Quantum Mechanics and Global Determinism

Emily Christine Adlam

Basic Research Community for Physics, Leipzig, Germany. E-mail: eadlam90@gmail.com

Editors: Ken Wharton & Peter J. Lewis

Article history: Submitted on February 27, 2018; Accepted on July 5, 2018; Published on July 16, 2018.

It is proposed that certain features of quantum mechanics may be perspectival effects, which arise because experiments performed on locally accessible variables can only uncover a certain subset of the correlations exhibited by an underlying deterministic theory. This hypothesis is used to derive the no-signaling principle, thus resolving an open question regarding the apparently fine-tuned nature of quantum correlations. Some potential objections to this approach are then discussed and answered. Quanta 2018; 7: 40–53.

1 Introduction

As scientists we usually aim to formulate laws which predict facts about the future from facts about the present, and this goal places strong constraints on the type of laws which we typically propose. But we should be wary of allowing ourselves to slip into a subtle and pervasive version of the anthropomorphic fallacy: as Ken Wharton puts it, “There’s one last anthropocentric attitude that needs to go, the idea that the computations we perform are the same computations performed by the universe.” Adherents of any reasonably strong form of scientific realism will surely agree that it is meaningful to distinguish between the laws that scientists write down and the true underlying laws of nature, and it would be naive to suppose that the two exactly coincide. For example, we know from quantum mechanics that the laws of nature may be spatially nonlocal [2, § 12], and special relativity indicates that if the laws of nature are spatially nonlocal then they are likely to also be temporally nonlocal. [3], and together these two forms of nonlocality open up a large range of possibilities for fundamental laws which go beyond our standard paradigms for theory-building. For example, our current scientific laws take as their arguments variables which can be accessed and manipulated by local observers such as ourselves, but if the true laws of nature are spatially and/or temporally nonlocal then they might also invoke global variables which are defined with respect to the whole history of the universe, and this possibility casts new light on a number of old scientific questions. In particular, it is common to regard the issue of determinism as a dichotomic dilemma: the universe must either be deterministic, meaning that the laws of nature together with a specification of the conditions on a single hyperplane of simultaneity suffice to fix all conditions to the future of that hyperplane, or fundamentally probabilistic, meaning that there exist facts about the world which are not fixed by the laws of nature together with initial conditions. [4, 5]. But if the laws of nature do in fact take global variables as their arguments, then we have a third option: the universe could be deterministic in a generalized sense when the global variables are taken into account, whilst appearing probabilistic to agents such as ourselves who do not have access to the whole picture. This is not purely a matter of foundational interest: a more mature understanding of the true form of the laws of nature is likely to help guide the development of a new generation of theories, and thus moving away from the
paradigm of purely predictive laws might actually lead to new experimental predictions. But of course, according to the hypothesis proposed, observers such as ourselves cannot directly measure global variables, so it is not immediately obvious how to go about narrowing down the range of possibilities for global laws. The approach I will pursue here involves determining how various features of global laws might be expected to constrain the local behaviour that we observe. In particular, I will consider the hypothesis that the global laws are deterministic in an appropriately generalised sense, and I will show that this idea is capable of providing a simple, physically realistic explanation for the fine-tuned nature of quantum no-signaling. This suggests that quantum no-signaling may be described as a perspectival effect, which is not to say that it is not real, but simply that it does not hold universally—it is true only of the variables which happen to be accessible to observers such as ourselves.

The basic components of this approach are set out in section 2 and a short outline of the argument is presented in section 3. The argument is developed in greater detail in section 4; section 5 deals with some conceptual objections and section 6 discusses the underlying assumptions.

2 Preliminaries

2.1 The fine-tuning problem for no-signaling

For much of the history of physics it has been taken for granted that causal influences are necessarily local in space and time. Quantum physics, however, has challenged that assumption. Experimentally verified violations of Bell’s theorem demonstrate that, ignoring loopholes, the universe must contain correlations which are in a specific technical sense nonlocal in space [6]. As a result, physicists have begun to take seriously the possibility that the universe contains spatially nonlocal causal influences—a choice of measurement direction for a measurement at one spacetime point can have a direct causal effect on the outcome of a measurement performed at some spacelike separated point.

Yet we also know none of the nonlocal correlations which can be produced within quantum physics can be used to send signals: we cannot, by choosing a measurement direction at one spacetime point, influence the probability distribution over the outcomes of a measurement performed at some spacelike separated point. It is not uncommon to feel that there is something uncomfortably conspiratorial about this state of affairs [7], and indeed, this intuition can be made formal: Wood and Spekkens have shown that any causal model for nonsignaling correlations which violate a Bell inequality must involve fine-tuning of the causal parameters, [8] i.e. given any such model, the conditional independences between variables exhibited by the probability distribution induced by the model will not remain the same when the values of the parameters of the model are varied.

It is a standard methodological principle of theoretical physics that where we see fine-tuning we should seek an explanation for it [9, 10], and thus there is a clear imperative for us to provide an explanation for the fact that all quantum correlations obey the no-signaling principle. Simply postulating a global prohibition on signaling is not sufficient, because signaling is an agent-relative concept and thus such a prohibition would seem uncomfortably close to making an in-principle distinction between agent and world, which is distasteful from the point of view of realism about science. Likewise it is preferable to avoid resorting to claims that physics is in some sense about information, which frequently give rise to fallacies [11].

In this article, I consider carefully the distinction between no-signaling in the quantum and relativistic contexts and propose a unifying definition which I believe to capture the important content of both versions of no-signaling, then propose an alternative, physically realistic explanation for this property, thus suggesting a resolution to the fine-tuning problem.

