides the correct formalism to encode information transmitted within the limits enforced by the relativistic spatio-temporal structure; the meaning of Bell's theorem is to show that such a unified description proper of classical physics does not exist.[7]

I have also listed the pragmatist view of quantum mechanics under this second point because it seems to make a move analogous to the relational interpretation. Although the pragmatist view does not hold that quantum mechanics is directly modelling an ontology, it is not an anti-realist approach in general. There are agent-independent matters of fact which are correctly accounted for by science. The manner in which quantum mechanics is conceived in this view has similarities with the relational view, with the difference that in this case the theory has prescriptive value with respect to what a situated agent ought to believe with respect to a matter of fact (Healey, 2017b, ch. 4). Despite the different role attributed to the quantum state, the meaning of Bell's theorem is the same as in the relational view: there is no possible view from nowhere that uniquely describes a physical situation, but every event is a matter of fact relative to — in this case — a situated agent.

To summarize, premises (a)-(d), based on a serious consideration of the empirical evidence, lead to two possible paths: **1** holds that quantum phenomena are non-local, while **2** attributes a particular function to the quantum state, namely, to encode information about a system relative to another or relative to an agent.

With this argumentative path under our belt, let us move now to a consideration

[7] There are objections to the view of Bell's theorem within the relational framework, precisely spelled out for instance in Laudisa (2019).

of Bell's theorem at a more general level.

5.4.2 A subset in the main set

The two interpretations that are compatible with premises (a)-(d) are not the only ways in which Bell's theorem can be interpreted. Since an overview of its alternative interpretations has been provided in detail already in chapter 3, section 3.2.4, I am not repeating here an in depth exploration. It will be merely recalled that the manner of exposition has followed the premises of the theorem that each interpretation rejects in light of the quantum results.

The superdeterministic interpretation refuses the assumption of "no superdeterminism". Interestingly, it was forged in an attempt to account for the entanglement correlations, rather than to interpret quantum mechanics per se. Another set is constituted by those that reject (P_4) – "Every observed event is an absolute single event not relative to anything or anyone", such as the two theories we have mentioned, relational quantum mechanics and the pragmatist interpretation, but also the anti-realist QBism, and, more importantly, the family of the Everettian interpretations, such as Many Worlds. Finally, we have also considered the interpretation that has received the least attention in the philosophical literature, namely, that which rejects an assumption of *classicality*, which enables the formulation of a Bell's inequality in the first place. This interpretation takes the theorem — together with Kochen-Specker theorem — to be a proof of the radical inadequacy of the formalism of classical physics to model the world.

What strikes about this list — except from the theories subsumed under **2** — is the fact that they are not directly sustained by the set of premises (a)-(d). In the Many Worlds Interpretation, every experiment realizes all the possible outcomes each in a separate branch of the universe, which means that every copy of the observer in a different branch is witness to a different state of affairs. This does not directly contradict (a)-(d), but it makes the Many Worlds Interpretation more distant from these premises as opposed, e.g., to the views subsumed under **1** and **2**. QBism denies (a), but we already clarified that the anti-realist option would not have been considered here. Superdeterminism and the retrocausal interpretation are not directly rejecting any of (a)-(d), but they are not directly supported by either of them. In particular, the fact that our experience of the world and the experiments we perform in it are embedded in three spatial dimensions does not *directly* entail that future-to-past interactions are as valid as the past-to-future ones, which we test and experience. Similarly, none of these assumptions directly supports the option that the value of *every* single variable associated to every single system has evolved deterministically since the beginning of the universe, in such a way that everything is correlated with everything else, as in the superdeterministic universe. Furthermore, the arguments related to the rejection of the pre-quantum formalism to model physical systems can be developed in almost complete independence from empirical tests, and premises (a)-(d) seem to be quite remote from the perspective of these arguments.

This series of remarks intends to bring attention to the fact that the theories subsumed under **1** and **2** are only part of a bigger group containing all the viable views of the theorem. With respect to this fact, I would like to advance three remarks.

First, the theories and interpretations grouped under **1** and **2** have all been developed as ways to address the measurement problem and to provide a coherent picture of the world which would accommodate the theory's structure. Their extent is therefore strictly independent of Bell's theorem and of its interpretation. Nevertheless, the fact that they are *also* directly supported by premises (a)-(d), which are all facts apart from (a), can be used as a strong argumentative basis in their favour. The first remark I want to put forth is the following. I argued that premises (a)-(d) underlie an analysis of the theorem *starting from the evidence*. If one believed that they constitute a solid argumentative basis to favour a scientific theory over another, they would be able to single out the four interpretations we subsumed under **1** and **2** as preferable over the others. Conversely, these four interpretations are those that fare most naturally with the perspective of empirical evidence.

Nevertheless, these premises and their compatibility with the theories we mentioned are not sufficient ground to refute or to rule out any of the other theories and interpretations mentioned above. The second observation I am advancing is that there is another side of the coin to the previous point, namely, the potential power that premises (a)-(d) have to single out a particular subset of theories could be detrimental, as all the alternative views we have recalled are as equally empirically adequate, and they could not be "reached" via reliance upon these premises. In other words, if on one hand they work as a basis to individuate theories that fare harmoniously with this set of facts, on the other they risk to stand in the way of accessing the full range of available theories and interpretations. Whether this should be seen as a problem or not will be addressed in a minute.

Finally, recall the claim advanced in sections 4.4 and 5.3.3, that there seems to be a measure of distance between the formalism of quantum mechanics and actual specific objects modelled by its means. This form of independence of the theory's structures from a specific ontology opens interpretational problems related to quantum mechanics that largely exceed the empirical significance of the theory. For the exploration of the latter type of problem, the study of empirical evidence may be of very limited relevance.

These three remarks seem to individuate two macro-schools of thought. One is represented by the idea that empirical evidence and the way we experience the world are what scientific theories account for. Premises (a)-(d) and the theories directly supported by them are an example of this school. The other endorses the idea that scientific theories can be studied beyond their empirical adequacy, and that the image of the world would emerge as a *consequence* of their structure rather than the other way around. Examples of theories that seem to be part of the latter school are the Everettian interpretations and the refusal of classical physics as a universal theory.

These schools of thought seem to distinguish two very different approaches to the interpretation of the theory, which have consequences on which theories one may endorse. In the next two sections I am going to (a) further explore their features, and (b) advance a hypothesis about different normative grounds where they have their roots.

5.4.3 Two schools of thought and their theoretical virtues

In this section I will outline more precisely the character of these two schools of thought, and then I will individuate what are their main theoretical virtues. Let us then start from the first school, which gathers the theories more directly compatible with premises (a)-(d). Since in this thesis I am focussing on quantum mechanics, I am concentrating mostly on Bohmian Mechanics and GRW theory, as they endow quantum mechanics with ontological primacy, so I am not going to focus further on the theories grouped under **2**.

As we mentioned, each of them proposes a solution to the measurement problem by providing a physical account for the phenomena we witness when we perform experiments aimed at revealing quantum behaviours. One crucial trait they share is the fact that both of them postulate a primitive ontology for the world, for which they describe the behaviour via new or modified dynamical laws modelled on the blueprint of the original quantum formalism. The silver lining of the primitive ontology program is that physics accounts for matter in three-dimensional space, or four-dimensional space-time, an idea that in the pertinent literature is traced back to John Bell (Bell, 1987), but that has been widely developed after: (Allori, 2013), (Allori et al., 2008), (Esfeld and Deckert, 2017), (Gao, 2018).

The effect of this well-developed philosophical foundation is that it provides a further motivation for the structure of these two theories, and for the additional postulates they rely upon. For instance, the particle ontology of Bohmian mechanics is naturally described by the pilot wave equation, which explicitly models the evolution of configurations of particles, and accounts coherently for the precise outcomes obtained upon measurement. GRW theory, on the other hand, can be specified with respect to two possible primitive ontologies, but in both cases it accounts for the same set of empirical phenomena. Both Bohmian mechanics[8] and GRW theory adhere quite closely to the world as we experience it, and in which we perform scientific experiments: it is a long standing position, as clarified in the beginning of each reference cited above, *that physics should account for matter in three dimensional space*, and this is what these theories achieve.

The second school is less considered in the philosophical literature, and it voices more closely how quantum mechanics is studied in physics departments. Rather than starting from the assumption that physics accounts for matter in space-time and modelling the theory accordingly, they start form the fact that quantum mechanics provides a complete set of axioms — comparable for instance with the axioms of classical mechanics —, and from the fact that the predictions it yields are corroborated by evidence, and they study the features and applications of the theory as originally formulated. With the latter expression I am referring to textbook quantum mechanics, for instance Le Bellac (2011) lists axioms and postulates of the theory and their applications, or to the "bare quantum formalism" as phrased in the first pages of Wallace (2008). Examples of applications of the theory are the concrete experimental research in the quantum context, but also quantum information theory, and the study of interpretational problems of mathematical nature.

