Many Worlds as Anti-Conspiracy Theory: Locally and causally explaining a quantum world without finetuning

Siddharth Muthukrishnan

University of Pittsburgh
siddharth@pitt.edu

November 5, 2024

Draft of paper for the collection *Everett and Locality*, editor: Alyssa Ney, *forthcoming* with Oxford University Press.

Abstract

Why are quantum correlations so puzzling? A standard answer is that they seem to require either nonlocal influences or conspiratorial coincidences. This suggests that by embracing nonlocal influences we can avoid conspiratorial fine-tuning. But that's not entirely true. Recent work, leveraging the framework of graphical causal models, shows that even with nonlocal influences, a kind of fine-tuning is needed to recover quantum correlations. This fine-tuning arises because the world has to be just so as to disable the use of nonlocal influences to signal, as required by the no-signaling theorem. This places an extra burden on theories that posit nonlocal influences, such as Bohmian mechanics, of explaining why such influences are inaccessible to causal control. I argue that Everettian Quantum Mechanics suffers no such burden. Not only does it not posit nonlocal influences, it operates outside the causal models framework that was presupposed in raising the fine-tuning worry. Specifically, it represents subsystems with density matrices instead of random variables. This allows it to sidestep all the results (including EPR and Bell) that put quantum correlations in tension with causal models. However, this doesn't mean one must abandon causal reasoning altogether in a quantum world. When decoherence is rampant and there's no controlled entanglement, Everettian Quantum Mechanics licenses our continued use of standard causal models. When controlled entanglement is present—such as in Bell-type experiments—we can employ recently-proposed quantum causal models that are consistent with Everettian Quantum Mechanics. We never need invoke any kind of non-local influence or any kind of fine-tuning.

1 Introduction

Particularly puzzling are certain correlations that appear in certain quantum experiments—the so-called EPR/Bell correlations (henceforth, simply *Bell correlations*).[1] What is puzzling about these correlations is what they seem to demand from any explanation of them. It seems as if, to explain them, we must invoke nonlocal influences or invoke delicately selected coincidences. Nonlocal influences are unpalatable since they would conflict with relativity. Delicate coincidences are unpalatable on broad methodological grounds.

This is a standard way of phrasing the puzzle of quantum correlations, and it presents the problem of quantum correlations as a dilemma: Either we admit nonlocal influences (as Bohmians or collapse theorists do) or we admit that our experimental settings or outcomes are unavoidably fine-tuned (as superdeterminists and retrocausalists do). However, the recent work of Wood and Spekkens (2015) employs the framework of causal models (Spirtes, Glymour, and Scheines 2000; Pearl 2000) to show that this isn't really a dilemma, for we can't avoid the fine-tuning horn of the dilemma by accepting the nonlocality horn. Even if we admit nonlocal influences, some sort of fine-tuning persists. This suggests that the main puzzle posed by quantum correlations is that they require a kind of fine-tuning, no matter what.[2]

How does Everettian Quantum Mechanics (a.k.a. the Many Worlds Interpretation) fit into this dialectic? Everettians will argue that they can explain Bell correlations without invoking nonlocality.[3] But what about the fine-tuning objection? While much has been written about how the Everett interpretation can avoid nonlocality (or indeed whether it does),[4] to the best of my knowledge nothing has been written about how or whether Everett avoids fine-tuning. This is important to engage with because if Everett falls prey to a fine-tuning objection despite requiring only local influences, then, all else equal, it is not clearly better than superdeterminist and retrocausalist views at explaining Bell

1. The name deriving from the work of Einstein, Podolsky, and Rosen (1935) and Bell (1964).

2. Those who would argue for the presence of nonlocal influences will say, however, that the amount of fine-tuning required when nonlocal influences are present is far lesser than the amount of fine-tuning required without them.

3. Some defenders of non-realist views about quantum mechanics, such as QBism and Pragmatism (see Healey (2023)), will also argue that they can explain Bell correlations without appeal to nonlocality. However, in this essay, I will set aside non-realist views and focus on realist approaches.

4. See, e.g., the papers in this volume.

correlations, which also preserve locality but admit fine-tuning.[5]

So, does Everett avoid the fine-tuning problem along with avoiding the nonlocality problem? I will argue that Everettian quantum mechanics (EQM) does not face a fine-tuning challenge. This is because of something much stronger: Everettian quantum mechanics avoids the fine-tuning challenge because it rejects the core principles of the causal-modeling framework that lead to the fine-tuning challenge in the first place.[6]

But is that too high a price to pay? The causal modeling framework is a powerful and valuable framework to represent, analyze, and discover causal structure. If EQM demands we jettison it, that might be a net weakness of EQM, despite whatever benefits such jettisoning confers concerning fine-tuning. But I will argue that this isn't a worry for EQM. For one, I will argue that EQM does not require us to abandon the classical causal modeling framework in all contexts, but only in contexts with controlled entanglement, such as the contexts that lead to Bell correlations; decoherence licenses the use of classical causal models in most ordinary contexts, thus allowing us to retain that successful and well-tested framework.