2.2 The process framework

I will make the simplifying assumption that the world is composed of a large set of pointlike events, which consist simply of a particular variable taking a particular value at a particular spacetime point. I will suppose that the laws of nature govern the relations between these events directly, thus dispensing with the concept of state.

This simple picture of the world does not distinguish between different types of variables, but in order to proceed further it is necessary to make a distinction between input variables, whose values can be freely chosen by local agents, and output variables, whose values are not fully controllable by local agents. For example, in a measurement of particle spin, the choice of measurement direction is an input variable, and the measurement outcome is an output variable. It is important to keep in mind that this distinction depends crucially on the notion of agency and on the specific physical manifestation of the relevant class of agents, since different types of agents with different physical capabilities might well pick out a different set of input variables, and hence most of our definitions and commentary from here on will be implicitly relativized to a class of agents. Indeed, under the hypothesis presented forward in this article, the distinction between input and output variables is entirely perspectival: there is no difference between these variables from the point of view...
of view of the universe, only from the point of view of agents who are able to set the values of some variables but not others. However, one could imagine theories in which this is not the case. For example, in a collapse model of quantum mechanics, the results of quantum measurements (i.e. output variables) differ in a fundamental way from variables which can be controlled by local agents, because they are determined by the probabilistic collapse process rather than by deterministic local processes.

With this distinction in hand, I define a local experiment as a pair of events: an input variable is set to some value and an output variable assumes some value. I will assume that an experiment is effectively instantaneous, so input and output variables are defined at the same spacetime point, but this idealization could be relaxed without significantly changing the arguments. I will also assume that there exist spacetime points at which these events take place, and that at least some pairs of spacetime points stand in spatiotemporal relations to one another, but because I want to avoid presuppositions about the nature of spacetime, I will make only minimal assumptions about the nature of these relations. In particular, I will assume that these relations have at least enough structure to render coherent the distinction between symmetric and asymmetric processes which I introduce in section 2.4 and I will assume the existence of an arrow of time, so that it is sometimes possible to say that one event is in the past of another. Since both of these distinctions are instantiated in our actual universe, it seems reasonable to assume that whatever the underlying reality of spacetime may be, it does indeed realize these distinctions, albeit perhaps only in an emergent sense.

Local experiments may be combined to form processes, as in the process framework introduced in [12], which specifies probabilities for the outcomes of local experiments without assuming the existence of a global causal order between experiments. In this framework, a process is defined by specifying the set of conditional probabilities for the relevant input and output variables, i.e. a complete set of probabilities of the form \( p(O^A = a, O^B = b, \ldots | N^A = a', N^B = b', \ldots; w^{AB\ldots}) \) where \( N^A \) is an input variable, \( O^X \) is an output variable, and \( w^{AB\ldots} \) is the variable that defines the global setup in which the individual experiments take place. Hence these conditional probabilities are necessarily identical for any implementation of the process.

Since a given process may have multiple input variables, henceforth I will write the input variables for a given process as a vector \( \mathcal{N} \), and the complete set of possible values for these variables as a set of vectors \( \mathcal{N} \). Likewise I will write the output variables for a given process as a vector \( \mathcal{O} \), with the complete set of possible values for these variables given by the set of vectors \( \mathcal{O} \).

### 2.3 Accessible and global variables

Because we are spatiotemporal beings, we can only perform experiments involving accessible variables, i.e. variables which are defined over a proper subset of the spacetime points in the universe and can therefore be accessed by observers localized over proper subsets of the spacetime points in the universe. However if the world is governed by global laws it is possible that experimental outcomes may also depend on global variables which cannot even in principle be accessed or manipulated at any finite set of spacetime points, perhaps because they describe the relation of a given instance of a process to the rest of the universe, or because they are defined by global topology; for example, if it were possible to define a classical action for the entire universe over the whole of history, that action would be a global variable. While this proposal may at first seem a significant conceptual leap, note that it is already almost a commonplace that entanglement implies the existence of properties of distributed systems which are not reducible to any collection of properties of subsystems; [11,13,14] our proposed global variables take this notion one step further, since, according to the hypothesis proposed, such properties cannot be accessed even by a sufficiently large and distributed set of observers, but nonetheless the phenomenon of entanglement provides at least some precedent for this way of thinking about distributed properties.

I will therefore extend the process framework by allowing that in addition to the usual accessible input variables, a process may also have global input variables, so we now have a full set of input variables \( V \) which may be split into a vector of accessible input variables \( N \) taking values in the set \( \mathcal{N} \), and a vector of global input variables \( Q \) taking values in the set \( \mathcal{Q} \). However, since global input variables cannot be accessed or manipulated at any small region of spacetime, it is still true that as in the standard process framework, the accessible input variables \( N \) provide the only means of information exchange between local experiments.

A word of caution: the distinction made here between accessible variables and global variables is not the same as the distinction made in relativistic quantum theory between local variables and nonlocal variables. In relativistic quantum theory, a local variable is a variable related to only one small region of space whilst a nonlocal variable is a variable related to several distinct small regions of space—but since such nonlocal variables in quantum mechanics (apart from variables related to the spread-out fermionic wave function) can be measured by local observers who coordinate their actions and subsequently share information [15], these nonlocal variables are still accessible variables in my terminology. Indeed,
no variables currently defined by quantum mechanics or quantum field theory are global variables in my sense, since all can be measured by appropriately distributed collections of observers such as ourselves.