In the foundations of quantum mechanics, the interpretations that can be associated with this school of thought are the Everettian interpretations. As the theories associated with the first school of thought, these interpretations as well provide a solution to the measurement problem, but they start form a different basis. Namely, they take the theory to be correct, on the basis of its strong empirical adequacy, and they "adapt" the picture of the world in function of such a basis. The Everettian interpretations are realist about the quantum mechanical formalism, and in particular they take the quantum state to represent the structure of the world. The most

[8] I am not entering in the matter of debate of whether Bohmian Mechanics is actually a Many Worlds Interpretation in disguise. The primitive ontology of matter points associated with it is postulated to live in three-dimensional space.

famous of these interpretations is Many Worlds, which takes the "branches" of the wave-function to account for branches of the universe. This metaphysical picture of the world comes out as the object that is directly described by the quantum formalization of systems and states. Silver lining of these interpretations is the fact that they study a successful theory and try to understand why it succeeds, and they make hypotheses about what the formalism may be *revealing* about the world.

If one asked what these two schools of thought *do* in a nutshell, a possible reply could be that the first holds on to a deeply rooted intuition about how the world is, namely, matter in space-time, and uses physics to account for the world so-conceived. For this reason, the formal structure of the theory is moulded and complemented with the necessary postulates to achieve this aim. The second school follows the guidance of how systems and dynamics are modelled in fundamental contemporary physics, and moulds the metaphysics of the world in a way that is compatible with the structure of the theory. For this reason, the cumbersome metaphysics resulting from a realist interpretation of the quantum formalism is not a major obstacle, but a possible discovery about the structure of the universe.

The neutral manner in which they have been presented reflects the fact that these theories and interpretations are actually underdetermined by evidence, and that none of them is affected by obvious flaws which would make their rejection compelling. Which brings us to the next step of the argument. Even if underdetermined, these views can be differentiated on the basis of their theoretical virtues. In philosophy of science there is a long standing debate on what are the theoretical virtues of theories, and what role they play in theory choice. Classical reference for the origins of this debate is Kuhn (1996), but I am following the recent work by Samuel Schindler (2018). Let us consider what these virtues are and whether the three theories we are considering possess them.

Empirical adequacy is the first, which refers to the theory's ability to account both for measurement outcomes, and for the features of observable phenomena. Proponents of both schools of thought have argued for the empirical adequacy of their views. Theories can also be internally consistent, which is a property possessed by all the mentioned interpretations, but they can also be assessed on the basis of their external consistency, namely, with respect to other accepted scientific theories. In this case, the theory to which quantum mechanics is usually related is relativity, and, again, every interpretation specifies differently its relation with it. The unificatory power of a theory is a virtue that applies more properly to new theories, as it refers to the ability to explain a range of phenomena previously thought to be unrelated.

Quantum mechanics in general has this virtue, but its interpretations have not added new insights on a unified explanation of phenomena. The ability of a theory to lead to *new* testable predictions is referred to as fruitfulness; all the mentioned theories hypothesize phenomena which could be tested in principle,[9] but, even when these tests have been performed, they have not yet produced sufficient evidence to break the underdetermination.

The last two theoretical virtues are more controversial for the case at hand, and they are that the theory should not be ad hoc or contain ad hoc hypotheses, and simplicity. There is a debate on how to define something as "ad hoc", and whether it is a problem of hypotheses or of theories more generally (Schindler, 2018, ch. 5). Precisely given the underdetermination among the current quantum interpretations, some of the hypotheses they advance to phrase a solution to the problem must be ad hoc. Thus, also the three theories we are considering contain ad hoc hypotheses, although they have different characters. Let us provide representative examples.

One of the postulates of Bohmian Mechanics has already been mentioned in detail in the previous pages, namely, the equilibrium hypothesis, which justifies why the theory does not allow to send superluminal signals.[10] The frequency rate at which spontaneous "collapse"— i.e. non-unitary evolution — of the Schrödinger's equation takes place in the GRW theory is a parameter inserted *by hand*, and thus considered ad hoc — see Ghirardi et al. (1986, sect. 7) for an explicit justification of the choice of the spontaneous localizaton parameter. All the Everettian interpretations face the "preferred basis problem", namely, they ought to justify the fact that even if we live in a branching universe, we seem to experience only one branch, within which we reliably use a probabilistic algorithm to account for measurements. The theory then has to account for this appearance via "seek[ing] sets of preferred, effectively classical but ultimately quantum states that define branches of the universal state vector, and allow observers to keep reliable records." (Zurek, 2010, p. 410). The various proposed solutions to this problem have been repeatedly criticised for being ad hoc.

Despite the general undesirability of ad hocness, it is not straightforward to

[9]Weak measurements may be a path to reveal Bohmian trajectories (Mahler et al., 2016), interference experiments on macroscopic bodies could rule out spontaneous collapse theories and verify Many Worlds (Vaidman, 2018, sect. 5), and GRW yields a variety of potentially testable predictions (Ghirardi and Bassi, 2020, sect. 7). The claims advanced with respect to these experiments are however not uncontroversial, and they are not shared by the entire physical community.

[10]Another example is the postulate of the quantum potential, which plays a role in the force determining the trajectory of particles — see Belousek (2003) and references therein for an exposition of the *ad-hocness* of this hypothesis and the proposal of a solution. This hypothesis would be ad hoc because, apart from a strong analogy with the potential in classical mechanics, there aren't further motivations for its postulation. In other words, it does not fare as naturally in the theory as its other postulates.

declare these ad hoc hypotheses as *obvious* downsides of the theory. The reasons why they were introduced in the first place, and the fact that they are justified within each theory's framework reduces their arbitrariness. Which shows that a rejection of the theory on the sole basis of its reliance upon ad hoc hypotheses is not possible: the ad hocness of a hypothesis becomes positively or negatively characterized only within a broader argument.

Finally, a very popular theoretical virtue for the evaluation of physical theories is simplicity. This virtue acquires meaning only from the comparison between different theories and the specification of a parameter with respect to which to assess the theories' simplicity. In the present context relevant notions could be syntactical and ontological simplicity (Schindler, 2018, p. 14). The first refers to the formal simplicity of a theory: for the case at hand, the Everettian interpretations are the simplest syntactically, as they merely interpret the "bare" formalism. In comparison with this one, Bohmian mechanics and GRW appear intuitively as less syntactically simple, but defenders of these latter would hardly agree that their interpretations are not formally simple. On the other hand, ontological simplicity is perhaps even less straightforward to argue for, as, again, there is no universal agreement in the community upon an objective criterion according to which a theory is more or less ontologically simple than another. The unique kind of entity postulated by Bohmian mechanics seems the simplest, if compared with the number of branches of the Everettian universe. The Many Worlds supporter may rebut that the plurality of branches does not complicate the ontology, since it does not affect the uniqueness of the universe, and because the branches themselves are not fundamental, but emergent (Wallace, 2013). In any case, as for non ad hocness, a comparison of these theories' simplicity *alone* does not yield a particular judgement of value that allows to discard or to opt for one particular theory. It should also be added that the extent to which these theoretical virtues are possessed or lacked by the mentioned theories and interpretations is object of discussion in the philosophy of physics; therefore, the exposition I made in the previous lines is not universally agreed upon.

To make a point of the situation: in this section I have outlined the features of what I have called two schools of thought which group, respectively, Bohmian Mechanics and GRW theory, supported by premises (a)-(d), and the Everettian interpretations. Despite these theories are underdetermined, they possess distinctively different theoretical virtues, which are generally summoned in the context of theory choice. But, we have stressed how the possession or lack of some of these virtues does not entail *per se* the endorsement or rejection of either of these theories. Yet, different members of the philosophical and physical communities are very clear about expressing strong preferences for one theory over another, and also in arguing why a particular theory *ought* to be rejected. The emphasis on this point is not casual: what I advance in the next section is precisely that at the basis of these disagreements there are judgements about what a theory *should or shouldn't do*, namely, judgements of a normative character.

5.4.4 A normative apple of discord

If one asks: why do you endorse the Everettian interpretation? A standard reply will be: because it is formally minimalistic, elegant, and it makes sense of the original quantum theory. Aren't you worried that it depicts the world in a metaphysically cumbersome and counter-intuitive manner? No, as it lays out a clean and straightforward way to interpret the formalism in such a way as to solve the measurement problem, and by maintaining a high level of generality. But, don't you think that the simple ontology of theories such as Bohmian Mechanics and GRW fares better overall with our intuitions about the physical world, rather than the postulation of a multiplicity of universes that we will never interact with? Maybe they have indeed a stronger phenomenological valence, but this is not what quantum mechanics ought to do. *The enterprise of science is to describe and explain reality, and quantum mechanics is "the most predictively powerful, most thoroughly tested, and most widely applicable theory in scientific history"* — (Wallace, 2012, pp. 1, 35). Thus, the dilemma between replacing such a theory and changing the enterprise of science is a false one: we already have a theory that fulfils this duty, and the Everettian interpretations simply take quantum mechanics as a straightforward description of the world.