For another, I will argue that in the contexts where controlled entanglement is present, there is another framework that allows us to represent and analyze causal relations, namely the framework of quantum causal models recently developed by Allen et al. (2017) and Barrett, Lorenz, and Oreshkov (2021), which is compatible with EQM, by virtue of being compatible with pure unitary quantum mechanics. It is unclear how a framework like this could be available to non-Everettians.

Let me also remark that while this framework was motivated by the result of Wood and Spekkens (2015), they did not explicitly articulate how their model explains Bell correlations. Further, they did not engage with the question of how one might realistically interpret their framework. In my paper, I advance the discussion in these directions.

Taken together, I will argue that EQM provides non-fine-tuned explanations of quantum correlations, traffics only in local interactions, licenses the use of the standard causal modeling framework in most classical contexts, and fits well with a quantum causal modeling framework when controlled entanglement is

5. Of course, all else is not equal, and there are other reasons one might favor or disfavor Everett over superdeterminism or retrocausalism.

6. It's interesting to note that the way in which a multiversal picture avoids fine-tuning here is quite different from the way in which a multiversal picture avoids fine-tuning in the context of fine-tuning of physical parameters for life or in the context of potential violations of naturalness in particle physics. See Friederich (2023) for more on these kinds of fine-tuning.

present. This is a constellation of virtues that EQM enjoys. While I don't argue in this paper that rival interpretations of QM don't or can't enjoy the same constellation of virtues, it is hard to see how they will be able to. At any rate, my goal in this paper is primarily to highlight the virtues of EQM with regards to causal explanation and non-fine-tuning. I leave to future work the question of comparing EQM with rival views.

Before we get there, though, I'll need to motivate why we should at all care about how well the causal modeling framework fits with quantum mechanics. So, after briefly introducing the framework of causal models in the next section (Sec. 2), I'll show in the two following sections (Sec. 3 and 4), how the famous arguments of EPR and Bell can be quite naturally phrased as arguments against the inapplicability of physically plausible causal models. This will show that applying the framework of causal models to try and explain quantum phenomena isn't at all new or alien—rather, it is a long tradition among philosophers and physicists going back at least to Einstein. This provides the motivation for us to take seriously the failure of faithfulness in quantum mechanics (Sec. 5). Further, given the problems faced of attempts to satisfactorily causally model quantum correlations, we have greater motivation to consider truly quantum causal models.

2 A very brief introduction to graphical causal models

Graphical causal models (Spirtes, Glymour, and Scheines 2000; Pearl 2000) are a powerful framework to represent, analyze, and understand causal relations. They provide a clear mathematical representation of causal relations between variables, they support an interventionist semantics (Woodward 2003), and they greatly aid in discovery of causal explanations (see, e.g., Malinsky and Danks (2018)). So, if there's a phenomenon which we are trying to give a causal explanation of, then it makes sense to try to represent that phenomenon using a causal model. Quantum correlations are one such family of phenomena. My goal in this section is to give a very brief introduction to the framework of classical causal models before we apply it to quantum correlations in the following sections.

The key structure used to represent causal relations in this framework is a *directed acyclic graph*, and we will soon encounter examples of such graphs. The variables in the graph represent the different systems or degrees of freedom that we think are causally related to each other. An arrow represents a causal influence between the variables. We are here interested in *probabilistic* causal models. Thus, we associate a probability distribution over the values the variables of the graph can take. So if the vertices of the graph are $V = \{A, B, C, \dots\}$, then we have a joint distribution over the variables $P(ABC\dots)$. From the joint distribution we can obtain the distribution of any individual variable or calculate the correlations between variables. The structure of causal influences in the graph—represented by the arrows—constrains the structure of conditional probability relations between the variables in the graph. Specifically, there are two central constraints one places on these causal models: the Causal Markov Condition and Faithfulness.