2.4 Spatiotemporally symmetric processes

It is an important feature of our experience that many types of processes which we encounter have a natural temporal direction, namely, the inputs and outputs are correlated in a certain way only if the parts at which the parts of the process occur stand in a certain asymmetric spatiotemporal relation to one another. For example, in a process where one device sends an electromagnetic signal to a second device, the correlations depend on whether or not the second device is operated after the signal from the first has had time to arrive. However, in quantum mechanics we encounter processes which are spatiotemporally symmetric:

Definition 2.1. A process is spatiotemporally symmetric if the correlations between input and output variables remain unchanged under any permutation of the sites at which the parts of the process take place.

For example, the process of performing measurements on an entangled state is spatiotemporally symmetric, because the only constraint on the spatiotemporal relation between the measurement sites is that both must lie in the future lightcone of the point at which the entangled state was prepared, and this constraint is symmetric with respect to the two measurements, so if we interchange the two measurement sites, the same correlations will still be observed. Indeed, any process where we perform measurements on two distinct quantum systems will be spatiotemporally symmetric, since the probabilities for the outcomes of quantum measurements performed on distinct quantum systems do not depend on the time or place of the measurement. The fact that measurements on two distinct non-interacting quantum systems are spatiotemporally symmetric is implied by the fact that the algebras of observables for distinct non-interacting quantum systems commute (this is sometimes known as microcausality or kinematic independence [16]). However, the commutativity of observables is a stronger constraint than spatiotemporal symmetry: one could imagine a spatiotemporally symmetric process where the algebras of observables would fail to commute, such as a set of measurements on some post-quantum entangled state of systems A and B where the outcome of the measurement on B depends on the choice of the measurement made on A even if the measurement on A is spacelike separated from or in the causal future of the measurement on B.

Of course, the distinction made here depends on our earlier distinction between input and output variables—given a process which is not spatiotemporally symmetric, if we could simply swap the inputs and the outputs, then the asymmetry would disappear. Thus the distinction between spatiotemporally-symmetric and spatiotemporally-asymmetric processes is, like the distinction between input and output variables, an agent-relative one: agents with a very different physical realisation might pick out different classes of spatiotemporally-symmetric processes.

Note that given a process which is not spatiotemporally symmetric, we can always use the asymmetry to assign a unique temporal order to the parts of the process, so without making additional assumptions about the nature of the embedding spacetime we can say that, given any pair of such events, one event must lie in the past of the other. The same may not necessarily be true of pairs of events which form part of a spatiotemporally symmetric process.

2.5 Global determinism

From Einstein onwards, [17] many physicists have hoped to find a way of making quantum mechanics compatible with determinism. However this project has encountered numerous difficulties, not least because the very notion of determinism is in need of some updating.

Giving a definition of determinism is notoriously difficult, [5, § 3] but typical proposals require that all spacetime points to the future of a given spacelike hyperplane \( \Sigma \) are fully determined by the conditions on \( \Sigma \) together with the laws of the theory. But the assumption that the entire universe can be understood as an initial value problem [18] seems unwarranted: it has been argued elsewhere [13] that we should take seriously the hypothesis that “the universe (runs) not as a computer, but as a global four-dimensional problem that (is) solved all at once.” [1]

In light of such possibilities, it seems we would be better served by a less restrictive notion that one might refer to as global determinism.

To give an intuitive picture of what the notion of global determinism entails, I suggest the following analogy. In a globally deterministic world, the laws of nature prescribe a unique course of history in the same way as a properly set up game of sudoku has a unique solution; the laws of nature govern the whole course of history at once, rather than moment-by-moment, in the same way as the rules of the game of sudoku apply to the whole grid at once, rather than dictating the entries column by column. Indeed, if a completed sudoku square were to be revealed column by column to an observer who could not see the whole square at once, each column would appear to be
related to the previous one in a probabilistic way (there would be obvious patterns of dependence which could be described by probabilistic rules, but there would not usually be enough information available to determine the next column exactly), and I suggest that similarly, the apparently probabilistic nature of quantum theory is a consequence of our inability to see the whole picture.

The task of giving a precise definition for global determinism is likely to inherit many of the difficulties of defining conventional determinism, with the additional difficulty that we currently have little idea what form globally deterministic laws of nature might take. Thus rather than providing a formal definition of global determinism, I will instead make the notion mathematically precise by working with what I take to be a necessary condition for global determinism:

**Necessary condition for global determinism:** For any process \( P \) with output variables \( O \) taking values in \( \mathcal{O} \), there exists a set of (accessible and/or global) input variables \( V \) taking values in \( \mathcal{V} \), and a function \( f^P: \mathcal{V} \rightarrow \mathcal{O} \) such that if process \( P \) is performed with the (accessible and/or global) input variables taking values \( v \in \mathcal{V} \), then we obtain output variables \( o \in \mathcal{O} \) such that \( o = f^P(v) \).

**Notes**

1. There exist well-defined relative frequencies for the values of variables defined with respect to processes, so we can specify frequentist probability distributions over variables and hence employ entropic quantities [19] such as \( H(O), H(O|N) \) and \( I(O : N) \) without invoking a notion of objective probability. In this language, the condition above may be restated as the requirement that if process \( P \) has input variables \( V \) and output variables \( O \), then \( H(O|V) = 0 \).