The reference to *reality* above seems to hide the further assumption that our experience of the world is not relevant for how the scientific account of reality is constructed. Of course, it is important that the theory is empirically adequate, and the supporters of the Everettian interpretations argue that it is. But if the resulting image is that of a branching universe, then it is of little or no importance that it is not graspable via human experience.

The supporters of the primitive ontology program start off from a very different assumption:

- Any fundamental physical theory is supposed to account for the world around us (the manifest image), which appears to be constituted by three-dimensional macroscopic objects with definite properties.

- To accomplish that, the theory will be about a given primitive ontology: entities living in three-dimensional space or in space-time. (Allori, 2013, p. 60)[11]

One may disagree specifically with the arguments presented by the proponents of this philosophical program, as they lend themselves to a possible array of general objections; these however exceed the purposes of my analysis, which is focussed on singling out the conceptual justifications put forth as reasons to defend such a program. As already stressed in the description of the first school of thought, to which belong the theories that model a primitive ontology, this assumption about what scientific theories ought to do fares very naturally with the manner in which Bohmian Mechanics and the GRW theory have been modelled. To the question of whether a supporter wouldn't be concerned with the additional structure of these theories, which accounts precisely for matter in three dimensional space, they could respond that no, these extra structures and modifications of the original theory are justified as far as they are in line with what is the aim of the scientific enterprise.

It should not be difficult to see that the schools of thought I have characterized above can be associated with these two normative positions about what science *ought* to do. We have linked the first school with premises (a)-(d), which we used to frame the perspective one may adopt about a theory starting from the empirical evidence. The similarities between those premises and the quote above about the conceptual foundations of the primitive ontology program should not surprise: both are directly connected with Bohmian Mechanics and GRW theory. The second school contains as a *realist* view precisely the Everettian interpretations, which define a different general aim for the scientific enterprise.

The reference to a normative sphere in the assessment of different scientific theories has been explored in the philosophy of science, in particular when arguments based on the analysis of the history of science have been brought into the discussion. As argued again by Schindler (2018, ch. 7), to understand how science works a pure analysis of facts does not provide a complete picture; representative defender of this position is Larry Laudan's "normative naturalism" (Laudan, 1987), which has been largely criticized. It is however not necessary to go so far to find reference to normative roots to the different positions in the debate at hand. A sharp analysis of the current undertetermination of the various interpretations made recently by Craig Callender points at the fact that "The choice of research program ... hangs on deep, hotly contested and familiar matters in philosophy of science—in particular,

[11]See also Esfeld (2014, p. 99) and Allori (2015, p. 107), where the normative character of this program is explicitly stated.

the relationship, if any, between explanatory virtues and confirmation." (Callender, 2020, p. 75). In this section I am arguing that this relationship is seen differently by supporters of the different theories in light of their normative assumptions about what science, and quantum mechanics too, should do. Callender's paper is part of an anthology on realism and quantum mechanics, and it is listed within the section about the underdetermination of the quantum interpretations. A paper from David Wallace in the same section aims to provide arguments in favour of the Many Worlds interpretation in light of its success in supporting intra-theoretic relations among different specific quantum theories. What is relevant here is that he claims that his advocacy of the Many Worlds Interpretation is "not a coincidence: this chapter is a codification and development of what I have long thought to be the strongest reasons for that interpretation (building on preliminary versions of these arguments presented in Wallace (2008, 83–5) and Wallace (2012, 33–5))." (Wallace, 2020, p. 80). The second reference contains precisely the statement we have quoted in the beginning of this section, about what the enterprise of science is, while the first is a justification for the Everettian solution of the measurement problem.

❈

One may disagree that the ultimate reasons to endorse a scientific theory in a case of underdetermination are *necessarily* normative. This is indeed not the point I am trying to make. Rather, in this section I have tried to bring attention to the fact that an evaluation of the different interpretations' theoretical virtues does not build a sufficient reason to single out the objectively best theory. We have seen that different theories have different virtues, but this remains an unhelpful fact until an argument is made as to *why* some virtues are to be preferred over others, and the replies to this demand for reasons have normative character. I have shown that this point has been made in the literature, but it hasn't perhaps been emphasized enough in the quantum interpretations controversy. The answers to the why-question just asked appear in fact in the simple form of unchallenged assumptions.

5.5 Conclusion

In this chapter I have put in context the role of observation with respect to the general analysis on Bell's theorem. In particular, I have shown what role this process plays for the interpretation of the theorem, by individuating what are the elements that are directly observable, and by tracing a clear boundary to the observable. By referring clearly to a broad range of the theorem's interpretations, I have shown which ones are closer to the effective empirical evidence collected via Bell tests, and which ones are instead mainly based on assumptions that do not strictly rely upon evidence. At all times throughout the chapter it has been made clear that the interpretations of the theorem are directly correlated with interpretations of quantum theory, and this analysis has brought me in fact to address the state of underdetermination into which the latter pays. After an overview of their main theoretical virtues and how the interpretations can be distinguished in terms of the latter, I have appreciated the fact that the mere possession or lack of virtues does not constitute alone a sufficient basis to reject or to prefer any of the interpretations. I have concluded with the suggestion, backed by the relevant literature, that currently the choices to endorse one theory over another are normatively justified in terms of judgements about what science *ought* to do.

Ceci n'est pas une conclusion

This section is in fact merely the snapshot on a still-developing set of ideas. Although it does effectively end the thesis, it does not conclude the research that has been initiated and developed in it. Nor does it point to particular *results*, rigorously dependent on a strict deduction from the presented arguments. What it does is rather to point the attention towards two general considerations, respectively on realism and progress, which I advance here on the basis of what has been done in the previous pages. They won't be representative of all the topics touched, and they explicitly connect to general debates in the philosophy of science, which is the primary area of research of this thesis. Without further ado, let us proceed towards them.

Let us start from acknowledging that, for the research in the foundations of quantum mechanics, observation has rather limited relevance. We have seen this at every stage of this work: the examples reported in the second chapter, Heisenberg's uncertainty relations and entanglement, posit limits to observability that are independent of the actual realizability of an experiment to test them. Moreover, it has emerged very clearly how the derivation of Bell's theorem and the violation of the inequalities by quantum mechanics are both achieved completely on formal grounds. As explored in the last chapter, the role of observation in the Bell tests is relegated to the *minimal observational units*, but the crucial empirical element corroborating the quantum predictions is the reproduction of the correlations at space-like separation. The various interpretations of quantum mechanics and quantum theories we have considered in the third and in the fifth chapter are both answering to the observation question posited in the previous pages, and providing an account of Bell's theorem significance. Importantly, the interpretations of both the "minimal observational units" (see ch. 5) and the general experimental results are all legitimate but mutually inconsistent. Namely, they are underdetermined by evidence.

Independently of quantum mechanics moreover, one major constraint on observation is represented by relativistic locality. Our theory of space-time does not contemplate the possibility of non-local interactions, and this fact posits an interpretational obstacle in particular to Bell's theorem's result. As we have seen in fact, a number of interpretations see the violation of Bell's inequalities by quantum mechanics as a rejection of "local causality", which clashes with relativistic locality unless one endorses the instrumental no-signalling principle. Given that the the hypothetical intrinsic non-locality of the correlations is empirically unattainable, these interpretations resort to additional assumptions that make the available evidence still only an alleged witness of this local effect.

All these remarks lead up to the first point I want to advance, namely, a claim in favour of scientific realism. From a philosophical point of view, the interpretational issues related to the connection between quantum mechanics and the ontology, and in particular the problem of underdetermination, may constitute arguments in favour of antirealism. The theory is empirically adequate, it saves the phenomena *in any case*, but we have difficulties to point at one particular account of the world, as there is a plurality of equally valid interpretations among which to choose. This means that, on the one hand, one could resort to the non-realist interpretations *of quantum mechanics*, but the more interesting claim is that one may point at the underdetermination among the quantum interpretations to defend an antirealist attitude towards scientific theories in general. And yet this is not a common position in the philosophy of quantum mechanics. The quest for defending realism is not fatally hindered by the current underdetermination, but it is actually pursued on non-empirical grounds.

