Chapter 1

WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

1.1 INTRODUCTION: REALISM AND LOCALITY

The classical picture of physical reality rests upon, among others, two intuitively compelling principles. The first of these reflects our natural belief in the objective and independent existence of the physical world. Skeptical arguments notwithstanding, we are inclined to believe that there are objects that constitute our physical environment, and that they possess certain properties independently of whether we are aware of them or not. In philosophical parlance this ontological intuition is typically presented under the multifarious heading of “realism”. Of course realism need not be “naïve”—we know very well that the very act of observation can disturb or even destroy the elements of reality under consideration. However, although there is probably no such thing as an ideal observer or an ideal observation, realists would insist that we can either make the disturbances associated with observations arbitrarily small, or at least take them into account and modify the outcomes of our observations accordingly. To use a standard textbook example: if I observe a stick which is partly submerged in water, and it appears bent to me, I can take into account the fact that light bends upon hitting water and infer that the stick is actually straight. In general, the crux of realism lies in the conviction that it makes sense to speak about objects and their properties even if no one is currently perceiving them, measuring them or is in any other way directly aware of them.

The second classical principle concerns the structure of the external world and the law-like connections between its elements. We believe that objects furnishing our world influence one another in many different ways—changing their behavior, properties, etc. In other words, we perceive a multitude of causal interactions among elements of physical reality. Those interactions and correlations are not entirely chaotic or capricious. They seem to display certain patterns, and the most conspicuous one is that the ability to influence seems to be dependent on proximity. If I want to move an object—a chair for example—I have to come close to it. And if I want
to act from a distance, I have to find a mediating tool, like a broomstick or a fully automated and computerized artificial arm. It seems, then, that the interactions we are familiar with display the feature of being “local”—roughly meaning that to influence a faraway object the causal chain has to somehow pass in a certain amount of time through the entire distance separating us from the object.

In spite of its intuitive appeal, the locality of cause-effect relations has not always been accepted unconditionally. Isaac Newton in his gravitational theory assumed, without much consideration, that the gravitational attraction between distant bodies is immediate and unmediated. The idea of action at a distance apparently wasn’t a cause of concern for the founding father of classical physics. However, his followers were not so sure. Physicists felt that it is a legitimate question to ask whether, for example, the effects of the Sun’s sudden disappearance would be felt instantaneously throughout the Solar system. Would the planets of the Solar system “feel” freed from the Sun’s gravitational bonds immediately as a result of its vanishing or would they follow their usual paths until the disturbance in the gravitational field caused by the sudden annihilation of the Sun reached them? As is well known, Albert Einstein opted strongly for the second answer. He built into his theories of relativity a fundamental principle expressing the locality of the physical world. This principle, of an undoubtedly ontological character, can be formulated in terms of the special theory of relativity as the claim that all events causally relevant to a particular occurrence $O$ have to occupy a specific part of space-time, called “the absolute past of $O$” (“the past light cone of $O$”). In other words, causal influences cannot propagate at arbitrary speeds, the speed of light being the upper limit for all transmissions of mass and energy.¹

Quantum mechanics calls into question the validity of both of the aforementioned principles, however. Probably the most notorious feature

¹ We should mention, however, that there is a slightly different intuition associated with the notion of locality, which does not focus on the limitation of speed at which all causal interactions can propagate, but rather on the idea of the spatiotemporal contiguity of causal chains. This alternative sense of locality can be presented informally in the form of the requirement that for any event $e$, no matter how small a spatiotemporal interval stretching in $e$’s past and containing it we consider, there will be always a set of complete causes for $e$ occurring in this interval. This meaning of locality is obviously different from the one we are presenting in the text, for it is possible that no interactions propagate faster than the speed of light, and yet that there are “gaps” in causal chains leading from the cause to the effect. For more on that issue, see (Lange 2002), in particular pp. 13-17.
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

of quantum mechanics is that it casts doubt on the objective existence of properties that characterize a particular physical system. A complete quantum description of a state for a physical system consists not of functions ascribing precise values to all physical parameters, but rather of a probability distribution function over all possible values for the given parameters. The literature on this subject, professional as well as popular, offers numerous examples and thought experiments illustrating this particular feature of the quantum world. Thus, we learn that when we want to measure a property of light called “polarization”, the result of this measurement for a particular photon is not determined, and we can only speak about a numerically given probability that the photon will pass through a specifically oriented polarizer or that it will get absorbed. However, after the measurement has been done and the photon has successfully passed through the polarizer, its polarization along that particular direction becomes well defined. Hence the idea of a photon having a particular polarization even when nobody is watching begins to fade away.

Even more astoundingly, the same disturbing phenomenon occurs with mundane parameters such as position. The famous two-slit experiment teaches us that it makes no sense to ask exactly which slit the electron went through, even though it manages to reach the detecting screen where it “materializes”. Again, the quantum mechanical description of this situation invokes the probability function known also as the wave function. The wave function enables us to calculate the probability of finding an object in any particular region of space. In the two-slit experiment the wave function consists of two components; one describing the electron passing through the first slit, and the other through the second one. A combination (or superposition) of these two components creates the famous interference pattern observed on the screen.²

However, it should be stressed that realism is not contradicted by the standard formulation of quantum mechanics. Quantum formalism operates with the probability function, but by itself offers no definitive answer to the question of how to interpret this probability. It is logically possible to maintain that quantum probability reflects only our ignorance regarding the exact state of a system, just as in statistical mechanics. This brings us to the perennial debate regarding the correct interpretation of quantum mechanics, with the choice being between the Copenhagen interpretation, the

² (Hughes 1989) offers an excellent and accessible, yet rigorous introduction to these problems. For a more historically-oriented exposition of the foundations of quantum mechanics, see (Jammer 1974).
various hidden variable theories and several other interpretations. Hidden variable theories, which claim that the physical parameters have objective values that go beyond what is given in the standard quantum-mechanical description, have been seriously undermined by several no-go results, some of which will be later analyzed. Still, no one has succeeded in showing their impossibility, so realist alternatives to quantum mechanics, such as David Bohm’s famous mechanics, remain an option, albeit an unlikely one in the opinion of the majority of scientists.

The anti-realist leaning of quantum mechanics has been its hallmark almost since its inception. This wasn’t the case with non-locality. Philosophical interest in the possibly non-local character of quantum mechanics arose only after John Bell’s groundbreaking studies of entangled systems in the sixties, even though the problem with locality in the context of the interpretation of quantum mechanics was suggested much earlier by Einstein and his collaborators. What is particularly fascinating is that these two apparently separate issues—the issue of realism and of locality—turn out to be deeply interconnected within the area of quantum physics. This connection can be hinted at with the help of the following, simple example. Let us suppose that we have an electron in the state in which its position is described by a wave function \( \Psi \) that has only two non-zero components \( \Psi_1 \) and \( \Psi_2 \), one concentrated around a particular location \( A \), and the other around a distant location \( B \). We can imagine that location \( A \) is situated in a lab on Earth, whereas distant location \( B \) is on the Moon. No principle in quantum mechanics forbids the existence of such systems, although for obvious reason their practical realization would encounter serious difficulties. Translating this description into the language of probability, we can say that in these initial conditions the probability of detecting the electron on Earth equals the probability of detecting it on the Moon, which in turn equals one-half. Now let us suppose that we have set up some sort of a detecting device in our lab, and that the detector successfully localized our initially prepared electron as being in area \( A \). What impact does this result of measurement have on the physical situation in area \( B \)? The answer depends on whether we interpret the initial wave function \( \Psi \) as describing merely our ignorance, or as reflecting the fundamental property of the quantum world. In other words, we have to decide if we choose to embrace ontological realism with respect to quantum properties, or to reject it and admit that before the measurement the position of the electron was not objectively defined.
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

If we opt for the first solution, nothing mysterious follows. We detected the electron in our lab, because it had been there all along. The function $\Psi$ represents only our limited knowledge with regard to the exact position of the electron. The fact that after the measurement the second component $\Psi_2$ disappears reflects merely the change in our knowledge (we now know that the electron hadn’t been on the Moon at all), not a change in the objective world, so there is no reason to believe that the measurement on Earth has had any physical impact on the Moon. However, things are different when we agree, as is suggested by the orthodox interpretation of quantum mechanics, that the function $\Psi$ affords a complete and objective description of all there is in the physical world. Suppose then—contrary to the realist stance—that before the measurement the electron wasn’t physically in area $A$ or $B$ separately, but rather “in both”, at least “partially”$^3$. Now we can agree that by detecting the electron in one location $A$, and by “forcing” it to reappear in our lab, we have made some change in the physical situation on the Moon—we have eliminated the “partial”, or “incomplete” presence of the electron there. But the question is whether this change, whose nature is still somewhat mysterious, has to occur instantly, or maybe—as it was the case with the gravitational influence between the Sun and the planets—only after the time required for the signal carrying the information about the outcome of our experiment to travel the distance between the Earth and the Moon. A little thought reveals that the second option is not tenable. If we allowed for even the slightest delay in the flow of signal from the earthly lab to the Moon, we would open the possibility of detecting the electron on the Moon in the short period of time allowed before the signal was received and in spite of its detection on Earth, which is obviously impossible. The change in the physical situation on the Moon has to occur simultaneously with the detection on the Earth, otherwise it would be possible to end up with two electrons rather than one. This means, however, that this purported influence has all marks of non-locality; that is it does not obey the restrictions put forward by Einstein. It propagates at infinite speed, and it also seems not to be transported by any physical means. Hence, if this is correct, it seems that once we reject realism while accepting the quantum-mechanical formalism, we have to reject locality as well.

Interestingly, the foregoing example is rarely taken seriously by philosophers and physicist as a genuine case of non-locality implied by the orthodox interpretation of quantum mechanics. P. Gibbins for example re-

---

$^3$ Robert Griffiths describes this quantum state of a particle as “delocalized” (Griffiths 2001, p. 18).
fers to a similar case as a “naïve” instance of non-locality (Gibbins, 1983, p. 191). The reason for this belittling remark is that the non-locality in this case supposedly follows only under an additional, unreasonable assumption of the physical character of the wave function. According to the standard presentation, quantum measurement is always accompanied by the so-called collapse of the wave function—i.e. the transformation of the initial probability distribution, which reflects the indeterminacy in the parameter to be measured, into a distribution representing the definite result of the measurement (in our situation the collapse transforms the initial function $\Psi = \Psi_1 + \Psi_2$ into its component $\Psi_1$). If the wave function were to be interpreted analogously to electromagnetic or gravitational fields, as some sort of a field-like entity, then an instantaneous disappearance of its distant part would obviously count as a genuine non-local phenomenon. However, most interpreters and commentators of quantum mechanics warn specifically against construing the wave function as having any objective physical existence. Instead, we are instructed to treat it merely as a “calculation device”, a mathematical instrument used to compute probabilities. Consequently, the collapse is not to be interpreted as a physical process, and the transition from the function $\Psi$ to its first component $\Psi_1$ ought not to be seen as involving any physical change in the distant area where $\Psi_2$ used to be non-zero.