Here, \( H \) denotes the Shannon entropy, which is defined as
\[
H(O) := -\sum_o p(O = o) \log(p(O = o));
\]
\( H(O|Q) \) denotes the conditional Shannon entropy, which is defined as
\[
H(O|Q) := \sum_{o,q} p(O = o|Q = q) \log(p(O = o|Q = q));
\] and \( I(A : B) \) denotes the Shannon mutual information, which is defined as
\[
I(A : B) := \sum_{x,y} p(A = x, B = y) \log \frac{p(A=x,B=y)}{p(A=x)p(B=y)}.
\]

2. The statistical data on which quantum mechanics is based comes from experiments in which some parameter is varied and the correlations between inputs and outputs are observed. Because the hypothesis of global determinism is supposed to be an explanation for these statistics, I will henceforth assume that in any such experimental process, the apparent dependence of the outputs on the inputs is real and that the apparent level of dependence is as it appears in the experimental statistics, i.e. if we find experimentally that \( I(O : N) = x \) for \( O \) some output variable, \( N \) some input variable, and \( x \) some constant, then the relation \( I(O : N) = x \) is implied by the function \( f^P \). I will also assume that the accessible input variables are statistically independent of one another and of all other variables on which the function \( f^P \) depends, e.g. for a process with a single accessible input variable \( N \) and a single accessible output variable \( O \), if we denote the set of all other variables on which \( f^P \) depends by \( Q \), we have \( I(ON : Q) = I(O : Q|N) \) since \( I(N : Q) = 0 \). See section [6.1] for further discussion of these assumptions.

3. Since we require that the functional relation between inputs and outputs is the same for any instantiation of a given process, it follows that if two processes \( P_1 \) and \( P_2 \) are combined, with the accessible output variables \( O_1 \) for process \( P_1 \) being used as the accessible input variables \( N_2 \) for process \( P_2 \), then the accessible output variables \( O_2 \) for process \( P_2 \) must be a function of the full set of accessible and global input variables \( N_1, Q_1 \) for process \( P_1 \) together with the global variables \( Q_2 \) for process two, i.e. if the composition is performed with the accessible input variables for \( P_1 \) taking values \( n_1 \) and the global input variables for the two processes taking values \( q_1 \) and \( q_2 \), then for process \( P_1 \) we obtain output variables \( o_1 = f^{P_1}(v_1, q_1) \) and for process two we obtain output variables \( o_2 = f^{P_2}(f^{P_1}(v_1, q_1), q_2) \) for some functions \( f^{P_1}, f^{P_2} \).

However there exist more general ways of composing processes—we may choose to use the output variables of one process to determine only a proper subset of the input variables for another process, and indeed, in later sections we will see that it is sometimes possible to use a subset of the output variables of \( P^1 \) to determine a proper subset of the input variables of \( P^2 \) and also use a subset of the output variables of \( P^2 \) to determine a proper subset of the input variables of \( P^1 \). For these more complex compositions it is natural to require that the full set of output variables for the joint processes should be a function of all remaining input variables, i.e. we require the following:

**Definition 2.2. Global determinism under composition:** Consider two processes \( P_i : i \in \{0, 1\} \), with accessible input variables \( N_i \) taking values in \( \mathcal{N}_i \), global input variables \( Q_i \) taking values in \( \mathcal{Q}_i \), and accessible output variables \( O_i \) taking values in \( \mathcal{O}_i \), for which it is possible to construct a composi-
A spatiotemporally symmetric process \( P \) can be used to produce a loop by performing the same process twice over and allowing the accessible output variables from one process to determine the accessible input variables for the other, and vice versa.

For example, consider a process which can be modelled by two black boxes, with the first box taking an input \( N \) at some spacetime point and the second box producing an output \( O \) at some other spacetime point. If global determinism holds, the output \( O \) for process \( P \) must be a function of \( N \) and possibly also some global input variables \( Q \). Now consider constructing two copies of the boxes required to implement process \( P \): boxes \( X_1 \) and \( X_2 \) and boxes \( Y_1 \) and \( Y_2 \). We may then compose the two processes as depicted in Fig. 1. Alice takes boxes \( X_1 \) and \( Y_2 \) and Bob takes boxes \( X_2 \) and \( Y_1 \), then at some spacetime point, Alice obtains \( O^x \) from \( X_1 \) and uses the value of \( O^y \) to choose the input \( N^y \) for box \( Y_2 \), while at some other spacetime point, Bob obtains \( O^y \) from \( Y_1 \) and uses the value of \( O^x \) to choose the input \( N^x \) for box \( X_2 \).

Under the composition rule for global determinism, the outcome variables \( O^x, O^y \) are a function of the global variables \( Q^x, Q^y \) only. But if the output variables \( O \) for this process depend too strongly on the accessible input variables \( N \) for the process, then under this composition there will not be enough information in the global variables \( Q^x, Q^y \) to fix the values of the variables \( O^x, O^y \), so global determinism will fail. Hence global determinism imposes strict constraints on the correlations which can hold between output variables and accessible input variables for spatiotemporally symmetric processes.

One consequence is that spatiotemporally symmetric processes must obey the no-signaling principle. For if a spatiotemporally symmetric process were to violate the no-signaling principle, a composition as in Fig. 1 could be used to produce a closed causal loop: Alice could receive a message, then use the signaling process to pass the message on to Bob, who could then use the signaling process to return the message to Alice. The content of the message would be its own cause and therefore would not be a function of the values of variables elsewhere in the universe, in violation of the necessary condition for global determinism. We have thus given a heuristic argument suggesting that in a globally deterministic universe the correlations exhibited by accessible variables in spatiotemporally symmetric processes must always satisfy the no-signaling principle; this argument is formalised in the following section.
and thus places no limits on the transmission of information by other means, such as physical signals travelling from one point to another. Nor does the relativistic prediction imply the quantum-mechanical theorem, because the relativistic prediction has implications only for the results of measurements made on systems at a spacelike separation, whereas the quantum theorem is independent of the spatiotemporal location of the measurements in question. Indeed, the analogue of the no-signaling theorem in quantum field theory is normally taken to be the statement that observables defined on regions of spacetime separated by a spacelike interval commute [23], but once again this statement only has implications for the results of measurements made at a spacelike separation, and is thus in some sense weaker than the original quantum no-signaling theorem.