The problem of how to assess scientific theories when empirical evidence is either not currently achievable or impossible to gather in general has been discussed historically in connection with particle physics and cosmology. A very interesting analysis of the relation between scientific theories and the *unobservable* is provided by Shapere (2000), where he argues in particular that the idea that scientific theories ought to be directly connected to currently observable phenomena in order to be scientific is an empiricist stronghold. This claim is grounded on the general view that for an empiricist, theories ought to account for the observable, which they often define via anthropomorphic criteria. In his analysis Shapere suggests that this attachment to what can be experienced by humans is a potential insurmountable limit for the development of theories at the two extreme scales of magnitude (fundamental and cosmological). He rebuts that there are solid grounds to support the non-empirical development of the latter, based on what he calls background knowledge, namely other scientific theories that are empirically corroborated. Shapere argues that the non-empirical development of theories in these domains is directly supported by scientific realism. More recent arguments in favour of non-empirical testability and non-empirical theory assessment have been presented in Dardashti et al. (2019). Regardless of the details of the various accounts of non-empirical testability — a constroversial account is developed and defended by Richard Dawid (2013) — these arguments are relevant because of their strength especially in cases of underdetermination by evidence. Whereas some years ago the realist reply to underdetermination was to delimit its scope and power, or to characterize it as inconsistent — see for instance Laudan (1990), Laudan and Leplin (1991), and Okasha (2002) — the study of non-empirical theory assessment promotes a possible *solution* to underdetermination that also reinforces the realist stance.

The point of connection between this line of arguments and the present work is the fact that, in the case of the underdetermination of the quantum interpretations, the different arguments in support of the realist interpretations are based on normative assumptions on what science ought to do, which are considerations that ground also the development of techniques of non-empirical theory assessment. In the previous chapter we have evaluated the considered quantum interpretations in the light of their theoretical virtues and how they are then weighted on the basis of normative assumptions. In the mentioned research on non-empirical theory assessment, the argument is more general as it is grounded on a normative assumption about what constitutes scientific progress, and which techniques ought to be employed in order to fulfil this definition of progress. Whereas reference to the history of science cannot be used as a set of assumptions from which to deduce what constitutes progress *today*, the acknowledgement that the scientific enterprise is carried out by historically situated communities moving towards unfixed targets is crucial for a definition of progress that underlies which methodologies of theory assessment are elaborated. The normative grounds of non-empirical theory assessment emerges in the opening pages of Daniele Oriti's article on his account:

> Indeed, we are interested in which non-empirical assessment practices can be argued to be *fruitfully* applied by scientists to achieve progress, thus with an inevitable *prescriptive* aspect, implicitly suggesting what scientists 'could be doing' to achieve progress. We do not shy away from this aspect of the discussion, exactly because we approach the methodological issue from the scientist's point of view (it is not philosophers' task to tell scientists what to do, but it is certainly scientists' duty to try to do it better, while listening carefully to the insights of the philosophers). (Oriti, 2019, p. 126)

The claim I am advancing in these pages thus is that underdetermination is not currently seen as a threat to scientific realism. In the case of fields where underdetermination by evidence is more probable, such as fundamental physics and cosmology, current methodologies for evaluating competing theories on non-empirical grounds are in place. It is interesting to notice that, on the one hand, these methodologies have necessarily normative grounds, as they depend on a definition of progress which is in itself a purely normative notion. On the other hand, if progress in these disciplines is to be possible at all, currently it has non-empirical character, which means that it cannot be supported by an anti-realist stance, but it *needs* a realist attitude towards scientific theories to be coherently endorsed — as already noted by Shapere in the article mentioned above.

The second remark I intend to advance concerns more closely progress. It is a fact that the realist interpretations of quantum mechanics explored in particular in the last chapter have been elaborated at least thirty years ago, and that the interpretational problems related to the empirical significance of the theory have been discussed since the beginning of its life. The philosophical discussion on the foundations of the theory has been sharpening conceptual issues, and solving interpretation-related problems. In particular, the attention has focussed on the metaphysical properties of the possible ontologies the theory may model, but it has not added something *new* with respect to the classical study of the theory which started in the 30s — if we take the seminal EPR paper as a symbolic first philosophical problematization of the implications of the quantum formalism. And the situation is not much better with respect of the scenario "post" quantum: one major natural step is the elaboration of a theory that provides a quantum account of gravity and space-time. Examples of theories that purport to do this are string theory and the other versions of quantum gravity, the most well-developed proposal being loop quantum gravity (Weinstein and Rickles, 2020), which have also been worked on for decades.

As we have appreciated, the impasse caused by the underdetermination of the interpretations of quantum mechanics considered in the last chapter is pragmatically overcome via resort to normative assumptions on what science ought to do. On the one hand, it is on the grounds of these assumptions that different members of the physical and philosophical community are able to lay fully-fledged arguments in favour of their preferred view. On the other hand, these assumptions are not unified, and in turn mutually inconsistent.[12] The fact that there may be a plurality of options also at this different level of theory evaluation had already been noticed by Laudan in the exposition of his account of normative naturalism (Laudan, 1987). He advises to

[12] I am referring to the examples we made about the contrast between the normative assumptions of the primitive ontology program as opposed to those expressed by Wallace in his version of the Many Worlds Interpretation.

overcome a stalemate as we have argued happens in the case of underdetermination at hand, by choosing the methodology that best promotes the cognitive *ends* over the alternatives (pp. 25-26). The point here is that it is not clear whether the different assumptions we have individuated silently presuppose particular cognitive ends, and what, exactly these ends would be. In other words, it is not clear whether these normative assumptions on what science *ought to do...* also assume *...in order to reach end x*. Let me unpack this claim. *By itself*, the simple assertion that science ought to do something rather than something else does not imply that science ought to do that particular thing *in order to reach a particular goal*. We have formulated the normative assumptions in the previous chapter with special care precisely not to presuppose some sort of teleological view of the development of science, which in fact none of the proponents supports explicitly. Nevertheless, it is still a fairly neutral claim to think of science as an enterprise that progresses, and to wonder then about which view of progress may underlie these assumptions.

As we have already mentioned in these pages, the notion of progress is eminently normative. It is not in fact a mere synonym of change, but it implies that a particular development is evaluated positively with respect to particular standards or criteria (Niiniluoto, 2018);[13] it is also possible to develop negatively with respect to the latter, as expressed by the notion of regress. The individuation of such standards or criteria is not objective, as it cannot be derived from first principles or from "nature itself", and the history of science shows that its choice has varied greatly over the centuries. Moreover, there is a classical philosophical debate on what is the aim of science, and on what then would constitute scientific progress. The suggestion of Laudan above is to measure different methodologies against the same aim, and to opt for the one that fares better to reach this end. The claim I am advancing is that it is not clear what this aim is supposed to be now, and, from looking at the debate we have considered so far, it is not clear that different groups in the community working on the foundations of quantum mechanics would agree on a unique view on what this aim is.[14]

With respect to the reflections advanced so far, I submit the following. The fact that the disagreement in the current debate on the different quantum interpretations boils down to a disagreement on the normative foundations of these views may have

[13]"To say that the passage from A to B is progress is to make a judgment of value, in the sense that there is an improvement relative to some standards or criteria. The notion of progress should be based on the aims of good science and the task to say what is good science should be made by philosophy." p. 188.

[14]With the latter word I am not committing to a particular view of what is the end of science, for instance as the discovery of the Truth, or the maximization of knowledge based on a succession of approximate truths. The option that scientific work may amount to endless problem solving is as viable as an option as the others in order to define progress.

two-folded significance. First, *thanks to* acknowledging and endorsing a particular normative assumption on what physics ought to do, it is possible to restrict the number of viable interpretations and to discard others, breaking in this way the underdetermination on non-empirical grounds. Second, however, the fact that the viable alternatives have been developed already for decades, and that the matter of disagreement can be currently phrased in terms of a normative disagreement may be the symptom of a stalemate in the current historical period. The debate over how quantum mechanics is to be understood has been developing since its birth, but even the youngest interpretations are now some decades old. What seems to be missing is a clear, unified view on what is the direction in which the current debate should evolve. Both because it is not fully clear what are the current views on progress in the foundations of quantum mechanics, but also because perhaps a clear statement regarding how research at this foundational level should proceed hasn't been issued yet. With this claim I am not calling for action in this direction, as it is not my place to attempt such a task. Rather, I am suggesting that, until this point will be clear, or unless new, decisive evidence will definitively confirm or falsify a particular theory, it is not to be expected that the current stalemate will be overcome any time soon.

Bibliography

Albert, D. Z. (2000). Special Relativity as an Open Question. In H.-P. Breuer and F. Petruccione (Eds.), *Relativistic Quantum Measurement and Decoherence*, Berlin, Heidelberg, pp. 1–13. Springer Berlin Heidelberg.