However, I think that this argument against the genuineness of the non-locality in the case of a single particle is not conclusive, and that by drawing our attention to the issue of the ontological interpretation of the wave function it draws us away from the real source of non-locality here—the interpretation of probability. I can readily agree that the wave function is nothing more but a calculating device to obtain probabilities. The question, however, is how are we supposed to interpret the probabilities obtained in this way. I believe that one unquestionable way of avoiding non-locality is to admit that the probabilities are to be interpreted subjectively, as reflections of our ignorance. But, as Tim Maudlin rightly pointed out, if “these probabilities are not reflections of our ignorance but of a basic indeterminism in nature, then we must take an event’s having a particular probability

---

4 John Bell, for example, wrote: “One of the apparent non-localities of quantum mechanics is the instantaneous, over all space, ‘collapse of the wave function’ on ‘measurement’. But this does not bother us if we do not grant beable status to the wave function. We can regard it simply as a convenient but inessential mathematical device for formulating correlations between experimental procedures and experimental results” (Bell 1987, p. 23; see also Griffiths 2001, p. 17, p. 261).
as a basic *physical* fact” (Maudlin 1994, p. 147) And an instantaneous change in a basic physical fact on the Moon brought about by a measurement on Earth has to involve some type of non-local influence. Denying this, while accepting some sort of a non-subjective interpretation of quantum probabilities, seems to be intellectually dishonest (Maudlin 1994, p. 148).

The reluctance to see the single-electron case as a strong case of non-locality may also be due to the fact that we don’t have here a clear ontological intuition regarding what entity physically constitutes a “bearer” of the property to be non-locally changed. If the probability of the occurrence of the electron on the Moon is an objective physical property, what is it a property of? An electron? This would only muddle the issue, because before the measurement the electron has no defined localization, so how could we be sure that the change in the probability happened on the Moon? After all, the localization of a property should be primarily tied to the localization of its bearer. The possibility of treating the wave function as the bearer of properties has to be excluded because of the above-mentioned arguments against the objective existence of a physical counterpart of the wave function. A possible response to this challenge could be that it is reasonable to accept space-time regions as bearers of the properties reflected in the appropriate ascription of probability. So, under this interpretation, it would count as a property of spatiotemporal region $B$ on the Moon that the probability of the occurrence of the electron there is one-half, and this property of region $B$ would be subsequently changed by the measurement on the Earth. Still, we should agree that the case for non-locality would look stronger, or at least more compelling, if we could find a material object whose physical property would be changed by action-at-a-distance. And this is exactly what happens in the famous EPR case involving two entangled particles.

### 1.2 THE EPR ARGUMENT: INCOMPLETENESS OR NON-LOCALITY?

As it is widely known, Albert Einstein was very keen on showing that quantum theory, although empirically accurate and amazingly fruitful, does not offer a complete description of physical reality. In other words, whenever a quantum formalism presents us with a (non-trivial) probability distribution over a set of possible values of a physical quantity as the ultimate description of a real physical system, we should expect that reality
holds more—that the quantity in question is actually characterized by a unique value. In order to argue that this is the case, Einstein together with Podolsky and Rosen turned their attention to what is known as entangled quantum systems (Einstein et al. 1935). They noticed that the quantum formalism allows for the existence of pairs of objects whose parameters (observables) $A_1$ and $A_2$ are such that although $A_1$ and $A_2$ may not have separate determinate values, there is a strict functional dependence between their values. As an illustration, they chose two particles whose positions and momentums were correlated as follows: the difference between the location of the first particle $X_1$ and the location of the second particle $X_2$ was guaranteed to be constant ($X_1 - X_2 = a$), and the total momentum of two particles was constant as well ($P_1 + P_2 = b$).

Later, Nils Bohr suggested a thought experiment in which it would be possible to experimentally create a situation like that (Bohr 1935, p. 697). The experiment involves a macroscopic plate with two slits, suspended on springs attached to a rigid frame. When two electrons pass through the slits, their locations relative to the rigid frame are not defined, because the plate can move freely with respect to the frame, while the difference between their relative positions remains determined by the distance between the slits. Hence, if we put a detecting screen right behind the plate and record the position of one electron, we would be able to immediately derive the position of the other electron without detecting it directly. The same applies to the momentum: the recoil of the plate after the passing of the electrons can give us information regarding their total momentum along one axis, so measuring the momentum of one particle in this direction is sufficient to infer the value of the momentum for the second one.

In modern discussions of the EPR argument momentum and position are typically replaced with different components of spin or with polarizations of pairs of photons. This is due to the fact that correlated systems involving position and momentum are very short-lived—the correlation between parameters quickly fades away with the evolution of the system (Dickson 2002). However, the basic idea remains the same. If we create a pair of electrons in the singlet state, in which their total spin equals zero, we can be sure that when we measure the spin of one of them along a specific direction, the spin of the other one in this direction will have to be the opposite. Using this example, we can now reconstruct the original EPR argument aiming to show that the quantum formalism is incomplete. The

---

5 This reformulation of the EPR argument is due to David Bohm (1951, pp. 611-622).
logical structure of this argument is such that one of its premises expresses the locality assumption, and the conclusion is the thesis that there must be an element of physical reality not described by the standard quantum formalism. The authors invite us to imagine an experiment, performed on one of two electrons created in the singlet state, that measures its spin along an arbitrarily chosen direction $x$. Knowing the result of this experiment, which we can denote as $a$, we can now infer, using the assumption of the perfect correlation between the spins of both particles, that the second electron has to have the opposite spin value along the same direction, which we can write as $\sigma_x^2 = -a$. The crucial thing is that we have arrived at this conclusion without physically interacting with the second electron. In order to ensure that this is the case, we have to appeal to the principle of locality, which forbids the existence of instantaneous influences at a distance. So, at the exact time $t$ when the experiment was performed on the first particle no physical change could occur in the vicinity of the second one. Because the inferred property $\sigma_x^2 = -a$ of the second electron cannot be created by the distant measurement, it must have characterized the electron even before time $t$ (this is the essence of Einstein’s so-called criterion of reality). But this plainly shows that the quantum-mechanical description is incomplete, because before the measurement both particles are supposed to be in a totally unpolarized state, i.e. in the state in which the objective probability of obtaining any value of spin in a given direction is $\frac{1}{2}$.

The above sketch of the EPR argument can raise legitimate doubts. The main problem is that it implicitly assumes the existence of absolute simultaneity when it “transfers” the time $t$ of the experiment to the second particle. Yet absolute simultaneity violates the principles on which the entire Einstein’s “local” worldview has been built. This drawback can be corrected by introducing the relativistic relation of space-like separation between distant particles. Now we can replace the statement about what property the second electron should possess at the time of the measurement, with the contention that the second electron has to have the $x$-component of its spin defined in all regions which are space-like separated from the measurement. Because according to the relativistic principle of locality no physical signal can connect two space-like separated regions, it can be conjectured that the second electron should have been characterized by the objective property even in the common past of the two particles (this being a relativistic equivalent of the non-relativistic notion of “the moment right before the measurement”).
Alternatively, it can be claimed that the same physical characteristic of the second electron, which was inferred but not obtained by any physical interaction with the particle, should also exist in the possible situation in which no measurement was performed on the first particle. This should be intuitively clear when we take into account that, because of the locality assumption, the measurement on the first particle should not make any difference in the distant region of space-time; so it seems legitimate to assume that in a possible situation in which no measurement is performed, all the properties of the second particle would remain the same—including the previously derived element of physical reality pertaining to the $x$-component of spin. The second variant of the EPR argument notably makes use of counterfactual reasoning, entertaining a possible but not actual situation when no measurement is carried out. Whether this argument is in fact valid remains to be seen later, but for the moment let us accept it without further analysis. If we accept the above derivation, then it should be quite clear that it subsequently leads to the incompleteness of quantum mechanics, because with no measurements performed whatsoever the particles are not supposed to have their spins defined. Consequently, the net result of the EPR argument can be presented in the form of the following implication:

(EPR) Locality $\Rightarrow$ Incompleteness

which is logically equivalent to

(EPR) Completeness $\Rightarrow$ Non-locality

Let us be more specific as to what the meaning of the Completeness assumption is. It means, basically, that the probabilities provided by quantum-mechanical formalism are all there is to say about reality—that there is no physical reality beyond what is given in the quantum probability distribution. Consequently, quantum systems sometimes do not possess the definitive properties that might meaningfully characterize them at other times. But this amounts precisely to the non-realist viewpoint we talked about earlier. Hence, the result of the EPR argument can be presented as

(EPR) Anti-realism $\Rightarrow$ Non-locality
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

If we give up realism, then quantum formalism forces us to abandon locality as well—that is the direct lesson of the EPR argument. However, there still seems to be room to keep both these cherished principles intact. Implication (EPR) does not exclude the possibility that both anti-realism and non-locality could be false. That is exactly what Einstein wanted to argue for. The only thing which is impossible in light of the EPR argument is to have locality while rejecting realism.

1.3 BELL’S THEOREM AND THE PLIGHT OF LOCALITY

Unfortunately for Einstein, it turns out that the possibility of the peaceful coexistence between realism, locality and the quantum-mechanical formalism vanishes as a consequence of the famous result achieved by Bell in (1964). Bell’s argument proceeded from the hypothetical assumption of realism (in his terminology: hidden variable hypothesis) coupled with the locality principle. In the mathematically simplified version of Bell’s argument, popularized by E. Wigner and B. d’Espagnat (Wigner 1970), we start with \( N \) pairs of electrons prepared in the same quantum state—namely the previously mentioned singlet state. Let us now select three directions in space \( \alpha, \beta, \gamma \) and consider three components of spin along those directions: \( \sigma_\alpha, \sigma_\beta \) and \( \sigma_\gamma \). According to the realist assumption, all particles in question have their spins pre-defined, so we can divide the sample into eight basic categories, depending on the ascription of one of two possible values (let’s symbolize them with + and – signs) to each spin component \( \sigma_\alpha, \sigma_\beta, \sigma_\gamma \). Let us now focus on the following three categories: \( (\alpha, \beta) \), \( (\beta, \gamma) \) and \( (\alpha, \gamma) \), where \( (\alpha, \beta) \) denotes all pairs of electrons such that the value of the \( \alpha \)-component of spin for the first electron is the same as the value of the \( \beta \)-component for the second one (either both are +, or –), and the remaining \( \gamma \)-component has an arbitrary value (either + or –). The remaining symbols are to be interpreted in the same way. Now it is straightforward to notice that the number of the pairs in \( (\alpha, \beta) \) can be presented in the following way:

\[
N(\alpha, \beta) = N(+, -, +) + N(+, -,-) + N(-, +, +) + N(-, +, -)
\]

where a combination of pluses and minuses \( (a, b, c) \) denotes a pair of electrons for which the values of spin of the first electron in directions \( \alpha, \beta, \gamma \) equal respectively \( a, b, c \) (hence the values for the second electron are the
opposite.) In the same way we can argue that the following equations have to hold:

\[ N(\beta, \gamma) = N(+, +, -) + N(-, +, -) + N(+, -, +) + N(-, -, +) \]

and

\[ N(\alpha, \gamma) = N(+, +, -) + N(+, -, -) + N(-, +, +) + N(-, -, +). \]

Noticing that each element constituting the definition of \( N(\alpha, \beta) \) occurs either in the equation characterizing \( N(\beta, \gamma) \) or the equation defining \( N(\alpha, \gamma) \) we can arrive at the following inequality:

\[ (1.1) \quad N(\alpha, \gamma) \leq N(\alpha, \beta) + N(\beta, \gamma) \]