To close the gap between the quantum and relativistic formulations of no-signaling, we need to invoke the additional fact that measurements on entangled quantum systems exhibit the same statistical features independent of the spatiotemporal relation between the measurements, i.e. measurements performed on distinct quantum systems always constitute a spatiotemporally symmetric process. This suggests a way of subsuming the key ideas from these different formulations of no-signaling under a single principle:

Definition 4.1. Generalized no-signaling principle: Consider a spatiotemporally symmetric process involving a set of devices \( \{ D_i : i \in \{1 \ldots N\} \) operating at distinct spacetime points \( \{ P_i \} \), such that at point \( P_i \) device \( D_i \) accepts an accessible input variable \( N_i \) and produces an accessible output variable \( O_i \). Let \( J \) be any subset of \( \{1 \ldots N\} \), let \( O_J \) be the set of variables \( \{ O_j : j \in J \} \), let \( N_J \) be the set of variables \( \{ N_j : j \in J \} \), and let \( \mathcal{P} \) be the set of variables defined at points which are in the past of all the points \( P_i \). Then if the inputs \( \{ N_i \} \) are uncorrelated, the outcomes satisfy \( p(O_J|N_1, \ldots, N_n, \mathcal{P}) = p(O_J|N_J, \mathcal{P}) \).

Notes

1. As before, this definition depends crucially on the distinction between input and output variables, and therefore the definition must be understood relative to some set of agents. This is not particularly surprising, since the notion of signaling is implicitly dependent on notion of agency—a signal is simply a variable whose value a given agent is able to exert control over, perhaps locally or perhaps at some spatiotemporal distance.

2. It follows that a spatiotemporally symmetric process obeys the generalized no-signaling principle iff for any set of inputs \( N \) and any set of outputs \( O \) which are accepted and produced respectively at non-overlapping sets of points, the mutual information \( I(O : N|\mathcal{P}) \) is zero (provided that the process is performed with uncorrelated inputs) [24]. For brevity, \( \mathcal{P} \) will henceforth be omitted.

3. In formulating the relativistic no-signaling principle, it is common to conditionalize on the set of variables in the past lightcone of the points \( P_i \) in order to rule out any possibility of a joint common cause explanation for the correlations (e.g. see [25]). To avoid assumptions about the nature of the background spacetime in which the devices operate, that conditionalization is here replaced with a conditionalization on the set of all variables which are in the past of the variables at the points \( P_i \).

4. The generalized principle implies the quantum-mechanical no-signaling principle, because any pair of measurements performed on two distinct quantum systems is spatiotemporally symmetric. Of course consecutive measurements performed on the same system are not spatiotemporally symmetric, but this is not a problem for the approach set out here, because the algebras of observables on individual quantum systems are not commutative and hence may indeed be signaling.

5. The generalized principle does not imply that no signal can be transmitted faster than light, but if we believe that the universe contains a finite number of types of spatiotemporally asymmetric process, the generalized principle does imply that signals cannot be transmitted faster than some universal maximum speed, i.e. the fastest speed associated with any spatiotemporal process. We are free to hypothesize that in our universe this happens to be the speed of light, thus obtaining the relativistic no-signaling principle; the field-theoretic no-signaling principle could also be obtained by similar reasoning. (Note that this does not require us to make the unwarranted assumption that the universe itself is finite, as we could have an infinite number of instances of a finite number of types of process; essentially, the assumption required is that the laws of nature may be finitely specified, even if the universe is infinite).

I now demonstrate that the generalized no-signaling principle can be derived from the hypothesis of global determinism under the assumptions we have made here. A similar approach was used by Arntzenius [26] to derive a variant of no-signaling from a prohibition on closed causal loops, but that derivation applies to probabilistic theories and is therefore threatened by ambiguities about
the behaviour of probabilities under composition [27]: assuming global determinism allows us to circumvent these difficulties.

**Theorem 4.2.** If the world is globally deterministic, then any spatiotemporally symmetric process $P$ must satisfy the generalized no-signaling principle for all its accessible input variables.

We will need the following lemma:

**Lemma 4.3.** If $N$ and $Q$ are statistically independent and $O$ is a function of $N$ and $Q$, then $H(O) = I(O : N) + I(NO : Q)$.

**Proof.** Since $O$ is a function of $N$ and $Q$, $H(O|NQ) = 0$, and thus from Bayes’ rule $H(NQ|O) + H(O) = H(NQ)$. Using the chain rule for conditional entropy, we obtain $H(N|O) + H(Q|NO) + H(O) = H(NQ)$. Since $N$ and $Q$ are statistically independent, $H(NQ) = H(N) + H(Q)$. Thus $H(O) = H(N) - H(N|O) + H(Q) - H(Q|NO) = I(O : N) + I(NO : Q)$. □

Thus we may proceed with the proof of theorem 4.2.