Allday, J. (2009). *Quantum Reality; Theory and Philosophy*. Boca Raton (FL): CRC Press.

Allori, V. (2013). Primitive ontology and the structure of fundamental physical theories. In A. Ney and D. Albert (Eds.), *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, pp. 58–75. Oxford, UK: Oxford University Press.

Allori, V. (2015). Primitive ontology in a nutshell. (1), 107–122.

Allori, V., S. Goldstein, R. Tumulka, and N. Zanghì (2008, 07). On the Common Structure of Bohmian Mechanics and the Ghirardi–Rimini–Weber Theory: Dedicated to GianCarlo Ghirardi on the occasion of his 70th birthday. *The British Journal for the Philosophy of Science 59*(3), 353–389.

Anderson, C. D. (1933, Mar). The Positive Electron. *Physical Review 43*, 491–494.

Andreas, H. (2017). Theoretical terms in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). Metaphysics Research Lab, Stanford University.

Aspect, A., J. Dalibard, and G. Roger (1982). Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Physical Review Letters 49*(25), 1804–1807.

Aspect, A., P. Grangier, and G. Roger (1981, Aug). Experimental Tests of Realistic Local Theories via Bell's Theorem. *Physical Review Letters 47*, 460–463.

Aspect, A., P. Grangier, and G. Roger (1982). Experimental Realization of Einstein-Podolski-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. *Physical Review Letters 49*(2), 91–94.

Baas, A. and B. Le Bihan (2020). What does the world look like according to superdeterminism? *Forthcoming on: The British Journal for the Philosophy of Science x*(y), n–m.

Bahcall, J. B. and R. Davis Jr. (1982). An account of the development of the solar neutrino problem. In C. A. Barnes, D. D. Clayton, and D. Schramm (Eds.), *Essays In Nuclear Astrophysics*, pp. 243–285. Cambridge University Press.

Bassi, A. and G. Ghirardi (2003). Dynamical reduction models. *Physics Reports 379*(5), 257–426.

Bell, J. S. (1973, 1987). Subject and object. In J. S. Bell (Ed.), *Speakable and unspeakable in quantum mechanics*, pp. 40–44. Cambridge University Press.

Bell, J. S. (1976, 1987). The theory of local beables. In J. S. Bell (Ed.), *Speakable and unspeakable in quantum mechanics*, pp. 52–62. Cambridge University Press.

Bell, J. S. (1977, 1987). Free variables and local causality. In J. S. Bell (Ed.), *Speakable and unspeakable in quantum mechanics*, pp. 100–104. Cambridge University Press.

Bell, J. S. (1987). *Speakable and unspeakable in quantum mechanics*. Cambridge UK: Cambridge University Press.

Bell, J. S. (1990). La nouvelle cuisine. In E. S. Publishers (Ed.), *Between Science and Technology*, pp. 910–928. North Holland.

Bell, J. S., A. Shimony, M. A. Horne, and J. F. Clauser (1985). An exchange on local beables. *Dialectica 39*(2), 85–110.

Belousek, D. W. (2003). Formalism, ontology and methodology in Bohmian mechanics. *Foundations of Science 8*(2), 109–172.

Benitez, F. (2019). Selective Realism and the Framework/Interaction Distinction: A Taxonomy of Fundamental Physical Theories. *Foundations of Physics 49*(7), 700–716.

Bigaj, T. (2020). On quantum entanglement, counterfactuals, causality and dispositions. *Synthese 197*, 4161–4185.

Birkhoff, G. and J. Von Neumann (1936). The logic of quantum mechanics. *Annals of mathematics*, 823–843.

Bleuler, E. and H. L. Bradt (1948, Jun). Correlation between the States of Polarization of the Two Quanta of Annihilation Radiation. *Physical Review 73*, 1398–1398.

Bogen, J. and J. Woodward (1988). Saving the phenomena. *The Philosophical Review 93*(3), 303–352.

Bohm, D. (1952, Jan). A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I. *Physical Review 85*, 166–179.

Bohm, D. and Y. Aharonov (1957, Nov). Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. *Physical Review 108*, 1070–1076.

Bohm, D. and Y. Aharonov (1960). Further discussion of possible experimental tests for the paradox of Einstein, Podolsky and Rosen. *Il Nuovo Cimento (1955-1965) 17*(6), 964–976.

Bong, K.-W., A. Utreras-Alarcón, F. Ghafari, Y.-C. Liang, N. Tischler, E. G. Cavalcanti, G. J. Pryde, and H. M. Wiseman (2020). A strong no-go theorem on the wigner's friend paradox. *Nature Physics*, 1–7.

Born, M., W. Heisenberg, and P. Jordan (1926). Zur quantenmechanik. ii. *Zeitschrift für Physik 35*(8-9), 557–615.

Brukner, C. (2018). A No-Go Theorem for Observer-Independent Facts. *Entropy 20*(350).

Bub, J. and W. Demopoulos (1974). The interpretation of quantum mechanics. In R. S. Cohen and M. W. Wartofsky (Eds.), *Logical and Epistemological Studies in Contemporary Physics*, pp. 92–122. Springer.

Butterfield, J. (1992). David Lewis meets John Bell. *Philosophy of Science 59*(1), 26–43.

Callender, C. (2020). Can We Quarantine the Quantum Blight? In S. French and J. Saatsi (Eds.), *Scientific Realism and the Quantum*, pp. 57–77. Oxford, UK: Oxford University Press.

Carnap, R. (1956). The Methodological Character of Theoretical Concepts. In H. Feigl and M. Scriven (Eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Volume 1 of *Minnesota Studies in the Philosophy of Science*, Minneapolis, MN, pp. 38–76. The University of Minnesota Press.

Carroll, S. (2004). *Spacetime and Geometry; An Introduction to General Relativity*. San Francisco, CA: Addison Wesley.

Cavalcanti, E. G. and R. Lal (2014). On modifications of reichenbach's principle of common cause in light of bell's theorem. *Journal of Physics A: Mathematical and Theoretical 47*(42), 424018.

Cavalcanti, E. G. and H. M. Wiseman (2021). Implications of Local Friendliness violation for quantum causality. *arXiv preprint arXiv:2106.04065*.

Chang, H. and N. Cartwright (1993). Causality and realism in the EPR experiment. *Erkenntnis 38*(2), 169–190.

Clauser, J. F. and M. A. Horne (1974, Jul). Experimental consequences of objective local theories. *Physical Review D 10*, 526–535.

Clauser, J. F., M. A. Horne, A. Shimony, and R. A. Holt (1969, October). Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett. 23*, 880–884.

Conroy, C. Everettian Interpretations of Quantum Mechanics. In J. Fieser and B. Dowden (Eds.), *The Internet Encyclopedia of Philosophy*.

Cordero, A. (2000). Realism, and the case of rival theories without observable differences. In E. Agazzi and M. Pauri (Eds.), *The Reality of the Unobservable; Observability, Unobservability and Their Impact on the Issue of Scientific Realism*, pp. 191–206. Dordrecht: Springer Science + Business Media.

Corry, R. (2015). Retrocausal models for EPR. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 49*, 1 – 9.

Cowan, C. W. and R. Tumulka (2014, 12). Epistemology of Wave Function Collapse in Quantum Physics. *The British Journal for the Philosophy of Science 67*(2), 405–434.

Crowther, K., N. S. Linnemann, and C. Wüthrich (2019). What we cannot learn from analogue experiments. *Synthese*, 1–26.

Crupi, V. (2016). Confirmation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.

Dardashti, R. (2021). No-go theorems: What are they good for? *Studies in History and Philosophy of Science Part A 86*, 47–55.

Dardashti, R., R. Dawid, and K. Thébault (Eds.) (2019). *Why Trust a Theory? Epistemology of Fundamental Physics*. Cambridge, UK: Cambridge University Press.

Dawid, R. (2013). *String Theory and the Scientific Method*. Cambridge, UK: Cambridge University Press.

DeWitt, B. S. (1970). Quantum mechanics and reality. *Physics today 23*(9), 30–35.

Diósi, L. (1987). A universal master equation for the gravitational violation of quantum mechanics. *Physics Letters A 120*(8), 377–381.

Diósi, L. (1989, Aug). Models for universal reduction of macroscopic quantum fluctuations. *Physical Review A 40*, 1165–1174.

Dirac, P. A. (1930). On the annihilation of electrons and protons. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 26, pp. 361–375. Cambridge University Press.

Dirac, P. A. M. (1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 133*(821), 60–72.

Dorato, M. (2000). Measurability, computability and the existence of theoretical entities. In E. Agazzi and M. Pauri (Eds.), *The Reality of the Unobservable; Observability, Unobservability and Their Impact on the Issue of Scientific Realism*, pp. 207–218. Dordrecht: Springer Science + Business Media.