This inequality consists of elements which are not jointly measurable. We have no means of determining simultaneously all spin components for the particles in question. Moreover, numbers like \( N(\alpha, \gamma) \) cannot be directly determined by the quantum mechanical formalism, as they are defined in terms of hidden parameters, not described by the standard quantum theory. The only thing the theory is capable of predicting is the probability that if we decide to perform for a given pair of singlet-state electrons a combined measurement of spins \( (\sigma_\alpha, \sigma_\beta) \) (or any other combination), the results on both electrons will be the same. This probability has an obvious empirical interpretation—namely it is equal to the relative frequency of cases with joint outcomes being either \((+, +)\) or \((-,-)\), among all pairs of particles which underwent the joint measurements in the setting \((\alpha, \gamma)\). Now the most important question is: can we argue that the quantum mechanical probability (the relative frequency of the occurrence of the same outcomes with respect to the number of particles selected for a particular measurement) is numerically equal to the ratio \(N(\alpha, \gamma)/N\), where \(N\) denotes the total numbers of prepared electrons, and not only of those which were selected for the measurement? If that is the case, then we can replace the above, non-empirical inequality, with the following, empirically meaningful one:

\[ (1.2) \quad P(\alpha, \gamma) \leq P(\alpha, \beta) + P(\beta, \gamma) \]

The last inequality can be confronted with quantum-mechanical predictions, as well as directly with experience. It has been verified that there are
particular directions $\alpha, \beta, \gamma$ for which quantum-mechanical predictions significantly violate inequality (1.2). But the question remains, how can we justify the transition from (1.1) to (1.2)? It appears that we have to rely, among others, on a principle evoking the locality assumption. In order to make sure that the outcomes of particular joint measurements on both particles accurately reveal the possessed values that, according to the realist assumption, were already present before the measurement, we have to assume that a distant measurement is incapable of changing the spin value already possessed by the local electron. As unlikely as it may seem, if we allow for a possibility that every now and then when we measure spin in the direction $\alpha$ on the first particle, the second particle’s spin suddenly “flips”, then the transition from (1.1) to (1.2) is not justified, because what we measure does not reflect exactly what was objectively there before. On the other hand, if we accept locality, and if we add that the local undisturbed measurement always reveals the value objectively possessed by the system beforehand (which seems to be a necessary part of what we mean by a “faithful measurement” and “objectively possessed value”), then the transition from (1.1) to (1.2) is just a matter of a statistically valid inference from a smaller random sample to the whole ensemble. In other words, we can make the statistical error associated with the derivation of (1.2) arbitrarily small by selecting greater and greater numbers of measured particles. In conclusion, the Bell argument seems to lead to the following implication:

(Bell$_1$) Realism + Locality $\Rightarrow$ Bell’s inequality

Knowing that Bell’s inequality is violated by quantum-mechanical predictions, we can rewrite (Bell$_1$) as follows

(Bell$_2$) Realism + Locality $\Rightarrow$ Not-Quantum Mechanics

or, equivalently

(Bell$_2$) Quantum Mechanics $\Rightarrow$ Anti-Realism or Non-Locality

The above implication expresses the common wisdom regarding the lesson taught by Bell’s result. It is widely accepted, especially in popular presentations written by physicists, that Bell’s theorem leaves us with two options: either to abandon realism, contrary to Einstein and following
Bohr, or to face non-locality. Confronted with such a choice, physicists usually opt for the first of two evils, sometimes reluctantly adding that a small group of dissenters (with David Bohm in the forefront) decided to follow the second path, rescuing realism at the price of introducing mysterious non-local interactions (for example electrons that “know” instantaneously what is going on in remote parts of the universe). However, surprisingly many philosophers of physics claim that this popular view is deeply incorrect. Quoting Bell himself, who sympathized with the realistic point of view, they say that in fact Bell’s proven contradiction with the quantum-mechanical formalism should be diffused by sacrificing locality, no matter what stance one takes regarding the issue of realism (see Maudlin 1994, p. 19; Hawthorn & Silberstein 1995, p. 100; Bell 1987a; 1987c, p. 150). This can be bewildering when we have the result like (Bell$_2$) above; however, when we couple it with the previously accepted (EPR) things begin to clarify a bit. As we remember, according to (EPR) anti-realism implies non-locality. This means that actually the dilemma given in (Bell$_2$) is illusory. Combining together (EPR) with (Bell$_2$), we obtain, with the help of elementary propositional logic, the following implication:

$$(\text{Bell}+\text{EPR}) \quad \text{Quantum Mechanics} \Rightarrow \text{Non-Locality}$$

In other words, no matter what you decide about realism, locality has to go. If you opt for realism, the Bell result by itself shows that you have to accept non-locality. And if you reject realism in a futile attempt to rescue locality, the EPR argument moves in and forces you to abandon locality anyway. So it looks like we have eliminated the problem of realism altogether.

Accepting the above conclusion (at least as long as we accept the validity of both EPR and Bell arguments, of which more comes later), I don’t think that it exhausts all that can be said regarding the relation between locality and realism in the context of quantum mechanics. True, it looks like locality in one or another form has to be abandoned, but a question remains: What exactly is the nature of the non-locality which is imposed on

$^6$ Bell himself presented the conclusion of his theorem in the following way: “In a theory in which parameters are added to quantum mechanics to determine the results of individual measurements, without changing the statistical predictions, there must be a mechanism whereby the setting of one measuring device can influence the reading of another instrument, however remote” (Bell 1964, p. 20).
us by the combined forces of the EPR and Bell theorems? And I am going to argue that when we look more closely at the type of non-locality implied by the EPR result and by the Bell theorem, we reach the conclusion that the decision to choose or abandon realism can make a difference after all. But first we have to notice that the non-locality referred to in (EPR) differs significantly from the non-locality implied by the Bell result. Let us first have a look at the EPR case. Einstein claims that when we accept a non-realistic interpretation of quantum mechanics, we have to admit the existence of non-local interactions between distant elements of an entangled system. What kind of non-local influence is it? Remember that the crux of Einstein’s argument is that by measuring a given parameter of one particle, we can learn about the value of this parameter for a distant particle. If we agree that before the measurement the value of the observable in question was not determined, then it means that our measurement was capable of changing the state of the distant particle from an indeterminate to a determinate one. Hence we can generally characterize the type of non-locality in question as follows:

\[(\text{EPR-NonLoc})\quad \text{A distant measurement can instantaneously alter the state of a particle } p \text{ from being undetermined with respect to the value of a given parameter } A \text{ to being characterized by a precise value.}\]

Alternatively, we can formulate the above version of non-locality as admitting that a distant measurement can create an objective element of physical reality that was previously absent (or, alternatively, that was present only “potentially”). The word “distant” is to be understood as denoting the relativistically invariant space-like separation between the act of measurement and the appearance of the precise value on the second particle. As it should be clear, we characterize the type of non-local influences by identifying two elements: the kind of event which can exert non-local influence (“the cause”) and the kind of physical change brought about by the non-local influence in question (“the effect”). In the EPR case, the cause is an act of measurement, whereas the effect is a transition from being undetermined to possessing a determinate value.

Let us now turn to the Bell case. In order for the argument to go through, we had to make sure that a distant measurement cannot change a possessed value which is already there. This was the case, because the Bell
argument works under the assumption of realism. Hence, the non-locality occurring in the (Bell$_2$) formulation should be interpreted as follows:

(Bell-NonLoc) A distant measurement can instantaneously alter the state of a particle $p$ from possessing one value of a given parameter $A$ to possessing a different one.$^7$

Speaking loosely, Bell-nonlocality allows for changing elements of physical reality, and not just picking them out of the great number of initial possibilities. Now let us look at the consequences of the distinction we have just made. First, we have to observe that by combining (EPR) with (Bell$_2$) we now obtain the following dilemma:

(Bell+EPR') Quantum Mechanics $\Rightarrow$ EPR-NonLocality or Bell-NonLocality

Although this formula doesn’t mention the viewpoint of realism, it appears that realism plays an important part in our decision as to which non-locality to choose. If we reject realism, then in order to rescue the quantum mechanical formalism we have the option of only accepting the EPR-type non-locality, and rejecting the Bell-type. However, when we opt for realism, we have to accept both types of non-locality—the Bell-type because

$^7$ The idea of distinguishing different versions of the locality condition in the contexts of the EPR argument and Bell’s theorem is not a new one. M. Redhead in (1987) formulates one version of the locality condition LOC$_1$ as “An unsharp value for an observable cannot be changed into a sharp value by measurements performed ‘at a distance’” (p. 77) which is very similar to my EPR-Loc, and locality condition LOC$_3$ as “A sharp value for an observable cannot be changed into another sharp value by altering the setting of a remote piece of apparatus” (p. 82), closely resembling my Bell-Loc. A. Fine, in turn, speaks directly about Bell-locality versus Einstein-locality; however his way of characterizing them differs somehow from the definitions I am proposing (Fine 1986, pp. 59-60). For Fine the main difference between these two types of locality lies in the fact that according to Einstein locality excludes the possibility of influencing the “real” physical state of a distant system, while in the context of Bell’s theorem the locality implies only the impossibility of changing outcomes of measurements for quantum observables from a distance. Using his distinction Fine argues next that Bell’s result does not directly affect Einstein’s locality, as it may be maintained that quantum observables do not represent real physical states of systems. It seems to me that this is a pretty desperate way of defending Einstein, implying that quantum-mechanical “measurements” actually don’t measure anything real.
of the dilemma contained in (Bell$_2$), and the EPR-type because of the implication (EPR). In conclusion, it looks like rejecting or accepting realism makes a difference to the issue of locality. This observation is all the stronger when we note that EPR-nonlocality seems to be weaker than Bell-nonlocality. In other words, the departure from the classical, local world is probably less radical in case of the acceptance of the former than the latter. EPR-nonlocality allows the determining of a parameter when it has no prior value; once this value is defined, however, EPR-nonlocality doesn’t by itself allow changing it. To put it differently, EPR-nonlocality leaves room for a limited locality principle, in which it is maintained that no distant influence can instantaneously change a definite, existent element of reality. Using philosophical terminology we may say that non-local interactions can only transform potentiality into actuality, but not one actuality into another one.8 On the other hand, the Bell-type non-locality goes further in violating our ontological intuitions: an objective, possessed characteristic of a physical system can be changed from far away. Later on we will see that this off-hand assessment can be supported by a counterfactual analysis of both the Bell and EPR versions of non-locality. If the above argument is correct, there may be a philosophical basis for rejecting realism and, thanks to that, admitting only a minimal version of non-locality.

1.4 GENERALIZED BELL’S THEOREM

The above conclusion from the joint EPR and Bell arguments does not exclude a possibility that there may be other arguments that directly show strong non-locality of quantum mechanics, with no reference to the realist assumption whatsoever. One such argument was actually proposed and analyzed by Bell himself (Bell 1987b; 1987c).9 His goal was to produce the most general inference possible that would proceed from some version of the locality principle without any additional, philosophically “suspicious” premises, such as the principle of realism (in Bell’s terminology “determinism”)10, the condition of perfect correlation,

8 Shimony uses the same terminology in (1986, p. 153).
9 A particularly elegant exposition of this version of Bell’s theorem, with an extensive discussion of its consequences for the issue of non-locality, and containing a section illustrating experimental methods of its verification, can be found in (Shimony 1990).
10 The common, yet philosophically peculiar practice of using interchangeably the terms “determinism” and “realism” in the context of quantum mechanics can be
or even the assumption that there are some objective and physically separated particles involved. The conclusion of this inference was supposed to be a result (in the form of an inequality) that violated quantum-mechanical predictions. If successful, this argument would show unambiguously that the principle of locality is untenable, as the assumption of realism is actually irrelevant to the derivation of the contradiction, and that, therefore, one is free to accept or reject realism according to one’s own ontological preferences.