**Proof.** Consider any spatiotemporally symmetric process $P$ in which an accessible input variable $N$ is accepted at one point and an accessible output variable $O$ is produced at a distinct point. We assume that the world is globally deterministic, so $O$ is a function of $N$ and possibly also some additional variables $Q$. Since process $P$ is symmetric, it can be used to construct a cyclic process as in Fig. 1. From the composition rule (2.2), we have $H(Q^NO^1Q^2\bar{Q}^2) = 0$, and hence $H(O^1O^2) = I(O^1O^2 : Q^1Q^2)$. From the definition of the mutual information, $H(Q^0Q^1) = H(Q^0) + H(Q^1) - I(Q^0 : Q^1)$. In this construction, $I(Q^0 : Q^1) = I(Q^0 : N^0) = I(O : N)$, so $H(Q^0Q^1) = 2H(O) - I(O : N)$. Since $O^1$ is a function of $N^1$ and $Q^1$ and $O^2$ is a function of $N^2$ and $Q^2$, $I(O^1O^2 : Q^1Q^2) \leq I(N^0O^0 : Q^0) + I(N^3O^3 : Q^2) = 2I(NO : Q)$. Since $O$ is a function of $N$ and $Q$, and $N$ and $Q$ are independent, from lemma 4.3 we have $H(O) = I(O : N) + I(NO : Q)$. Combining the above results, we obtain $2H(O) - I(O : N) \leq 2H(O) - 2I(O : N)$. Hence $I(O : N) \leq 0$. But the Shannon mutual information is nonnegative, so $I(O : N) = 0$, and thus by applying the first note on [4.1] we conclude that this process must be non-signaling. □

Note that the proof applies for the case of a process which takes an input at some point and produces an output at another point. It is more common to discuss no-signaling with respect to processes where at each of two separate points an input is provided and an output is produced; for example, when we perform measurements on a Bell pair, the two choices of measurement direction are the inputs and the measurement results are the outputs. However, clearly the former type of process can always be produced from the latter by simply fixing the value of one input and ignoring the value of one output, so the same proof applies in the more general case.

5 Objections

5.1 Records

One possible criticism of the concept of global determinism is that if the value of a variable can be determined by states of affairs at any other point in space or time, then the outcome of every experiment would be trivially determined by any subsequent records of the outcome, so any world in which records exist would qualify as globally deterministic. I have a simple response to this objection: such an arrangement would indeed be consistent with the notion of global determinism, but if the outcome of a reproducible process were fixed by some subsequent record of that outcome, that record could not also be fixed by the outcome, since otherwise the outcome and record would together form a closed causal loop, and as argued in section 3, closed causal loops cannot exist in a globally deterministic universe. Hence under such circumstances the record would itself necessarily be fixed by the value of some other variable(s), $X$, and hence we could equally well say that the outcome and record are jointly determined by $X$, thus rendering the notion of global determinism nontrivial.

5.2 Global variables vs hidden variables

Since the nature of the postulated global variables played no explicit role in the derivation of the no-signaling principle, it is reasonable to ask why these variables need to be global, rather than simply inaccessible like hidden variables in interpretations such as the de Broglie Bohm interpretation [28].

However, the defining feature of the global variables, postulated here is that when a global variable is an input to some process, it is not possible to set that global variable to be a function of the output to some other process, so global variables cannot be used to form closed loops in constructions like that employed in section 3. Although I described the construction of these loops in terms of the actions of agents Alice and Bob, these arguments would apply in the same way to natural or spontaneous evolutions in which the output of one process becomes the input to another, and therefore in a globally deterministic universe, the variables that I refer to as global variables must be inaccessible not only to observers such as ourselves.
5.3 Learning about global variables

If it is the case that global variables help to fix the outcomes of local experiments in certain kinds of processes, then those processes must have nonzero mutual information $I(O : Q)$ between the accessible outputs and the relevant global variables. Yet mutual information is symmetric, so this implies that the outcomes of local experiments gives us information about the values of the global variables, which seems to contradict the stipulation that global variables should be strictly inaccessible.

But in fact there is no contradiction here. Global variables are composed from accessible variables, and therefore the accessible variables must certainly provide information about these global variables, but the information is of such a kind that it is not usable unless we have access to the full set of relevant accessible input variables. As a simple example, imagine a type of process which has an apparently random outcome labelled by $+1$ or $-1$, and consider a global law which requires that the parity of the set of all the outcomes for all such events occurring throughout history must be equal to $+1$. A sufficiently complex set of such constraints could fix the outcome for every individual process uniquely, and yet the outcome of any individual process gives no information whatsoever about the parity of the whole sequence unless we already know the outcomes of all the other iterations throughout history. The global variable here is not the parity $P$ of the sequence, since for any individual outcome $I(O : P) = 0$, but rather a complete description of all the other constraints which apply to events in the sequence and which jointly fix all the outcomes in the sequence uniquely.

This might sound counterintuitive, but behaviour of this kind is in fact very common in quantum mechanics. As noted in section 2.3, global variables are somewhat analogous to entangled states in the sense that they are not reducible to properties of the individual subsystems, and indeed, a similar phenomenon occurs with entangled particles: if we are given a pair of qubits $A, B$ in an unknown maximally entangled state $\psi$, and we measure both particles in the same direction, then specifying the state $\psi$ and the outcome of the measurement on qubit $A$ uniquely determines the outcome of the measurement on qubit $B$, yet the outcome of the measurement on qubit $B$ alone gives no information whatsoever about the state $\psi$. Another example is the phenomenon of quantum secret sharing—it is possible to divide a quantum message into $N$ parts such that for some integer $m \in [2, N]$, any collection of fewer than $m$ parts gives no information at all about the message, but the message is determined uniquely by any collection of $m$ or greater parts. Our concept of global variables may be regarded as a natural extension of these well-understood features of quantum theory.