Duarte, F. J. (2019). *Fundamentals of Quantum Entanglement*. Bristol, UK: IOP Publishing.

Dürr, D., S. Goldstein, and N. Zanghi (1992). Quantum equilibrium and the origin of absolute uncertainty. *Journal of Statistical Physics 67*(5-6), 843–907.

Earman, J. (2015). Some Puzzles and Unresolved Issues About Quantum Entanglement. *Erkenntnis 80*(2), 303–337.

Eberhard, P. H. (1993, Feb). Background level and counter efficiencies required for a loophole-free einstein-podolsky-rosen experiment. *Physical Review A 47*, R747–R750.

Egg, M. (2018). Entity realism. In J. Saatsi (Ed.), *The Routledge handbook of scientific realism*, pp. 120–132. London: Routledge.

Egg, M. and M. Esfeld (2014). Non-local common cause explanations for EPR. *European Journal for Philosophy of Science 4*(2), 181–196.

Einstein, A. (1919). Time, Space, and Gravitation. *The Times (London) 28*, 13–14.

Einstein, A. (1936). Physics and Reality. *Journal of the Franklin Institute 221*(3), 349–382.

Einstein, A. (1948). Quanten-Mechanik und Wirklichkeit. *Dialectica 2*(3-4), 320–324.

Einstein, A., B. Podolski, and N. Rosen (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review* (47), 777–780.

Esfeld, M. (2004). Quantum entanglement and a metaphysics of relations. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 35*(4), 601–617.

Esfeld, M. (2014). The primitive ontology of quantum physics: Guidelines for an assessment of the proposals. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 47*, 99–106.

Esfeld, M. and D.-A. Deckert (2017). *A minimalist ontology of the natural world*. New York: Routledge.

Evans, P. W. and K. P. Y. Thébault (2020). On the limits of experimental knowledge. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 378*(2177), 20190235.

Everett, H. (1957, Jul). "Relative State" Formulation of Quantum Mechanics. *Reviews of Modern Physics 29*, 454–462.

Flores, F. (1999). Einstein's theory of theories and types of theoretical explanation. *International Studies in the Philosophy of Science 13*(2), 123–134.

Franson, J. D. (1989, May). Bell inequality for position and time. *Physical Review Letters 62*, 2205–2208.

Frauchiger, D. and R. Renner (2018). Quantum theory cannot consistently describe the use of itself. *Nature communications 9*(1), 1–10.

Freedman, S. J. and J. F. Clauser (1972, Apr). Experimental Test of Local Hidden-Variable Theories. *Physical Review Letters 28*, 938–941.

Frisch, M. (2020). Causation in Physics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 ed.). Metaphysics Research Lab, Stanford University.

Fuchs, C. A., N. D. Mermin, and R. Schack (2014). An introduction to QBism with an application to the locality of quantum mechanics. *American Journal of Physics 82*(8), 749–754.

Fuchs, C. A. and R. Schack (2014, Dec). QBism and the Greeks: why a quantum state does not represent an element of physical reality. *Physica Scripta 90*(1), 015104.

Gao, S. (2018). *Collapse of the Wave Function: Models, Ontology, Origin, and Implications*. Cambridge, UK: Cambridge University Press.

Genovese, M. (2005). Research on hidden variable theories: A review of recent progresses. *Physics Reports 413*(6), 319–396.

Ghirardi, G. and A. Bassi (2020). Collapse Theories. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 ed.). Metaphysics Research Lab, Stanford University.

Ghirardi, G. C., A. Rimini, and T. Weber (1986, Jul). Unified dynamics for microscopic and macroscopic systems. *Physical Review D 34*, 470–491.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of science 71*(5), 742–752.

Giustina, M. (2017). On Loopholes and Experiments. In R. Bertlmann and A. Zeilinger (Eds.), *Quantum [Un] Speakables II*, pp. 485–502. Springer.

Giustina, M., M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger (2015a, Dec). Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Physical Review Letters 115*, 250401.

Giustina, M., M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger (2015b). Supplemental Material: Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. https://journals.aps.org/prl/supplemental/10.1103/PhysRevLett.115.250401/Supplemental_material_final.pdf.

Glick, D. and G. Darby (2018). In Defense of the Metaphysics of Entanglement. In D. Glick, G. Darby, and A. Marmodoro (Eds.), *The Foundation of Reality: Fundamentality, Space, and Time*, pp. 169–180. Oxford University Press.

Goldrei, D. C. (2017). *Classic Set Theory: For Guided Independent Study*. Routledge.

Goldstein, S. (2017). Bohmian Mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). Metaphysics Research Lab, Stanford University.

Goldstein, S., T. Norsen, D. V. Tausk, and N. Zanghì (2011). Bell's theorem. *Scholarpedia 6*(10), 8378.

Greenberger, D. M., M. A. Horne, and A. Zeilinger (1989). Going Beyond Bell's Theorem. In M. Kafatos (Ed.), *Bell's Theorem, Quantum Theory and Conceptions of the Universe*. Dordrecht: Springer.

Gühne, O. and G. Tóth (2009). Entanglement detection. *Physics Reports 474*(1), 1–75.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.

Handsteiner, J., A. S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H. T. Nguyen, I. Sanders, F. Steinlechner, R. Ursin, S. Wengerowsky, A. H. Guth, D. I. Kaiser, T. Scheidl, and A. Zeilinger (2017, Feb). Cosmic Bell Test: Measurement Settings from Milky Way Stars. *Physical Review Letters 118*, 060401.

Hanna, R. (1948). Polarization of annihilation radiation. *Nature 162*(4113), 332–332.

Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. London: Cambridge University Press.

Hardy, L. (1992, May). Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories. *Physical Review Letters 68*, 2981–2984.

Hardy, L. (1993, Sep). Nonlocality for two particles without inequalities for almost all entangled states. *Physical Review Letters 71*, 1665–1668.

Hausman, D. M. and J. Woodward (1999). Independence, invariance and the causal markov condition. *The British journal for the philosophy of science 50*(4), 521–583.

Healey, R. (2012). Quantum theory: a pragmatist approach. *The British Journal for the Philosophy of Science 63*(4), 729–771.

Healey, R. (2017a). Quantum-bayesian and pragmatist views of quantum theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University.

Healey, R. (2017b). *The Quantum Revolution in Philosophy*. Oxford, UK: Oxford University Press.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik* (43), 172–198.

Heisenberg, W. (1949). *The physical principles of the quantum theory*. Toronto, CA: Dover Publications.

Hempel, C. G. and P. Oppenheim (1948 (1965)). Studies in the Logic of Explanation. In C. G. Hempel (Ed.), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, pp. 245–290. New York, US: Free Press.

Hensen, B., H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, et al. (2015). Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature 526*(7575), 682–686.

Hess, K. and W. Philipp (2001). A possible loophole in the theorem of bell. *Proceedings of the National Academy of Sciences 98*(25), 14224–14227.

Hodges, W. (2018). Model theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

Hofer-Szabó, G. and P. Vecsernyés (2013). Bell inequality and common causal explanation in algebraic quantum field theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 44*(4), 404–416.

Holland, P. R. (1993). *The Quantum Theory of Motion: an account of the de Broglie–Bohm causal interpretation of quantum mechanics*. Cambridge, UK: Cambridge University Press.

Horodecki, R., P. Horodecki, M. Horodecki, and K. Horodecki (2009, Jun). Quantum entanglement. *Reviews of Modern Physics 81*, 865–942.

Hossenfelder, S. (2020). Superdeterminism: A Guide for the Perplexed.

Hossenfelder, S. and T. Palmer (2020). Rethinking Superdeterminism. *Frontiers in Physics 8*, 139.

Hájek, A. (2019). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 ed.). Metaphysics Research Lab, Stanford University.

Isham, C. J. (1995). *Quantum Theory; Mathematical and Structural Foundations*. London, UK: Imperial College PRess.

Jaeger, G. (2009). *Entanglement, information, and the interpretation of quantum mechanics*. Berlin-Heidelberg: Springer Science & Business Media.

Jaeger, G. (2014). What in the (quantum) world is macroscopic? *American Journal of Physics* (82), 896–905.

Janotta, P. and H. Hinrichsen (2014, jul). Generalized probability theories: what determines the structure of quantum theory? *Journal of Physics A: Mathematical and Theoretical 47*(32), 323001.

Jaroszkiewicz, G. (2017). *Quantized Detector Networks; The Theory of Observation*. Cambridge UK: Cambridge University Press.

Jarrett, J. P. (1984). On the physical significance of the locality conditions in the bell arguments. *Noûs 18*(4), 569–589.

Kasday, L., J. Ullman, and C. Wu (1975). Angular correlation of compton-scattered annihilation photons and hidden variables. *Il Nuovo Cimento B (1971-1996) 25*(2), 633–661.