![Figure 1.1 Spatiotemporal layout of measurements in the generalized Bell theorem.](image)

The initial assumptions of the argument in question are very parsimonious indeed. We have to presuppose only that there are two spatially separated measuring devices, capable of recording outcomes of measurements for two sets of observables, each set associated with one measuring device. Let us denote by \( A \) and \( B \) the two particular observables that can be measured, respectively, by our two devices, and let us use lower-case let-

explained by pointing out that the apparent indeterministic character of quantum mechanics reveals itself only in measurements. As it is well known, quantum mechanics is a deterministic theory when it comes to the description of the state evolution of a system which is not subject to measurement, meaning that the initial state at time \( t \) uniquely determines all later states, given the Hamiltonian governing all existing physical interactions. However, this determinism applies only to states understood as probability distributions. When a measurement which aims to reveal a precise value of a given parameter is performed, the transition from the initial state to the final state is believed to be stochastic. If, in spite of this, one insists that the value revealed in the measurement should have characterized the system beforehand (realism), then one is committed to the view that the measurement-induced transition is actually purely deterministic. (To avoid possible confusion, some authors use the term “determinateness” instead of “determinism”—cf Maudlin 2003, pp. 470-472).
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

ters $a$ and $b$ to indicate the results of the measurements (we assume that each measurement can yield one of two numerical values $+1$ or $-1$ as its outcome). The quantum system that is about to be subjected to this double measurement has to be created at a certain moment, lying in the absolute past of both acts of measurement (in other words, it has to be located in the intersection of the two backward light cones, each defined by the appropriate measurement) (see Fig. 1.1). Now let us assume that the letter $\lambda$ represents a complete description of the physical state of the quantum system at the moment of its creation. No specific hypothesis as to what this complete description can look like is necessary here. In fact, the generality of the result we are about to present relies on not deciding what kind of “hidden variables” the parameter $\lambda$ describes. In particular, $\lambda$ can be seen as determining completely and accurately outcomes of all of the measurements in question (so-called “deterministic hidden variables”), or only the objective probabilities of particular outcomes (“stochastic hidden variables”). One example of such a description $\lambda$ can simply be a standard quantum-mechanical state as given with the help of the wave function $\Psi$. In that case $\lambda$ would be obviously stochastic. Without deciding the particular nature of the description $\lambda$, we will nevertheless assume that $\lambda$ is complete, i.e. that there is nothing more in the physical situation of the common past that is relevant to the future outcomes of experiments.

The single mathematically formulated assumption on which the entire generalized argument hinges is the following relation concerning the probabilities of the measurements outcomes:

$$(F) \quad P(a, b| A, B, \lambda) = P(a| A, \lambda) \quad P(b| B, \lambda)$$

We will call this equation the “factorization assumption”, although it is known in the literature under various names. It specifies that the probability of the joint results $a, b$ of measurements $A$ and $B$ for a system in the initial state $\lambda$ can be presented as the product of two independent probabilities: that the result of the first measurement $A$ is $a$, and the second $B$ is $b$. Equation (F) evokes the well-known definition of stochastically indepen-

---

11 The exact nature of the conditional probabilities used in this expression will be discussed later.

12 It was A. Fine who coined the term “factorizability” to refer to condition (F) (Fine, 1981). J. Jarrett in turn calls it “strong locality” (Jarrett 1984, 1989). Some authors use the longer but obviously uncontroversial term “conditional stochastic independence”.

Unauthenticated
Download Date | 5/29/19 1:41 AM
ent events, and therefore can be seen as expressing the idea of the nonexistence of causal influences between the two measurements in question; however, we will have to return to the question of its proper interpretation later. For the moment let us examine where this equation leads us. It appears that (F) itself implies mathematically a simple restriction on the expectation values of the products of the two measurement outcomes. Suppose that we have chosen two observables (two “settings”) $A$ and $A'$ for the left-hand side apparatus, and two observables $B, B'$ for the right-hand side. Let $E(A, B)$ denote the expectation value (the average) of the product of results for both observables. This expectation value is calculated as the sum

$$E(A, B) = \sum_{i,j} P(a_i, b_j | A, B) a_i b_j$$

where the probabilities $P(a_i, b_j | A, B)$ are given as the averages over all possible values of the parameter $\lambda$:

$$P(a, b | A, B) = \int P(a, b | A, B, \lambda) \rho(\lambda) d\lambda$$

With the help of the factorization assumption (F), it is a matter of a relatively simple derivation to arrive at the following inequality (known as the Clauser-Holt-Shimony-Horne inequality)\(^{13}\):

$$(\text{CHSH}) \quad |E(A, B) + E(A, B') + E(A', B) - E(A', B')| \leq 2$$

As was the case with the original Bell inequality, it turns out that there are observables $A, A', B, B'$ for which quantum-mechanical predictions violate the above equation. In particular, we can choose suitable components of spin such that the right-hand side of the above inequality will have the numerical value of $2\frac{1}{2}$, which is a significant departure from the predicted value of less or equal to 2. So the contradiction with the quantum-mechanical formalism has been achieved.

The above result can be significant for the purpose of proving the non-locality of quantum mechanics only if we are able to give a clear, unambi-

\(^{13}\) This inequality can be derived in a more standard way, using assumptions of locality and realism, exactly as in the original Bell theorem. See (Clauser et al. 1969; Redhead 1987, pp. 82-84) and sec. 3.1 in this book for more details.
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

The ambiguous interpretation of the condition (F) and its connections with the issue of non-locality. Unfortunately, this is not an easy matter. Let us start first with the original derivation of factorization proposed by Bell. Bell claims that (F) is a consequence of another ontological principle, which he dubs “the principle of local causality” (Bell 1987b, p. 54). Its non-mathematical expression can be presented as follows: all causes of a particular event must lie in its backward (past) light cone (see Maudlin 1994, p. 90). Due to the notorious ambiguity of the word “cause”, the aforementioned principle has to be given a more precise formulation. The proposal here is to use once again the notion of the conditional probability. Suppose that we have a particular event $E$, and let us denote by $\Gamma$ the complete physical description of the entire past light cone of $E$. Then, once $\Gamma$ is assumed to be known, no additional information can be relevant to the probability of the occurrence of $E$—this can be expressed as:

(LC) \[ P(E | \Gamma, \Delta) = P(E | \Gamma) \]

where $\Delta$ denotes any additional information about physical reality. $\Delta$ can for instance include the complete description of a space-time region not overlapping with the past light cone of $E$. We should arguably exclude the possibility that $\Delta$ refers to the absolute future of event $E$, for conditionalizing on future, yet unrealized events may change the initial probability without violating the idea of non-locality (for example, taking into account the outcomes of future measurements could dramatically change my evaluations of the current probability distribution over possible values).\(^{14}\) So, let us assume that $\Delta$ can refer only to the regions space-like separated from event $E$.

The condition (LC) obviously does not exclude the possibility that there may be observable statistical correlations between space-like separated systems. In other words, we can change our estimation of the probability for a given event on receiving information about distant systems. But this may happen only if we don’t have the complete knowledge of the causal past of the event in question.\(^{15}\) In other words, the only admissible reason

\(^{14}\) To put it more succinctly: if we agreed that the symbol $\Delta$ is to be interpreted as referring to future events as well, then the only way of satisfying (LC) would be by postulating deterministic hidden variables which would imply that $P(E | \Gamma)$ equals either 0 or 1.

\(^{15}\) To illustrate this point, we may use Jarrett’s example (1989, p. 72). When we estimate the probability that a particular person will have a heart attack, the
for such pseudo non-local correlations may be that they are due to a common cause in the absolute past. Once all the relevant facts from the past are taken into account, the space-like separated events should become statistically “screened-off”.\footnote{Hans Reichenbach (1956) was the first to introduce the notion of the “common cause” using the screening-off condition. See also (van Fraassen 1982, Bigaj 2003a).} In connection with this point we can raise the question about the nature of the probability function used in (LC): should we interpret it subjectively, objectively, as the relative frequency, or maybe in an entirely different fashion? It seems natural that if we conditionalize on the complete physical state of the system in the absolute past of \( E \), then we do not have to restrict our interpretation of probability to a subjective one. Hence, we will accept that the probabilities in (LC) are to be interpreted objectively, as the only and ultimate description available which is relevant to the occurrence of event \( E \). Subjective probabilities reflect our ignorance, but probabilities conditional on complete knowledge should reflect no less than the objective “propensities” or tendencies existing in physical reality.\footnote{This agrees with the standard way of treating probabilities in quantum mechanics (see e.g. Jarrett 1985, p. 586, footnote 6). However, we have to admit that probability interpreted as a dispositional property creates many conceptual difficulties, as presented in (Sklar 1970 and 1979). For an excellent survey of other interpretations of probability, see a chapter in the same author’s book (Sklar 1993, pp. 96-120).}

Now, we have to make sure that the factorization condition (F) is derivable from (LC). Let us start with the left-hand side of the equation (F), which can be presented equivalently as

\[
P(a, b \mid A, B, \lambda) = P(a \mid A, B, b, \lambda) \cdot P(b \mid A, B, \lambda)
\]

(by the rules of the probability calculus, including the definition of conditional probability). We would like to take out the conditionalization on \( B \) and \( b \) from the first factor in the product, as well as the conditionalization on \( A \) from the second one. Can we do this on the basis of the local causality principle (LC)? In order to appeal directly to (LC) we have to assume

\[\text{information that the person regularly buys jogging shoes may influence our judgment. However, the negative statistical correlation in question is not due to a causal connection between buying a pair of jogging shoes and avoiding a heart attack, but is a result of the existence of a common factor, namely regular exercising, which at the same time causally lowers the probability of a heart attack and is statistically positively correlated with possessing athletic equipment.}\]
that the complete description $\Gamma$ of the absolute past of the outcome $a$ is
given by totally specifying the common past of the two measurements (para-
meter $\lambda$) and by specifying the setting of the measurement $A$: $\Gamma = \lambda + A$.
Maudlin in (1994, p. 91) seems to agree that we can make use of (LC) only
if we assume that the sole physical event statistically relevant to the out-
come and located in the part of the backward light cone of the
measurement outside the common past is the choice of the setting of the
first detector (observable $A$). Once we agree on that, we can eliminate
factors $B$ and $b$ from $P(a \mid A, B, b, \lambda)$. Obviously, by symmetry the same
applies to the second measurement and outcome $b$. However, this seem-
ingly insignificant detail can in fact ruin the entire generalized argument
against locality, as an ardent proponent of the local causality principle (LC)
may claim that the resulting inequality (CHSH) contravening quantum me-
chanics is not due to the failure of (LC), but due to the above-mentioned
minor assumption. After all, it is quite reasonable (and compatible with the
relativity) to expect that physical processes that take place in the proximity
of the measurement may affect its outcome, even if they do not belong to
the common past of both measurements. If that is the case, then (LC) can
be applied only when we take as our complete description $\Gamma$ not only $\lambda + A$, but also all possibly relevant states of affairs localized in the absolute
past of measurement $A$ but outside $\lambda$. However, the expression $P(a \mid A, B,
b, \lambda)$ conditionalizes only on $A$ and $\lambda$, so (LC) is not directly applicable,
and therefore the derivation of the factorization condition (F) fails.