5.4 Initial conditions

Historically, determinism has been taken to imply that the history of the universe is fixed by the laws of nature together with an appropriate set of initial conditions, usually identified with the state of the universe at the beginning of time, and it seems likely that a world which is globally deterministic would likewise require some sort of initial conditions in order for the course of history to be determined uniquely—returning to the earlier analogy, recall that a sudoku square does not have a unique solution unless an at least 17 squares are initially filled.

Detailed answers to questions about the nature of the initial conditions for global determinism must wait upon specific proposals for the form of the globally deterministic laws, but I will briefly mention some possibilities. It might be that our inability to conceive of laws of nature which do not demand initial conditions is simply a failure of imagination. Alternatively, it might be the case that there exist some small set of variables in the universe which are not fixed by the other variables: these would have the same status as initial conditions in traditional formulations of determinism, but, moving away from the explicitly temporal character of traditional determinism, they are no longer required to occur at the start of time. For example, in a world governed by laws...
of nature requiring the optimization of a global variable akin to a classical action for the entire universe over the whole of history, the relevant initial conditions would be the boundaries of the entire universe, i.e. conditions at both the start and end of time. The existence of these variables would not be inconsistent with the hypothesis of global determinism, because the need to postulate initial conditions is not usually viewed as an objection to the traditional notion of determinism, and the concept of global determinism is only intended to be an update of this notion, not a logically stronger or more demanding concept.

5.5 No-go theorems

Finally, it is common to object that one or another of the quantum-mechanical no-go theorems rules out the possibility of a deterministic model for quantum theory. Here I briefly review the relevant no-go theorems and explain why they do not pose a challenge to global determinism.

5.5.1 Bell’s theorem

Bell’s theorem [34] shows that no hidden variable theory, which is constrained to obey the principle of spatial locality, can reproduce all the predictions of quantum mechanics. But one of the governing principles behind the approach I have adopted here is that one should not expect hidden variables to obey constraints of spatial or temporal locality. Indeed, on this view of the world nonlocality is the norm, and the phenomenon that demands explanation is not the nonlocality which is manifested in violations of Bell’s inequality, but rather the specific limits on nonlocality which are manifested in the no-signaling principle. Thus Bell’s theorem does not pose difficulties for the proposed globally deterministic universe, for in such a universe one would naturally expect nonlocality to be ubiquitous.

5.5.2 Randomness in quantum theory

Colbeck [35] has observed that from the no-signaling principle, it follows that if some measurement outcomes violates a Bell inequality, these outcomes must be inherently random. This fact has since been employed to develop protocols for device-independent randomness expansion [36,37]. It might seem that such results are in tension, if not outright contradiction, with the claim that the no-signaling principle itself follows from global determinism. However, there is no real contradiction here, because the proof in [35] shows only that the outcomes must be random in the sense that they are independent of any variable which could be used to perform signaling, i.e. any accessible input variable. This is not inconsistent with the claim that they are deterministic when we take into account global variables which are not accessible. Indeed, these results indicate that the apparent randomness of quantum mechanics may actually be viewed as a consequence of the global determinism of the theory: multiple authors [35–37] show that randomness arising from measurements on entangled systems is a necessary consequence of the no-signaling theorem, so if it is accepted that no-signalling is a consequence of global determinism then this sort of randomness is also a perspectival effect which follows from global determinism.

5.5.3 The Colbeck–Renner theorem

Renner and Colbeck [38] have a theorem to the effect that no extension of quantum mechanics can have greater predictive power than standard quantum mechanics. At first glance, this would seem to rule out global determinism, since the extension of quantum mechanics supplemented with the full set of variables which make the outcomes deterministic would clearly have greater predictive power than standard quantum mechanics. However, as Landsman discusses in detail [39], the result of Renner and Colbeck is based on a number of assumptions about the nature of the extension which may seem unreasonably strong. Moreover, the term predictive power is used here to refer to predictability in practice rather than merely in principle, as the proof assumes that the putative extension is accessible to ordinary observers and therefore obeys no-signaling constraints. For example, Renner and Colbeck use no-signaling to argue that for a pair of measurements with settings A, B and outcomes X, Y, if the putative extension is described by variables C, Z, then we must have \( p(A|BCYZ) = p(A) \), since otherwise it would be possible to signal using A to a spacelike separated observer who chooses B, C and has access to Y, Z. But if the variables C, Z are global ones, then it is not possible for any single observer who is spacelike separated from A to learn all the values B, C, Y and Z, so the no-signaling constraint need not hold. The key point is that in a practical sense, a globally deterministic theory would not have greater predictive power than quantum theory, since the extra variables which make the theory deterministic are not available to be used for predictions, and thus Renner and Colbeck’s result does not pose a problem for the postulation of global determinism.

5.5.4 Contextuality

The Kochen–Specker theorem [40] rules out the possibility of outcome deterministic noncontextual hidden variable models, i.e. models employing “hidden variables
which fix the outcome of each unsharp measurement (in a context independent manner).” [41] That is, in Hilbert spaces of dimension greater than two, it is possible to construct sets of observables for which it is impossible to assign values 1 and 0 consistently to every observable in the set such that in every possible measurement which can be constructed from this set of observables, exactly one observable is assigned the value 1, [42] which means that any deterministic hidden-variable model for quantum theory must be contextual, in the sense that measurement results sometimes depend on the context of measurement.

This no-go theorem does not pose a problem for global determinism, because although it implies the existence of hidden variables which fix the outcome of each unsharp measurement, it does not follow that the resulting model should be noncontextual. Indeed, the existence of contextuality is very natural in this framework: the hypothesis of global determinism implies that in general, the outcome of an experiment will depend not only on accessible input variables but also on facts about the relation between that experiment and other events throughout history, and the context of measurement is precisely such a relational fact, so it is to be expected that different measurement contexts may give rise to different results. Note, however, that the context of measurement is not a global variable, in the sense in which I use that term, because observers such as ourselves can observe and manipulate variables describing which measurement is being performed.