Kennard, E. H. (1927). Zur Quantenmechanik einfacher Bewegungstypen. *Zeitschrift für Physik* (44), 326–352.

Kiess, T. E., Y. H. Shih, A. V. Sergienko, and C. O. Alley (1993, Dec). Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by type-II parametric down-conversion. *Physical Review Letters 71*, 3893–3897.

Kim, J. (2003). The american origins of philosophical naturalism. *Journal of Philosophical Research 28*(Supplement), 83–98.

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher and W. Salmon (Eds.), *Scientific Explanation*, pp. 410–505. Minneapolis, US: University of Minnesota Press.

Kolmogorov, A. and N. Morrison (2018). *Foundations of the Theory of Probability: Second English Edition*. Dover Books on Mathematics. Dover Publications.

Kosso, P. (1989). *Observability and observation in physical science*. Dordrecht: Kluwer Academic Publishers.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3 ed.). Chicago: University of Chicago press.

Landsman, K. (2017). *Foundations of Quantum Theory; From Classical Concepts to Operator Algebras*. Cham, CH: Springer Open.

Larsson, J.-Å. (2014). Loopholes in bell inequality tests of local realism. *Journal of Physics A: Mathematical and Theoretical 47*(42), 424003.

Larsson, J.-Å. and R. D. Gill (2004, Sep). Bell's inequality and the coincidence-time loophole. *Europhysics Letters (EPL) 67*(5), 707–713.

Larsson, J.-Å., M. Giustina, J. Kofler, B. Wittmann, R. Ursin, and S. Ramelow (2014, Sep). Bell-inequality violation with entangled photons, free of the coincidence-time loophole. *Physical Review A 90*, 032107.

Laudan, L. (1987). Progress or Rationality? The Prospects for Normative Naturalism. *American Philosophical Quarterly 24*(1), 19–31.

Laudan, L. (1990). Demystifying underdetermination. *Scientific theories 14*, 267–297.

Laudan, L. and J. Leplin (1991). Empirical equivalence and underdetermination. *The journal of philosophy 88*(9), 449–472.

Laudisa, F. (2019). Open problems in relational quantum mechanics. *Journal for General Philosophy of Science 50*(2), 215–230.

Laudisa, F. and C. Rovelli (2019). Relational Quantum Mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 ed.). Metaphysics Research Lab, Stanford University.

Lazarovici, D., A. Oldofredi, and M. Esfeld (2018). Observables and Unobservables in Quantum Mechanics: How the No-Hidden-Variables Theorems Support the Bohmian Particle Ontology. *Entropy 20*(5).

Le Bellac, M. (2011). *Quantum physics*. Cambridge University Press.

Leifer, M. S. and R. W. Spekkens (2013). Towards a formulation of quantum theory as a causally neutral theory of bayesian inference. *Physical Review A 88*(5), 052130.

Lutz, S. (2012). *Criteria of Empirical Significance. Foundations, Relations, Applications*. Utrecht: Zeno Institute of Philosophy, University of Utrecht, Zutphen.

Mahler, D. H., L. Rozema, K. Fisher, L. Vermeyden, K. J. Resch, H. M. Wiseman, and A. Steinberg (2016). Experimental nonlocal and surreal Bohmian trajectories. *Science Advances 2*(2).

Mallesh, K., S. Chaturvedi, R. Simon, and N. Mukunda (2012). States of physical systems in classical and quantum mechanics. *Resonance 17*(1), 53–75.

Maudlin, T. (2011). *Quantum non-locality and relativity: Metaphysical intimations of modern physics*. Oxford: Wiley-Blackwell.

Maxwell, G. (1962). The ontological status of theoretical entities. In H. Feigl and G. Maxwell (Eds.), *Scientific Explanation, Space, and Time: Minnesota Studies in the Philosophy of Science*, pp. 181–192. University of Minnesota Press.

Mayers, D. and A. Yao (1998). Quantum cryptography with imperfect apparatus. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)*, pp. 503–509.

Mermin, N. D. (1993, Jul). Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics 65*, 803–815.

Muller, F. A. (2004). Can a constructive empiricist adopt the concept of observability? *Philosophy of Science 71*, 637–654.

Muller, F. A. (2005). The deep black sea: observability and modality afloat. *The British journal for the philosophy of science 56*(1), 61–99.

Muller, F. A. and B. C. van Fraassen (2008). How to talk about unobservables. *Analysis 78*(3), 197–205.

Myrvold, W. (2016). Lessons of Bell's Theorem: Nonlocality, Yes; Action at a Distance, Not Necessarily. In M. Bell and S. Gao (Eds.), *Quantum Nonlocality and Reality*, pp. 238–260. Cambridge: Cambridge University Press.

Myrvold, W. (2018). Philosophical Issues in Quantum Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

Myrvold, W., M. Genovese, and A. Shimony (2019). Bell's theorem. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019 ed.). Metaphysics Research Lab, Stanford University.

Nerlich, G. (1979). What can geometry explain? *The British Journal for the Philosophy of Science 30*(1), 69–83.

Ney, A. and D. Albert (2013). *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*. Oxford, UK: Oxford University Press.

Niiniluoto, I. (2018). Scientific Progress. In J. Saatsi (Ed.), *The Routledge handbook of scientific realism*, pp. 187–199. London: Routledge.

Norsen, T. (2009). Local causality and completeness: Bell vs. Jarrett. *Foundations of Physics 39*(3), 273–294.

Okasha, S. (2002). Underdetermination, holism and the theory/data distinction. *The Philosophical Quarterly 52*(208), 303–319.

Oriti, D. (2019). No alternative to proliferation. In R. Dardashti, R. Dawid, and K. Thébault (Eds.), *Why Trust a Theory? Epistemology of Fundamental Physics*, pp. 125–153. Cambridge, UK: Cambridge University Press.

Ou, Z. Y. and L. Mandel (1988, Jul). Violation of Bell's Inequality and Classical Probability in a Two-Photon Correlation Experiment. *Physical Review Letters 61*, 50–53.

Papineau, D. (2021). Naturalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University.

Pearl, J. (2000). *Models, Reasoning, and Inference – Second Edition*. New York: Cambridge University Press.

Pearle, P. (1990). Toward a Relativistic Theory of Statevector Reduction. In A. I. Miller (Ed.), *Sixty-Two Years of Uncertainty. NATO ASI Series (Series B: Physics)*, Volume 226, pp. 193–214. Boston, MA: Springer.

Pearle, P. M. (1970, Oct). Hidden-Variable Example Based upon Data Rejection. *Phys. Rev. D 2*, 1418–1425.

Penrose, R. (1996). On gravity's role in quantum state reduction. *General relativity and gravitation 28*(5), 581–600.

Popescu, S. and D. Rohrlich (1998). The joy of entanglement. In H.-K. Lo, S. Popescu, and T. Spiller (Eds.), *Introduction to Quantum Computation and Information*, pp. 29–48. Singappore: World Scientific Publishing.

Price, H. (2008). Toy models for retrocausality. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 39*(4), 752 – 761.

Pryce, M. and J. Ward (1947). Angular correlation effects with annihilation radiation. *Nature 160*(4065), 435–435.

Robertson, H. P. (1929). The uncertainty principle. *Physical Review 34*, 573–574.

Robledo, L., L. Childress, H. Bernien, B. Hensen, P. F. Alkemade, and R. Hanson (2011). High-fidelity projective read-out of a solid-state spin quantum register. *Nature 477*(7366), 574–578.

Ross, D., J. Ladyman, and H. Kincaid (2013). *Scientific metaphysics*. Okford, UK: Oxford University Press.

Rovelli, C. (1996). Relational quantum mechanics. *International Journal of Theoretical Physics 35*(8), 1637–1678.

Rovelli, C. (2018). Space is blue and birds fly through it. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376*(2123), 2017.0312.

Saatsi, J. (Ed.) (2018). *The Routledge handbook of scientific realism*. London: Routledge.

Salmon, W. C. (1971). *Statistical explanation and statistical relevance*, Volume 69. Pittsburgh, US: University of Pittsburgh Press.

Salmon, W. C. (1989). *Four decades of scientific explanation*. Pittsburgh: University of Pittsburgh Press.

Salmon, W. C. (1990). Scientific explanation: causation and unification. *Critica: Revista Hispanoamericana de Filosofia*, 3–23.

Salmon, W. C. (1998). *Causality and Explanation*. Oxford: Oxford University Press.

Scarani, V. (2019). *Bell Nonlocality*. Oxford: Oxford University Press.

Scheidl, T., R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein, and A. Zeilinger (2010). Violation of local realism with freedom of choice. *Proceedings of the National Academy of Sciences 107*(46), 19708–19713.