There is no question that the factorization principle (F) by itself repre-
sents an intuition bordering on the locality of causal influences, and the
CHSH inequality shows that this intuition has to be abandoned. But, in the
light of the above-mentioned difficulties associated with an attempt to de-
rive (F) from the more fundamental principle (LC), the question about the
proper interpretation of (F) becomes more pressing. Precisely what kind of
non-local influences has to be admitted? As we recall, with the Bell and
EPR arguments we had a choice between the weaker EPR-nonlocality and
the stronger Bell-nonlocality. Now it appears that it is possible to analyze
the principle (F) into similar components. In what follows we will review
and discuss the famous attempt to distinguish two elements in the general
factorization condition developed by Jon Jarrett.

It is easy to prove that the factorization is equivalent to the conjunction
of the two following principles:

1. \[ P(a \mid A, B, b, \lambda) = P(a \mid A, \lambda) \]  
2. \[ P(a \mid A, B, b, \lambda) = P(a \mid b, \lambda) \]
Obviously, when we combine (I) and (II) we can take out all conditionalization on distant measurements and outcomes without changing the numerical values of the probabilities: \( P(a \mid A, B, \lambda) = P(a \mid A, \lambda) \) and \( P(b \mid A, B, \lambda) = P(b \mid B, \lambda) \). Hence, the factorization (F) follows immediately. But the crucial thing is to analyze the physical meaning of these two principles. The most commonly used nomenclature already reveals the meaning typically associated with those conditions: condition (I) is namely referred to as “parameter independence”, and (II) as “outcome independence”.\(^{18}\) Parameter independence supposedly expresses the idea that the physical state of one system (encapsulated in the objective probabilities of the possible measurement outcomes) cannot be influenced by the choice of an observable (“parameter”) to be measured for the other, distant system. The principle (II), on the other hand, seemingly amounts to the assertion that the outcome of the distant measurement cannot influence the physical state of the local system. Under this interpretation, both principles (I) and (II) represent some instances of the general locality principle. Surprisingly, however, this is not quite the way Jarrett interpreted them originally. According to his terminology, the first condition bears the name of “the locality condition”, but he calls the second “completeness”. Jarrett’s point is that only violation of the principle (I) poses a threat to the fundamental principles of the relativity, as it allows for the possibility of the existence of superluminal signaling. The violation of completeness, on the other hand, does not allow for any signaling, for we have no control over what outcome will be revealed by a particular measurement.

The motivation behind Jarrett’s choice of the term “completeness” may be that when (II) is violated, it means that the separate descriptions of the two components of the system in terms of outcome probabilities are going to be incomplete, as they will not contain any information about the mutual correlations between the distant outcomes. This can be illustrated with the help of the standard case of two electrons in the singlet spin state. Assuming, for the sake of argument, that the hidden variable \( \lambda \) contains only the

\(^{18}\) This terminology was introduced by A. Shimony in his (1986).
usual quantum-mechanical state, we can notice that the probability functions \( P(a \mid A, \lambda) \) and \( P(b \mid B, \lambda) \) do not contain the entire available information about the system, for they totally ignore the fact that the outcome \( a \) of the measurement of \( A \) will be always correlated with the opposite outcome \( b \) of the measurement of \( B \).

These above remarks may explain the motivation behind the terminology proposed by Jarrett, but his terminology seems to suggest that it is possible to avert the inevitable demise of the locality principle at the price of rejecting “completeness”.\(^{19}\) This may look (intentionally or not) like a repetition of the EPR argument, but this impression is thoroughly misguided. Jarrett’s completeness has in fact very little in common with Einstein’s one, in spite of his own remarks in this matter (see Jarrett 1989, p. 73; Hughes p. 243). Moreover, no terminology can obscure the fact that the denial of the principle (II) entails some ontological non-locality. If (II) is false, it means that the complete state of the absolute past of the measurement does not fully determine the probability distribution of the possible outcomes but that this objective probability distribution depends ultimately (and not through a common cause) on some physical facts space-like separated from the measurement itself. Hence, we have to agree with the common view that both principles (I) and (II) are relevant to the issue of locality, although it may be argued that the negation of (I) constitutes a greater departure from the idea of locality than the negation of (II).

1.5 OUTCOME-LOCALITY AND PARAMETER-LOCALITY

Tim Maudlin seems to agree with the view that both principles (I) and (II) involve ideas that bear on the issue of locality. However, he is generally suspicious as to the correctness of their standard interpretations. His main argument against treating (I) as expressing “parameter independence”, and (II) “outcome independence”, as well as—as I see it—generally against splitting the factorization principle (F) into two components, is that the combination (I) and (II) does not offer a unique way of representing (F). Maudlin points out that we can formulate two alternative principles, different from (I) and (II), whose conjunction is logically equivalent to (F), and

---

\(^{19}\) This is precisely the conclusion in Jarrett’s 1984 article (p. 585). In his later paper (1989) however, while still maintaining that only completeness should be abandoned, he admits that “incompleteness [...] appears to represent a connectedness of some sort between spatially distant events, which nevertheless does not directly contradict relativity” (p. 77)
which supposedly can express the ideas of “parameter independence” and “outcome independence” equally well as (I) and (II). Here is what they look like according to Maudlin (1984, p. 95):

\[
\begin{align*}
(I') & \quad P(a \mid A, B, b, \lambda) = P(a \mid A, b, \lambda) \\
& \quad P(b \mid A, B, a, \lambda) = P(b \mid B, a, \lambda) \\
(II') & \quad P(a \mid A, b, \lambda) = P(a \mid A, \lambda) \\
& \quad P(b \mid B, a, \lambda) = P(b \mid B, \lambda)
\end{align*}
\]

To see that with equations (I’) and (II’) we can get (F), we should observe that they allow us to take out the conditionalization on \(A\) from \(P(b \mid A, B, \lambda)\). According to the law of total probability we have

\[
P(b \mid A, B, \lambda) = \sum_i P(b \mid A, B, a_i, \lambda)P(a_i \mid A, B, \lambda)
\]

(where \(a_i\)'s are different outcomes of the measurement of \(A\), and therefore constitute a set of mutually exclusive and jointly exhaustive events). But using (I’) and (II’) we can replace \(P(b \mid A, B, a_i, \lambda)\) with \(P(b \mid B, \lambda)\), and because the sum \(\Sigma_i P(a_i \mid A, B, \lambda)\) equals 1, we finally get \(P(b \mid A, B, \lambda) = P(b \mid B, \lambda)\). Once we have this result, we can see that (F) is easily derivable from the formula \(P(a, b \mid A, B, \lambda) = P(a \mid A, B, b, \lambda) P(b \mid A, B, \lambda)\).

Having proven that (I’) and (II’) imply (F), we have now to turn to the issue of their proper “physical” interpretation. Maudlin maintains that (I’) deserves to be named “parameter independence”, because it allows us to eliminate conditionalization on the distant measurement, and (II’) can be seen as expressing the outcome independence, because it shows that conditionalization on the distant outcome is irrelevant for the probability of the local outcome. However, (I’) is not equivalent to (I), and (II’) is not equivalent to (II). As Maudlin points out, standard quantum mechanics (under the supposition that \(\lambda\) contains only available quantum-mechanical description of the state) violates (I’) but obeys (I), and violates (II) but obeys (II’). Hence if Maudlin is right, our basic ideas of what constitutes outcome independence and parameter independence are fundamentally ambiguous.

I don’t agree with the aforementioned conclusion for quite straightforward reasons. It seems to me quite clear that equations (I’) and (II’) do not represent any physically significant conditions which could be interpreted...
in an analogous way to the way principles (I) and (II) are commonly interpreted. To see this, let us first focus on equations (II'), supposedly expressing the idea of independence on the distant outcome. Equations (II') state that the information about the outcome of an unspecified distant measurement is statistically irrelevant to the probability of a given outcome of the local measurement. In other words, if we considered all cases of combined measurements such that the first apparatus is set to measure $A$, but the second apparatus records a given value $b$ no matter what the setting is (what observable $B$ is being measured), then the outcome $a$ will occur in the same fraction of cases as in the entire population of cases. But now the question is: what is the physical significance of the outcome of an unspecified measurement? To say that the second apparatus registered the value $-1$, without mentioning what particular observable was being measured carries no meaningful physical information. The very notion of an outcome is meaningful only when it is coupled with the specification of what it is an outcome of. Hence if we look for a clear-cut, physically meaningful description of a distant factor which may be responsible for an instantaneous, non-local change, we should not use the notion of the outcome of an unknown measurement. Such outcomes simply don’t constitute legitimate physical events.

Similar objections can be raised against principle (I'). Taken at face value, (I') says that conditionalization on a distant measurement (setting) is statistically irrelevant, provided that the result of this measurement is fixed. In other words, we are comparing cases in which a particular measurement is performed and the result is such and such, with cases in which any possible measurement is taken, but the result remains numerically the same. But, once again this means that we assign physical significance to “dangling” outcomes; outcomes detached from their appropriate measuring procedures. After all, that is exactly what happens on the right-hand side of equations (I'): we are asked to entertain the probability of a given outcome under the supposition that the distant result is known, but without specifying the procedure leading to this result. It is, therefore, highly doubtful that the principles (I') and (II') could express any clear idea of locality with respect to distant outcomes and distant settings.

However, this is not to say that conditions (I) and (II) are crystal-clear. On the contrary—Maudlin’s example teaches us that we should be extremely careful in attaching physical or philosophical meaning to otherwise
meaningful mathematical formulas. And it appears that there is certain amount of confusion associated with the proper interpretation of the statistical independence conditions given in (I) and (II). For example, Maudlin himself gives the following characteristics of (I) and (II), apparently reporting them as a commonly accepted interpretation: “parameter independence holding if the act of setting the distant device has no distant causal role; outcome independence holding if the measurement event itself has no distant effects.” (Maudlin 1994, p. 95) However, this is somewhat surprising. Usually, when we talk about outcome independence, we have in mind the fact that somehow the last phase of the measurement process, i.e. that the recording or determining in any way of a particular outcome has no bearing on the distant state of affairs. And yet Maudlin speaks about the whole process of measurement. But, if this is correct, why call this “outcome independence”, and not “measurement independence”? The proposed characteristic of parameter independence is no less confusing, in spite of being quite persistent in the literature. What do we mean by “setting the distant device” or, as it is sometimes put, “preparing the device”? The usual interpretation is that this setting amounts, for example, to the spatial orientation of a Stern-Gerlach magnet, so that it is ready to measure the spin of an incoming electron in a suitable direction. But let us suppose that for some reason the right-hand electron didn’t get to the measuring device (it was absorbed by an atom, or scattered away by some passing particle). Would we count such a case in order to empirically calculate the probability given in the left-hand side of equations (I)? It seems that in order to calculate $P(a \mid A, B, \lambda)$, we have to make sure not only that the distant apparatus was ready to measure $B$, but that it actually measured it (that the electron was fed in, and got reflected either up or down, although we don’t know which).

It appears then that we have to devote more effort to the task of properly interpreting the formulas given in (I) and (II) than it has been thus far. We have to make precise the meaning associated with the different conditionalsizations of the probabilities on various parameters that we have taken for

---

21 It speaks volumes that other authors display a certain amount of hesitation when characterizing physically both conditions (I) and (II). For example R.I.G. Hughes interprets explicitly symbol $A$ as “A-measurement is performed on an electron” (Hughes 1989, p. 244); however, when he explains the meaning of the conditionalization in (I) he uses the notorious phrase “given certain settings of the measurement apparatuses” (ibid.). One cannot help but wonder whether authors like Hughes intentionally blur the distinction between preparing measuring devices and actually performing the measurement.
Why Does the Quantum World Have to Be Non-Local?