### 6 Assumptions

Below, I discuss some of my assumptions in greater detail.

#### 6.1 Independent variables

The statistical data on which quantum mechanics is based comes from experiments in which some input parameter is varied and the correlations between inputs and outputs are observed. Of course it is always possible that the input and output are together a function of some third set of variables, so the apparent statistical dependence results from a common cause relationship, but in this paper I have assumed that the measurement settings for quantum mechanical experiments are always among the set of variables of which the outcome is a function, and are independent of the other variables in this set. This assumption should not be too controversial, since assumptions of this kind are ubiquitous throughout science. Indeed, independently chosen variables already play an important role in the contemporary understanding of nonlocality, since there is a meaningful notion of nonlocal correlations only when experimenters can make free choices of experiment in different spatial locations. In particular, it is possible to violate a Bell inequality only by performing measurements in different bases [43], and the derivation of a Bell inequality requires the assumption that the bases can be chosen freely and independently [34]. This assumption is sometimes referred to as free will, [44], but in fact it does not depend on any metaphysical assertions about free will, since it is simply a statement about the statistical relation between certain sets of variables. That said, the usual free will assumption requires that the values of freely chosen variables have implications only in their future light cones, which implies that these variables are uncorrelated with all variables in their past light cones. In this article I do not need to make such a strong unphysical claim: I assume only that the measurement settings for quantum experiments are independent from a specific set of variables, not from everything that has occurred in the past, and therefore the assumption I make is in fact substantially weaker than the widely accepted free will assumption.

Another possible concern is that in a globally deterministic world no two variables would ever be truly independent, since every set of variables would depend on every other set of variables in a global sense. However it is possible for two sets of variables to be statistically independent even if they become dependent when conditioned on some further information (consider for example the phenomenon of hidden nonlocality [43]) and therefore it is entirely consistent to make this assumption in conjunction with the assumption of global determinism.

#### 6.2 Closure

Our derivation requires the assumption that for a given process, the functional relation between the inputs $V$ and the outputs $O$ does not depend on the way in which the inputs are chosen. In particular, I assume that experimenters are free to connect up their boxes in any way compatible with special relativity, and that the boxes function in the same way no matter how they have been connected. This assumption is related to the requirement of closure under wirings, originally introduced in the study of non-local boxes [45], which is not usually regarded as controversial. Moreover, I argue that this assumption and the independence assumption discussed earlier are together the minimal assumptions which are consistent with the possibility of making meaningful counterfactual assertions about how our measurement results would have been different under different circumstances, and the scientific method depends crucially on this possibility.

I also assume a stronger closure principle, namely, the world must remain globally deterministic under all possible compositions of boxes. This assumption is a natural
one if it is accepted that global determinism is a fundamental property of the laws of nature, since then the world must remain globally deterministic under any composition that is actually performed, and thus if experimenters are free to perform any classically allowable composition it follows immediately that the world must remain globally deterministic under the cyclic compositions used in the derivation above.

6.3 Unique outcomes

Finally, I have assumed that when a quantum measurement is performed, only one outcome actually occurs. This is certainly the common-sense view, but it is not true within the Everett interpretation of quantum mechanics, and therefore the Everett approach is not compatible with this way of explaining the information-theoretic features of quantum theory. If the explanation is regarded as a plausible one, this might be viewed as an argument against the Everett interpretation, since the apparent fine-tuning of the no-signaling principle remains unexplained within the Everett picture.

7 Conclusion

I have given an argument to the effect that in a globally deterministic universe, processes which are spatiotemporally symmetric (with respect to a certain class of agents) must also be non-signaling (with respect to that same class of agents). Given that quantum measurement processes are spatiotemporally symmetric with respect to agents like us, this immediately implies that if the universe is globally deterministic, quantum measurement processes must be non-signaling with respect to agents like us, thus providing a simple, physically realistic explanation for the apparently fine-tuned nature of no-signaling correlations.

The explanation suggested here depends crucially on the specific capabilities of a class of agents, and hence this argument suggests that one might regard the no-signaling property as a perspectival phenomenon. However, I wish to reinforce that this perspectival approach is not intended to be an argument in favour of instrumentalism or antirealism, and it does not require us to suppose that agents are somehow unphysical or specially privileged relative to the rest of the universe. The idea is simply that agents have a certain physical realisation, and thus there are objective facts about which sorts of variables they can and cannot manipulate, which leads naturally to objective constraints on the kinds of phenomena that they can and cannot observe. Indeed, signaling is clearly an agent-relative concept, and hence it seems almost inevitable that the supposed universal prohibition on signaling should in some sense turn out to be a perspectival feature of the experience of agents rather than a truly universal feature of all possible correlations.

A number of directions for future research are suggested by the approach discussed here. A further implication of the paper’s main result is that in a globally deterministic universe, a process which does allow signaling for a certain class of agents must necessarily fail to be spatiotemporally symmetric for that class of agents. It follows that if it is possible to send information forwards in time via some signaling process, then that same process cannot be used to send information backwards in time. Further work would be required to fully cash out this argument, but it seems possible that the prohibition on closed causal loops could thus be used to explain why it is that we can signal only in one temporal direction, which might provide some new insight into the mystery of the arrow of time. I would also suggest that similar methods might be applied to explain other apparently fine-tuned features of quantum physics. For example, one might hope to derive the principle of information causality and thus obtain the Tsirelson bound [24].

References