Schindler, S. (2018). *Theoretical Virtues in Science: Uncovering Reality through Theory*. Cambridge, UK: Cambridge University Press.

Schrödinger, E. (1926, Dec). An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review 28*, 1049–1070.

Schrödinger, E. (1935). Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society* (31), 555–563.

Schrödinger, E. (1936). Probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society* (32), 446–452.

Schurz, G. (2018). Truthlikeness and approximate truth. In J. Saatsi (Ed.), *The Routledge handbook of scientific realism*, pp. 133–148. London: Routledge.

Schutz, B. F. (2009). *A First Course in General Relativity* (2nd ed.). Cambridge, UK: Cambridge University Press.

Sellars, W. (1961). The language of theories. In H. Feigl and G. Maxwell (Eds.), *Current Issues in the Philosophy of Science (1962)*, pp. 57–77. New York: Holt, Reinhart and Wilson.

Sellars, W. (1965). Scientific Realism or Irenic Instrumentalism: A Critique of Nagel and Feyerabend on Theoretical Explanation. In R. Cohen and M. Wartofsky (Eds.), *Boston Studies in the Philosophy of Science*, Volume II. Atlantic Highlands, NJ: Humanities Press.

Sellars, W. S. (1963). Theoretical Explanation. In W. S. Sellars and K. Lehrer (Eds.), *Essays in Philosophy and its History*, pp. 439–454. Dordrecht: D. Reidel Publishing Company.

Selleri, F. and A. Zeilinger (1988). Local deterministic description of Einstein-Podolsky-Rosen experiments. *Foundations of physics 18*(12), 1141–1158.

Sen, I. and A. Valentini (2020a). Superdeterministic hidden-variables models i: non-equilibrium and signalling. *Proceedings of the Royal Society A 476*(2243), 20200212.

Sen, I. and A. Valentini (2020b). Superdeterministic hidden-variables models ii: conspiracy. *Proceedings of the Royal Society A 476*(2243), 20200214.

Shalm, L. K., E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam (2015a, Dec). Strong loophole-free test of local realism. *Physical Review Letters 115*, 250402.

Shalm, L. K., E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam (2015b). Supplemental Material: A strong loophole-free test of local realism. https://journals.aps.org/prl/supplemental/10.1103/PhysRevLett.115.250402/LHFSupplementary.pdf.

Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of Science 49*(4), 485–525.

Shapere, D. (2000). Testability and Empiricism. In E. Agazzi and M. Pauri (Eds.), *The Reality of the Unobservable; Observability, Unobservability and their Impact on the Issue of Scientific Realism*, Volume 215, pp. 153–164. Dordrecht: Springer Science + Business Media.

Shimony, A. (1990). An exposition of Bell's theorem. In A. I. Miller (Ed.), *Sixty-Two Years of Uncertainty; Historical, Philosophical, and Physical Inquiries into the Foundations of Quantum Mechanics*, pp. 33–43. New York, US: Plenum Press.

Shrapnel, S. (2017, 08). Discovering Quantum Causal Models. *The British Journal for the Philosophy of Science 70*(1), 1–25.

Smerlak, M. and C. Rovelli (2007). Relational EPR. *Foundations of Physics 37*(3), 427–445.

Snyder, H. S., S. Pasternack, and J. Hornbostel (1948). Angular correlation of scattered annihilation radiation. *Physical Review 73*(5), 440.

Suppes, P. (1967). What is a scientific theory? In S. Morgenbesser (Ed.), *Philosophy of Science Today*, pp. 55–67. New York: Basic Books.

Suppes, P. (1972). *Axiomatic set theory*. New Jersey, US: D. Van Nostrand Company.

t Hooft, G. (2016). *The cellular automaton interpretation of quantum mechanics*. Springer Nature.

Teller, P. (1986). Relational Holism and Quantum Mechanics. *The British Journal for the Philosophy of Science 37*(1), 71–81.

Thébault, K. (2019). What Can We Learn from Analogue Experiments? In R. Dardashti, R. Dawid, and K. Thébault (Eds.), *Why Trust a Theory? Epistemology of Fundamental Physics*, pp. 184–201. Cambridge, UK: Cambridge University Press.

Tittel, W., J. Brendel, N. Gisin, and H. Zbinden (1999, Jun). Long-distance bell-type tests using energy-time entangled photons. *Physical Review A 59*, 4150–4163.

Tsirel'son, B. S. (1980). Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics 4*(2), 93–100.

Tumulka, R. (2016). The Assumptions of Bell's Proof. In M. Bell and S. Gao (Eds.), *Quantum Nonlocality and Reality*, pp. 79–90. Cambridge: Cambridge University Press.

Vaidman, L. (2016). The Bell Inequality and the Many-Worlds Interpretation. In M. Bell and S. Gao (Eds.), *Quantum Nonlocality and Reality*, pp. 195–203. Cambridge: Cambridge University Press.

Vaidman, L. (2018). Many-worlds interpretation of quantum mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

van Enk, S. J., N. Lütkenhaus, and H. J. Kimble (2007, May). Experimental procedures for entanglement verification. *Physical Review A 75*, 052318.

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

van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

Von Neumann, J. (1955). *Mathematical foundations of quantum mechanics: New edition*. Princeton university press.

Wallace, D. (2008). Philosophy of quantum mechanics. In D. Rickles (Ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*, pp. 16–98. Aldershot, UK: Ashgate Publishing Limited.

Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford, UK: Oxford University Press.

Wallace, D. (2013). A prolegomenon to the ontology of the everett interpretation. In A. Ney and D. Albert (Eds.), *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, pp. 203–222. Oxford, UK: Oxford University Press.

Wallace, D. (2020). On the Plurality of Quantum Theories: Quantum Theory as a Framework, and its Implications for the QuantumMeasurement Problem. In S. French and J. Saatsi (Eds.), *Scientific Realism and the Quantum*, pp. 78–102. Oxford, UK: Oxford University Press.

Weihs, G., T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger (1998, Dec). Violation of Bell's Inequality under Strict Einstein Locality Conditions. *Physical Review Letters 81*, 5039–5043.

Weinstein, S. and D. Rickles (2020). Quantum Gravity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.). Metaphysics Research Lab, Stanford University.

Werner, R. F. (2014). Comment on 'What Bell did'. *Journal of Physics A: Mathematical and Theoretical 47*, 424011.

Wheeler, J. A. (1946). Polyelectrons. *Annals of the New York Academy of Sciences 48*(3), 219–238.

Wheeler, J. A. and W. H. Zurek (1983). *Quantum Theory and Measurement*. Princeton (NJ): Princeton University Press.

Wigner, E. P. (1995). *Remarks on the Mind-Body Question*, pp. 247–260. Berlin, Heidelberg: Springer Berlin Heidelberg.

Winther, R. G. (2021). The Structure of Scientific Theories. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 ed.). Metaphysics Research Lab, Stanford University.

Wiseman, H. M. (2006). From Einstein's theorem to Bell's theorem: a history of quantum non-locality. *Contemporary Physics 47*(2), 79–88.

Wiseman, H. M. and E. G. Cavalcanti (2017). Causarum investigatio and the two bell's theorems of john bell. In R. Bertlmann and A. Zeilinger (Eds.), *Quantum [Un] Speakables II*, pp. 119–142. Springer.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2017). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). Metaphysics Research Lab, Stanford University.

Wu, C. S. and I. Shaknov (1950, Jan). The angular correlation of scattered annihilation radiation. *Physical Review 77*, 136–136.

Yin, J., Y. Cao, Y.-H. Li, S.-K. Liao, L. Zhang, J.-G. Ren, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, G.-B. Li, Q.-M. Lu, Y.-H. Gong, Y. Xu, S.-L. Li, F.-Z. Li, Y.-Y. Yin, Z.-Q. Jiang, M. Li, J.-J. Jia, G. Ren, D. He, Y.-L. Zhou, X.-X. Zhang, N. Wang, X. Chang, Z.-C. Zhu, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan (2017). Satellite-based entanglement distribution over 1200 kilometers. *Science 356*(6343), 1140–1144.

Zeilinger, A. (1999). Experiment and the foundations of quantum physics. In B. Bederson (Ed.), *More Things in Heaven and Earth*, pp. 482–498. Springer.

Zurek, W. H. (2010). Quantum Jumps, Born's Rule, and Objective Reality. In S. Saunders, J. Barrett, A. Kent, and D. Wallace (Eds.), *Many Worlds? Everett, Quantum Theory and Reality*, pp. 409–432. Oxford, UK: Oxford University Press.

Żukowski, M., A. Zeilinger, M. A. Horne, and A. K. Ekert (1993, Dec). "Event-ready-detectors" Bell experiment via entanglement swapping. *Physical Review Letters 71*, 4287–4290.