Granted so far. We will start this clarification with a confession: the formulas which have been used throughout our current discussion have actually been oversimplified in comparison with Jarrett’s original article (1984). This oversimplification is quite common in the literature, nevertheless it can obscure certain important issues and cause misunderstandings. In order to have a firm grasp on expressions like $P(a, b|A, B, \lambda)$, used freely in all discussions on the Bell theorem, we have to precisely define our probability space and probability function first. As we remember, our experimental setup consists of two measuring devices, each of which may be prepared in different settings. Let us denote by $A_i$ ($B_j$) all the possible settings of the measuring devices (left and right, respectively). To these we will have to add one more state of the devices: it is possible that a device is simply turned off—not prepared for measuring any observable (any component of spin). We will symbolize by $0^A$ and $0^B$ the states of the left-hand side and the right-hand side apparatuses in which they do not measure any observable. Hence the set of all states for the left-hand side device will be $M^A = \{0^A, A_i\}$ (and $M^B = \{0^B, B_j\}$ for the right-hand side apparatus).

Analogously, we will define the sets of possible outcomes. Each device can record two “normal” outcomes $+1$ and $-1$. To this we will once again add a non-standard “outcome” 0, associated with the non-measuring state of the apparatus. Hence the set of possible outcomes for the left-hand side device will be $O^A = \{+1^A, -1^A, 0^A\}$ (and $O^B = \{+1^B, -1^B, 0^B\}$ for the other system). Subsequently, we will omit superscripts whenever it doesn’t lead to ambiguities. Any theory that aspires to give a complete description of the physical reality in our example should produce probabilities of the following sort:

\[(P) \quad P(a, b|A_i, B_j, \lambda)\]

where $\lambda$, as usual, denotes the complete initial state of the system. The formula above uses the standard mathematical symbol for conditional probability, but the actual meaning of this conditionnalization has to be quite non-standard. In particular, we cannot resort to the formal definition of the conditional probability in order to explain the meaning of the probability function given above. The standard definition gives the following explication for $(P)$:
but the probabilities \( P(A_i, B_j \mid \lambda) \) cannot be calculated within the underlying theory. The decision to choose a particular setting of the device is made freely by an experimenter, and although in principle it is possible to select settings \( A_i \) and \( B_j \) randomly according to a previously defined statistical distribution, the resulting probabilities will be essentially extraneous with respect to the theory governing the behavior of the system in question (see Dickson 1998, p. 136 for a similar appraisal). For that reason it is better to interpret the conditionalization given in (P) as merely amounting to a parameterization of the probability function.\(^{22}\) In other words, the formula (P) should be actually read as describing a class of independent probability functions, each of which defines the probabilities of particular outcomes in a given experimental setting, as in the following notation:

\[
(P') \quad P_{\lambda}^{ij}(a, b)
\]

Remembering that only (P') provides the proper interpretation for the formulas (P), we will nevertheless continue using the conditionalization symbol (“the stroke”), as it is commonly done in literature. The domain on which functions (P') are defined consists obviously of four values: \{\langle +1, +1 \rangle, \langle +1, -1 \rangle, \langle -1, +1 \rangle, \langle -1, -1 \rangle\}. Now, as we remember, the formulation of both locality conditions (I) and (II) requires probabilities “conditionalized” only on one measurement: \( P(a \mid A_i, \lambda) \). How are we supposed to interpret these expressions? Let us start by presenting the standard interpretation of probabilities for only one selected outcome, given two settings of the measuring devices: \( P(a \mid A_i, B_j, \lambda) \). This function is obtained by simply summing the function (P) over all possible results of measurement \( B_j \):

\[
P(a \mid A_i, B_j, \lambda) = \sum_b P(a, b \mid A_i, B_j, \lambda)
\]

However, the question remains how we can “eliminate” the second conditionalization on measurement \( B_j \). The standard probabilistic equation

\[
P(a \mid A_i, \lambda) = \sum_j P(a \mid A_i, B_j, \lambda)P(B_j \mid a, A_i, \lambda)
\]

\(^{22}\) In the 1984 article Jarrett used no conditional probabilities. However, he adopted a terminology based on conditional probabilities in his later 1989 paper without much explanation.
won’t work for the purpose of explicating the expression $P(a| A_i, \lambda)$, because of meaningless probabilities $P(B_j| a, A_i, \lambda)$ of particular settings of the apparatus. In order to give the function $P(a| A_i, \lambda)$ an acceptable interpretation we have to make use of the extra state of the measuring devices, the non-measuring setting, and to stipulate that $P(a| A_i, \lambda)$ is meant to actually represent the following probability: $P(a| A_i, 0^B, \lambda)$, which in turn is identical to $P(a, 0| A_i, 0^B, \lambda)$. Hence, our class of probability functions has to be extended to include the following:

$$(P'') \quad P(a, 0| A_i, 0^B, \lambda)$$

Functions (P'') are defined on the following domains: $\{\langle +1, 0 \rangle, \langle -1, 0 \rangle\}$ and $\{\langle 0, +1 \rangle, \langle 0, -1 \rangle\}$, respectively. Their empirical meaning is quite straightforward – they numerically represent relative frequencies of a particular outcome of one measurement in case the other measuring device is simply turned off.

With these preparatory stipulations in hand, we can now turn to the task of the reinterpretation of the two components (I) and (II) of the factorization condition. The full version of the so-called parameter independence condition can now be expressed as follows:

$$(PI) \quad P(a, 0| A_i, 0^B, \lambda) = \sum_b P(a, b| A_i, B_j, \lambda)$$

The formula expressing the outcome independence will be a little more complicated. In order to avoid possible confusion, let us proceed in steps. The original version of the principle was:

$$(II) \quad P(a| A_i, B_j, b, \lambda) = P(a| A_i, B_j, \lambda)$$

The right-hand side of the equation has a suppressed summation over all the possible results of measurement $B_j$. The left-hand side in turn contains the new element on which the probability is conditional (namely the outcome $b$), and as it turns out it can be interpreted with the help of the usual definition of conditional probability (all elements in the definiens are meaningful). This brings us to the following reformulation:
Now, we can introduce the suppressed summation in the denominator of the left-hand side fraction. The resulting equation will look like this (compare Howard 1997, p. 127):

\[
(P) \quad \frac{P(a, b | A_i, B_j, \lambda)}{P(b | A_i, B_j, \lambda)} = \sum_b P(a, b | A_i, B_j, \lambda)
\]

Finally, we can give the full version of the factorization condition. You can easily verify that its precise formulation should look like this:

\[
(F) \quad P(a, b | A_i, B_j, \lambda) = P(a, 0 | A_i, 0^B, \lambda) P(0, b | 0^A, B_j, \lambda)
\]

So we have achieved our first goal: we have presented all three conditions in a uniform language containing only well defined, empirically meaningful probabilistic functions of the kind presented in (P) and (P″). This should help us to gain a better understanding of the physical and philosophical meaning behind these conditions.

However, the task is not yet complete. The problem is that we would like to understand exactly what types of non-local influence are associated with the negation of each condition (PI) and (OI). The analysis of these influences should preferably be given, as was the case with Bell-locality vs. Einstein-locality, in terms of “what causes what” (what kind of action on the distant system triggers what changes in the local one). Let us start with parameter independence (or the “proper” locality, as Jarrett would put it). The left-hand side of the equation (PI) represents, most typically, the objective probability (“propensity”) of obtaining outcome \(a\) of measurement \(A_i\) when no measurement is performed on the other system (with the “null” outcome). The right-hand side is usually interpreted, as we already indicated several times, as the probability of the same outcome \(a\) given that the measurement of \(B_j\) is performed on the distant particle. Hence, the typical interpretation associated with the negation of (PI) is that by selecting a particular observable for measurement we can change the objective state of the distant system (represented by the probability function) as compared with the situation when no measurement is performed. But is this really the correct interpretation? Several objections against this standard elucidation can be put forward. To begin with, we already know that the probability
$P(a | A_i, B_j, \lambda)$ is not a probability ‘in its own right’, but, rather, is defined as the sum of two more basic probabilities: $P(a, +1 | A_i, B_j, \lambda) + P(a, -1 | A_i, B_j, \lambda)$. Hence, it can hardly be seen as an objective characterization of the state of the distant system in the case where the observable $B_j$ is chosen to be measured. A more accurate explication would be that $P(a | A_i, B_j, \lambda)$ can represent our subjective estimation of the probability of outcome $a$ when we know that the measurement $B_j$ has been performed, but we don’t know its outcome. In other words, $P(a | A_i, B_j, \lambda)$ would merely reflect our ignorance rather than the objective propensity determining the likelihood of obtaining outcome $a$. But if that is the case, then the above “ontological” interpretation of not-(PI) has to be reconsidered.

Let us illustrate this point with an artificially created example. Suppose that the “initial” probability of a particular outcome $a$ for the measurement $A_i$ when no measurement is performed on the other system (given all the available knowledge about the past of both systems, of course) equals $P(a, 0 | A_i, 0^B, \lambda) = \frac{1}{2}$. Moreover, let us assume that when the distant measurement of $B_j$ reveals $+1$ the probability of obtaining $a$ in the situation is given the numerical value

$$P(a | A_i, B_j, b = +1, \lambda) = \frac{3}{4},$$

and let’s similarly stipulate that, when $B_j$ yields $-1$,

$$P(a | A_i, B_j, b = -1, \lambda) = \frac{3}{8}.$$  

Finally, we will assign particular values to the conditional probabilities of obtaining the results $+1$ and $-1$ in the $B_j$ measurement:

$$P(b = +1 | A_i, B_j, \lambda) = \frac{1}{3} \text{ and } P(b = -1 | A_i, B_j, \lambda) = \frac{2}{3}$$

(these numbers obviously have to add up to one). With these numerical values of the probabilities in question, it is not difficult to verify that the probabilities $P(a, +1 | A_i, B_j, \lambda)$ and $P(a, -1 | A_i, B_j, \lambda)$ will both equal $\frac{1}{4}$, so $P(a | A_i, B_j, \lambda) = \frac{1}{2}$. Thus the equation (PI) remains satisfied, which may be interpreted as expressing the fact that the choice of a distant measurement has no influence on the local outcome. But let us consider the individual case of two particles, such that the first underwent the measurement of $A_i$ with the result $a$, and the second was subject to the measurement of $B_j$. Are
we justified in the claim that in this particular, singular case the state of the first particle just before its measurement is exactly the same as in the situation when no measurement is made on the other particle? After all, measurement $B_j$ has to reveal some precise outcome, even if we don’t know what it will be. And it may be pointed out that if the outcome of $B_j$ is $+1$ then the objective probability of revealing $a$ equals $\frac{3}{4}$, which is different than $\frac{1}{2}$; and when the outcome of $B_j$ is $-1$, the objective probability of $a$ takes the value $\frac{1}{8}$, which again differs from $\frac{1}{2}$. So no matter which is the case (and one of them has to be the case) the probability of the outcome $a$ for the measurement $A_i$ will be different from the probability given that no measurement is performed on the distant system. But doesn’t this amount to saying that the distant measurement is capable of changing the objective state of the local system? Yet this conclusion does not agree with our initial reading of the condition (PI). Hence the association of (PI) with the lack of influence between the act of measurement on one particle and the outcome of the other one seems to be unjustified.

Jarrett in his article (1984) has proposed a particular argument in favor of the standard interpretation of the condition (PI). His idea was to prove that when the condition (PI) is violated, it is in principle possible to send information regarding the distant choice of measurement setting to the other, spatiotemporally separated experimenter, thus violating the relativistic restrictions. However, his example involves a large number of correlated pairs of particles, all prepared in the same initial state $\lambda$. The information about the distant setting can be “decoded” only when sufficiently many independent measurements ($A_i$, $B_j$) have been made. In such a case the statistical distribution of result $a$ for measurement $A_i$ in the ensemble will reflect the probability $P(a|A_i, B_j, \lambda)$ (different from $P(a, 0|A_i, 0^B, \lambda)$ by assumption) because the impact of individual outcomes of $B_j$ will be statistically “filtered out”. This will allow the local experimenter to learn that the distant collaborator chose to measure $B_j$ rather than making no measurement at all. However, it is not at all clear whether this proves that the supposed non-local correlation takes place in each pair independently. My point is that in order to prove that a non-local influence between the choice of setting and the particular outcome obtains when (PI) is violated, we would have to devise a method which would allow us to transfer super-

---

23 Jarrett’s argument obviously proceeds under the assumption that it is possible to create quantum systems in a desired state $\lambda$. However, it may be the case that the laws of nature forbid this. In this case the direct violation of Einstein’s restriction on superluminal signaling would not threaten.
luminally the information about the distant setting for a single pair of correlated particles. And the only conceivable way of accomplishing this would be to make a sufficient number of identical copies of the left-hand side particle, and then to measure the relative frequency of obtaining a as the result of the measurement of observable $A_i$ within the prepared ensemble. However, in that way we would not obtain a measure of the required probability $P(a|A_i, B_j, \lambda)$ but rather one of the following: $P(a|A_i, B_j, +1, \lambda)$ or $P(a|A_i, B_j, -1, \lambda)$, depending on what the actual outcome of measurement $B_j$ was. Hence, in some cases it would be impossible to tell whether the other particle underwent measurement $B_j$ or not. This means that information transfer appears to be unattainable in the case when we have at our disposal only one pair of entangled particles.

This problem does not occur when we consider the second component of the factorization condition, namely outcome independence. Here it turns out that the violation of the principle (OI) amounts to the change of the outcome of one measurement having an instantaneous impact on the physical situation of the other system. To see this, let us first present the condition (OI) in a different, but equivalent form:

$$(OI') \quad P(a|A_i, B_j, b, \lambda) = P(a|A_i, B_j, b', \lambda) \text{ for all } b \neq b'$$

That (OI’) follows from the original formulation (II) should be seen as pretty obvious when we take into account that there is actually a suppressed universal quantification over all possible outcomes $b$ in (II). To see the reverse implication (OI’) $\Rightarrow$ (II) let us note that the right-hand side of (II) can be presented as

$$P(a|A_i, B_j, \lambda) = \sum_b P(a|A_i, B_j, b, \lambda)P(b|A_i, B_j, \lambda)$$

by the law of total probability (outcomes $b$ constitute a set of mutually exclusive and jointly exhaustive events). Using assumption (OI’) and the fact that $\sum_b P(b|A_i, B_j, \lambda) = 1$ we arrive at the required equation $P(a|A_i, B_j, \lambda) = P(a|A_i, B_j, b, \lambda)$. We can see now that the violation of (OI’) means that a change of the outcome in the measurement $B_j$ from $b$ to $b'$ changes the objective probability of obtaining outcome $a$ in the distant system. Or, using Jarrett’s approach, we can argue that it is possible to let the faraway experimenter know what the result of the local measurement $B_j$ was. In order to decipher the information about the outcome, the experimenter has to
make a sufficient number of copies of his particle and then measure the relative frequency of the occurrence of outcome \( a \). If this number approaches \( P(a | A_i, B_j, +1, \lambda) \), the experimenter learns that the other end of the apparatus recorded +1; if the number approximates \( P(a | A_i, B_j, -1, \lambda) \) (and by the negation of (OI′) this has to be different from \( P(a | A_i, B_j, +1, \lambda) \)), he knows that the result was −1.\(^{24}\)

To sum up, we have argued that the first of the two “locality” conditions considered here gives rise to some interpretive problems.\(^{25}\) This is not to say that the idea of separating the two types of possible non-local influence—one triggered by an experimenter’s choice of an observable to measure, and the other linked to the outcome received in the course of the measurement—is unreasonable. The only thing that is questionable is the connection between these ontological ideas and particular mathematical formulas, such as formula (PI). And, ultimately, if we agree that it is by no means clear what condition is expressed in (PI), then the philosophical lesson from the generalized Bell theorem becomes quite ambiguous. Hence, it may be argued that there is still a need for a more decisive result showing the untenability of the notion of locality in the context of quantum-mechanical phenomena. This is precisely the task taken up by Henry Stapp, among others, and which we will scrutinize at length later in this book. One idea of approaching such a task may be to express “parameter locality” in the language of counterfactual conditionals rather than in the probability calculus language; for example in the form of a sentence stating that if we had chosen a setting of the measuring apparatus different from the actual one, no physical change would have occurred in the distant system. If we were able to derive Bell’s inequality from such a condition coupled with quantum-mechanical predictions, without any explicit or implicit reference to the realist condition, then we would unambiguously

\(^{24}\) Jarrett does not see this possibility as violating the prohibition of superluminal signaling, because, as he argues, the outcome of the measurement \( B_j \) does not depend on the other experimenter’s will, so he cannot use the outcome dependence as a means for sending meaningful messages to his partner. Regardless of this practical setback, the outcome dependence definitely allows for superluminal exchange of physical information about a distant system, and therefore counts as a non-local influence.

\(^{25}\) Dickson (1998, pp. 134-139) is similarly skeptical about the philosophical significance of Jarrett’s analysis. However, he has no objections to treating the factorization condition (F) as a direct representation of the locality requirement. I, on the other hand, am inclined to look for more appealing and straightforward representations of the principle of local causality than the statistical condition (F).
show that quantum mechanics is non-local. However, we will argue throughout this book that this ambitious task has never been accomplished to a satisfactory degree, and there are even some grounds for thinking that it may simply be incapable of being accomplished.

1.5.1 Non-locality and non-separability
It can easily be verified that when the hidden variable $\lambda$ is interpreted as consisting only in the standard quantum-mechanical state, quantum theory implies that parameter independence (PI) is satisfied, but outcome independence (OI) turns out to be violated. This fact is commonly seen as indicating that in standard quantum theory there is no measurement-induced non-locality, while some sort of non-local influence between outcomes revealed in distant wings of the measuring apparatus is present. Yet this last conclusion has been questioned. Some authors argue that the violation of (OI) (or (II)) has nothing to do with the existence of superluminal causal links, but instead is a consequence of the failure of an altogether different classical intuition, i.e. the assumption of separability. In what follows we will take a closer look at this claim.

The most ardent proponent of this way of interpreting the failure of outcome independence is Don Howard, who claims on the basis of the available historical evidence that Einstein himself distinguished separability from locality as two different metaphysical principles.26 Roughly speaking, the principle of separability asserts that complex systems can be broken up into smaller components, each of which is endowed with its own physical properties, and such that the properties of the entire system are somehow “reducible” to or “supervenient” on the properties of its parts. In the context of quantum mechanics it is claimed that entangled systems do not satisfy the requirement of separability, as the state of the complex system is not a simple combination of states of the components. For instance, the single spin state of the system of two spin-$\frac{1}{2}$ particles is given as the following (up to the normalization constant) superposition of pure states: $\psi_{\text{singlet}} = \psi_1^+ \otimes \psi_2^- - \psi_1^- \otimes \psi_2^+$, where $\psi_i^j$ represents the pure state of the $i$th particle in which the value of this particle’s spin in a given direction equals $j$ (either + or −). It turns out that $\psi_{\text{singlet}}$ cannot be represented as a simple product of pure states of the two components $\psi_1 \otimes \psi_2$, which is sometimes interpreted as indicating that entangled particles do not possess

well-defined states of their own. On the other hand, each particle taken separately does have a well-defined mixed state, i.e. there exists a probability distribution over all possible results of spin-measurements in every direction, however this state cannot be represented with the help of a vector (a ray) in the Hilbert space, but rather as a density operator on the same space (or alternatively as a weighted sum of pure states—for mathematical details see Hughes 1989, chapter 5).

I think that at this point we should distinguish two separate although closely related problems. The first issue is whether it is generally justified to claim that quantum entangled systems violate the metaphysical principle of separability; the second is the more specific question whether the outcome independence condition can be seen as a legitimate explication of separability. In terms of the first issue, Michael Esfeld points out that although technically it makes sense to speak of separate states of the components of an entangled system, still these states taken together do not determine the total state of the system, for the important information about the correlations between outcomes is lost (Esfeld 2001, 2004). To use the singlet spin example, the two mixed states that reduce the state $\psi_{\text{singlet}}$ can be combined in an infinite number of ways to create entirely different complex states, of which $\psi_{\text{singlet}}$ is only one. However, one may wonder if this underdetermination of the complex state by the component states is by itself sufficient to conclude that the metaphysical principle of separability is violated here (and that instead we have to adopt some sort of quantum holism). For example Richard Healey in (2004) gives one possible characterization of separability in terms of the “property determination” condition:

Every qualitative intrinsic physical property and relation of a set of physical objects from any domain $D$ subject only to type $P$ processes supervenes on qualitative intrinsic physical properties and relations in the supervenience basis of their basic physical facts relative to $D$ and $P$ [italics mine]

It is important to notice that the above condition does not assert that intrinsic base properties alone should determine the properties of the compound system, but rather that base properties plus base relations should accomplish this determination. But if that’s the case, then the singlet spin state clearly satisfies the condition of separability, for the specification of each particle’s own mixed state plus the perfect anti-correlation relation between outcomes of spin-measurements uniquely determine that the state of the entire system is $\psi_{\text{singlet}}$. It has to be added that Howard as well as Esfeld
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

...adopts a stronger requirement of separability in which it is demanded that the joint state of entangled systems be wholly determined by their separate states (Healey calls this requirement “state separability”). But it is open to a debate whether this is a reasonable claim, given that there may be some non-local relations between elements of the system that do not enter into their separate states.

Nevertheless, Howard maintains that the statistical condition of outcome independence that we presented in the form of the equations \( P(a \mid A, B, b, \lambda) = P(a \mid A, B, \lambda) \) and \( P(b \mid A, B, a, \lambda) = P(b \mid A, B, \lambda) \) represents the condition of separability rather than locality. To show this, he rewrites the outcome independence condition in an equivalent form as \( P(a, b \mid A, B, \lambda) = P(a \mid A, B, \lambda) P(b \mid A, B, \lambda), \) and then argues that this is essentially the assumption of the factorizability of the joint state represented by \( P(a, b \mid A, B, \lambda) \) into the component “contextual” states \( P(a \mid A, B, \lambda) \) and \( P(b \mid A, B, \lambda), \) equivalent to the assumption that the joint state is a tensor product of the separate states (see Howard 1997, pp. 126-127). However, Howard’s claim is countered by Maudlin, who points out that it is possible to construct models of the EPR situation which violate outcome independence and which explicitly employ superluminal signals while arguably conforming to the separability requirement (Maudlin 1994, p. 98). The violation of outcome independence does not per se imply non-separability, as the probability distributions for each particles separately may be well-defined, although due to non-local interactions those distributions may not uniquely determine the joint distributions for the outcomes of the entire system.

Finally, we may voice more general concerns regarding the proposed holistic interpretation of quantum entangled systems. As Dickson emphasizes (1998, p. 156), it is unclear what is precisely gained by postulating that the two parts of the entangled system are in fact one unseparable whole possessing joint and irreducible states. As long as we continue to use the notions of localized measurements and outcomes, there will always be a problem of how to explain apparent correlations between spatially separated readings of the measuring apparatuses. The very notion of a measurement is local: it is always a result of interaction of the separated measuring device with whatever we take as the measured system. Unless we change radically our ontology of events, allowing for instance for the existence of events that are simultaneously localized in two space-like separated regions, the measurement-events are local and separated, and hence any correlations between them call for some explanation in terms of...
influences, even if we subscribe to the holistic vision of the quantum world.²⁷

1.6 A PROVISIONAL TAXONOMY OF QUANTUM NON-LOCALITIES

At the end of this opening chapter we will try to introduce some order into the untidy world of the multitude of non-local influences. Earlier, we suggested that the most intuitive way to characterize any type of causal correlation is in terms of “what causes what”: what type of action precipitates what effects. We now know that in the context of quantum entangled systems there are two basic candidates for a cause of non-local interactions: the choice of a particular observable to be measured on one component of an entangled system, and the revealed outcome of the measurement. As far as the potential effects of these factors are concerned, we have the following possibilities, mentioned already before. First of all, one can try to non-locally induce a transition from an “unsharp” state of the particle to a sharp one (i.e. from a state with no defined value for a given observable to a state in which the observable takes a precise value). Another type of non-local interaction would lead to a transition from one definite value to a different one. We should also allow for a possibility, albeit to my knowledge only a theoretical one, of changing the state of the distant particle from possessing a determined value to an undetermined one. The non-locally induced change may also be of a more subtle kind,

²⁷ It should be stressed that our brief survey of the problem of non-locality in quantum mechanics and various approaches to it is by no means exhaustive. Many attempts to solve the problem of non-locality stem from adopting particular non-standard interpretations of quantum mechanics. Let me mention one such attempt: in a recently proposed variant of the many worlds interpretation, known as the many minds interpretation, the problem of the non-local correlations between distant outcomes in the EPR situation disappears altogether. According to the many minds interpretation, the physical evolution of quantum systems is always deterministic, and measurements are no exceptions (hence, there are no wave collapses). However, upon interacting with the measuring device, the mind of the experimenter “splits” into an infinity of different minds, each associated with a particular possible outcome. Of course under this interpretation there is no problem of how to explain the correlation between outcomes obtained in space-like separated locations, as there are physically no definite outcomes. The correlations predicted by the theory come into being only when two experimenters communicate the results of their respective measurements, thus interacting physically and locally with one another (see Albert & Loewer 1988; Albert 1992)
expressed in terms of probabilities, or relative frequencies only. We would definitely call it non-locality, if it were possible to instantaneously change from a distance the objective probability of revealing particular values of an observable.

<table>
<thead>
<tr>
<th>Change from an undefined value of $B$ to a defined one</th>
<th>Choice of a measured observable $A$</th>
<th>Obtained outcome of $A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Loc\textsubscript{11}</td>
<td>Non-Loc\textsubscript{21}</td>
<td></td>
</tr>
<tr>
<td>Change from a defined value of $B$ to an undefined one</td>
<td>Non-Loc\textsubscript{12}</td>
<td>Non-Loc\textsubscript{22}</td>
</tr>
<tr>
<td>Change from one defined value of $B$ to a different one</td>
<td>Non-Loc\textsubscript{13}</td>
<td>Non-Loc\textsubscript{23}</td>
</tr>
<tr>
<td>Change from one probability measure on all values of $B$ to a different one</td>
<td>Non-Loc\textsubscript{14}</td>
<td>Non-Loc\textsubscript{24}</td>
</tr>
</tbody>
</table>

Tab. 1.1. Classification of quantum non-localities

All this can be summarized conveniently in a table, where the columns indicate different causes precipitating non-local effects, and rows indicate the different effects of the causes. We assume that the causes relate to cases of measuring an observable $A$ on the left-hand side particle of a two-particle EPR system, while the effects involve a correlated observable $B$ and its values, characterizing the right-hand side particle of the same system.

So, for example, Non-Loc\textsubscript{13} represents a non-local influence such that by choosing a particular observable to be measured on a distant particle we can change an already possessed definite quantum property of another system (a sharp value of an observable) to another definite property, and Non-Loc\textsubscript{21} characterizes a situation in which by “forcing” a particular outcome to appear for one particle we could give some value to a previously undefined observable. It should not be difficult to identify in this table the versions of non-locality known from our previous discussion of the EPR and Bell theorems. The EPR non-locality will obviously be identical to Non-Loc\textsubscript{11}, and Bell non-locality to Non-Loc\textsubscript{13}. On the other hand, intended interpretations of “parameter non-locality” and “outcome non-
locality” would coincide, respectively, with Non-Loc\textsubscript{14} and Non-Loc\textsubscript{24}. However, as we recall from the previous section, at least the first interpretation raises serious concerns.

Non-localities Non-Loc\textsubscript{14} and Non-Loc\textsubscript{24} can be given a narrow interpretation, under which probability measures are only permitted to assume values other than 1, or a broad interpretation in which the probability distributions which give measure 1 to one value of the observable $B$ and 0 to the rest are included. However, in the latter case all of the remaining types of non-localities would count as subspecies of the broad categories Non-Loc\textsubscript{14} or Non-Loc\textsubscript{24}, so we will opt for a narrow interpretation which uses only “non-trivial” probability distributions.

We can now use the classification of non-localities given in Tab 1.1 in order to attach different “weights” to particular types of non-local interactions. It has already been remarked several times that not all cases of violation of the locality intuition are equal. The departure from the common-sense ontological view regarding the propagation of causal influences varies depending on the “severity” of accepted non-locality. It seems natural to treat one type of non-locality as being less “severe” than another, depending on two factors: on the degree to which the effect makes a difference in the world (for example, whether the effect can be observed, or detected), and on the ease with which the cause required to obtain such an effect can be produced. Regarding the latter factor, it can be argued that all non-localities placed in the first column under the heading of “Choice of a measured observable” are stronger than their counterparts in the second column. This is the case, because the choice to measure a particular observable can be made freely by an experimenter (or, more cautiously, because there are no strong arguments against such a contention), whereas the outcome obtained in the course of the experimenter’s action seems to be independent of their will. Hence, superluminal signaling is in principle possible in the first case, while the second case does not offer us such an opportunity. It may be also pointed out that “outcome non-locality” offends our common sense to a lesser degree, because it resembles cases of spurious non-causal correlations of the sort we discussed in footnote 14. According to this intuition, it may be tempting to explain away a correlation between outcomes of space-like separated experiments by pointing to a possible hidden factor which causally determines both outcomes and hence secures their connection. In other words, an outcome revealed in one experiment would reflect certain “preordained” mechanism, which is also responsible for the other result. Such a move is plainly impossible in the
WHY DOES THE QUANTUM WORLD HAVE TO BE NON-LOCAL?

case of a correlation between a choice of experiment on one system and a physical state of the other, because by assumption the experimenter’s decision doesn’t depend causally on any physical factor in the past of the system. However, we should not forget that this strategy of explaining away non-local correlations between outcomes has been seriously undermined by Bell-like results (see e.g. van Fraassen 1983). Therefore, we have to accept that there is an irreducible element of non-locality in the correlation between the outcome of one experiment and a physical situation leading to the outcome of the other experiment.

Let us now turn to looking at how the strength of a particular type of non-locality depends upon the sort of effects brought about. No effect given in the left-hand side column of the table can be seen as literally observable, yet it should be quite clear that changing one definite value into another one comes close. If we could only monitor a particular physical magnitude over an extended period of time, we would definitely be able to notice that this magnitude “switched” between two values at a certain instant. On the other hand, there is no easy way to detect an “undefined” or “superposed” state of a particular observable. Any properly conducted measurement in quantum physics has to reveal a particular value, regardless of whether the system was initially in an eigenstate for a given observable or not (in short: there is no “superposition” label on the measuring device). We can only indirectly conclude that the initial state was not an eigenstate when we conduct the same measurement on a statistical ensemble of identically prepared objects. Hence, all transitions involving quantum “potentialities” seem to be at least epistemologically less radical than the transition from one actuality to another (once again this relates to the possibility of sending and receiving signals at a superluminal speed). For that reason we may argue that the strongest type of non-locality is Non-Loc$_{13}$ (identical to Bell non-locality!) and next is Non-Loc$_{11}$ along with Non-Loc$_{12}$. The weakest variant of non-locality seems to be associated with the statistical transition from one probability distribution to a different one, and combining this with the analysis from the previous paragraph, we can see that the best candidate for the least radical departure from locality is Non-Loc$_{24}$.

This ordering can also be confirmed by considerations of a more ontological nature. There is a quite commonly accepted intuition which attaches greater ontological importance to well-defined states of a particular observable rather than undefined ones. According to this intuition, a genuine property attribution happens only when a system possesses a definitive
value $a$ for a given observable $A$. When, on the other hand, the system is not in an eigenstate with respect to $A$, it is claimed that the system lacks certain physical characteristics, rather than possesses a property of a different sort. Although this approach is not entirely uncontroversial, it seems that probabilistic propensities or dispositions are less clear-cut attributes of physical objects than sharp values of magnitudes, like position, momentum, or charm. Hence, a switch from, and to a state characterized only by some sort of “fuzziness” should count as less conspicuous compared with a transition between definitive properties.

Finally, a word about how to understand the causal relation connecting causes to effects in different types of non-local influences. Causality is often associated with repeatability: every time a cause occurs, it is followed by the appropriate effect. However, this approach ignores the fact that causal relations are highly sensitive to external conditions, meaning that the alleged repeatability holds only under certain strict conditions (an example: striking a match causes flames to occur only if there is enough oxygen in the atmosphere, if the match is not wet, if we strike at the proper angle, etc.). These conditions are typically difficult to characterize properly and exhaustively, so it is argued that a better approach is to define causality in terms of “necessary” conditions: if a cause hadn’t occurred, the effect wouldn’t have occurred. This characterization is a “singular” one; it defines a cause for a particular event in the context of its actual realization in particular conditions, and not as the regularity “events of type $A$ always cause events of type $B$”. Here we will be following this second method of

---

28 M. Redhead would probably disagree. See (Redhead 1987, pp. 48-49). However, this intuition can be further supported by the tendency within science to reduce dispositional properties, such as fragility or water-solubility, to more fundamental physico-chemical properties of objects. The only problem in quantum mechanics is that quantum propensities cannot be reduced to anything more fundamental. We will talk more about this issue in sec. 6.4.

29 Once again, as it happened previously, the following intuition may have influenced this particular judgment. Someone could namely interpret the “undefined” state merely epistemologically—as a reflection of our subjective ignorance regarding the exact value of the magnitude in question. Of course in such a situation a “transition” from such state to a fully defined state would not count as something that happens objectively “there”. However, we should remember that the ignorance interpretation of the quantum state is essentially equivalent to the hidden variable hypothesis, and as such is seen as highly improbable.

30 It is almost impossible to give here an exhaustive list of the relevant texts that deal with the philosophical problems of causation, so huge is the literature of this topic. An introductory survey of the main conceptions of causality can be found in the an-
interpreting causality, thus adding one more area for applying counterfac-
tual conditionals to the problems analyzed in this book. Because locality
conditions are given in terms of the non-existence of causal connections
between space-like separated events, no wonder that counterfactuals will
play a role in the precise formulation of the locality assumption. Hence it
may be a good idea to look more closely at the logical analysis of counter-
factual conditionals in order to prepare logical tools necessary for our
undertaking. The next chapter will be entirely devoted to this task.