Causal Markov condition (CMC).—This condition states that for any variable X in a causal model, conditional on its parents (i.e., its direct causal ancestors in the graph), X is independent of all the other variables in the model that are not descended from it.[7] Formally,

$$P(X|\text{parents}(X)\&\text{nondescendants}(X)) = P(X|\text{parents}(X)). \tag{1}$$

Roughly, the idea behind this condition is that all the causal influences on a given variable X are laundered through the X's immediate parents. If one can causally manipulate X's parents, then one gains no more causal mileage by being able to manipulate any other variable that isn't causally downstream of X.

It is also useful to have in mind another common formulation of CMC, which codifies the idea that the probability distribution over all the variables can be obtained by chaining the results of variables being influenced by their immediate parents. Formally, for a collection of random variables $\{X_1, X_2, \dots, X_n\}$ featuring in a causal model, this formulation of the CMC asserts that

$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i|\text{parents}(X_i)). \tag{2}$$

This is called the *factorization* formulation of the CMC, and it is equivalent to the screening-off formulation, Eq. (1).

Why believe the CMC? One might argue that it follows from what it even means for a collection of variables to stand in relations of causal influence (or lack thereof) with each other. It can also be seen as a generalization of Reichenbach's Common Cause Principle (Reichenbach 1956), and so deriving its plausibility

7. This is the so-called *screening off* formulation of the CMC.

from that principle. A different line of argument for the CMC is given by Pearl (2000), who argues that it follows if causal relations we see between random variables follow underlying deterministic causal relations along with some noise. Yet another line of argument is provided by Hausman and Woodward (1999, 2004), who argue that the idea that causes can be used to *manipulate* the effects lends support to the CMC. It also has a proven track record: it has been central to the powerful framework of causal modeling pioneered by Spirtes, Glymour, and Scheines (2000) and Pearl (2000). More pertinent to our purposes, we will see that we can interpret Einstein, Podolsky, and Rosen (1935) and Bell (1964) as arguing for instances of the CMC (though they don't think of it in those terms) via appeal to specific physical features of the system in question—features such as spatial separation and connection to random sources. Thus, in these contexts, the CMC is often be motivated by what Weinberger, Williams, and Woodward (2024) call "worldly infrastructure".

Let's turn now to the other central constraint that the causal graph places on the probability distribution over variables.

Faithfulness.—Faithfulness might be thought of as the dual of the CMC. While the CMC says that there should be no more *correlations* between variables than what we would expect from the causal diagram, faithfulness says that there should be no more *independences* between variables than what we would expect from the causal diagram. The CMC tells us that a variable will become independent of any nondescendant once we condition on its parents, but stays silent about the relations between a variable and its parents and descendants. Meanwhile, faithfulness tells us that the CMC encodes all the independences in the causal network: it says there should no more independences than what you expect from CMC.

For instance, applying the faithfulness condition to the causal graph depicted in in Fig. 1 would lead us to expect that X and Y are not independent because they have an arrow between them. The CMC does not deliver an independence between them; faithfulness says that, consequently, that they are not independent. (Note that the CMC doesn't make X and Y independent even conditional on Z because even though Z is a common cause, X is still a descendant of Y, and hence not a nondescendant.) Thus, if the distribution over XYZ is to be faithful to this causal model, then X and Y cannot be independent, conditional or otherwise. However, if X and Y are in fact independent, then the distribution is said to be unfaithful.

Distributions that violate faithfulness but respect the CMC are obtained

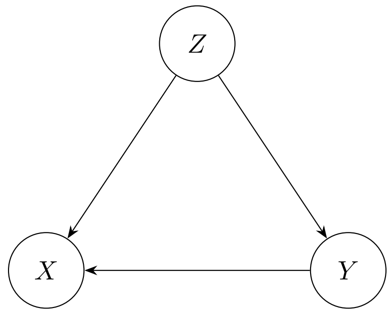

Figure 1: A causal model that is unfaithful to—i.e., fine-tuned for—a distribution in which X is probabilistically independent of Y.

by *fine-tuning*. That is, if we consider some parameters governing the relation between the variables (i.e., the relation between the values of the variables or the relation between the probabilities over those values) in the causal model, then faithfulness-violating values of these parameters will need to be very specially selected.[8] Furthermore, small deviations away from these special values will almost always restore faithfulness.

In sum, the CMC and faithfulness can be taken together as encoding the requirement that the probability distribution over the variables *respect the structure of the causal diagram*, containing all and only those correlations that the causal network leads to.

Having set out the main ideas of graphical causal models, we will now apply this framework to quantum correlations. We will see, in the following two sections, that the classic arguments of Einstein (a version of which was made famous in Einstein, Podolsky, and Rosen (1935)) and Bell can be naturally seen as ruling out certain classes of causal models as potential explanations of quantum correlations. We look at these arguments so as to motivate the applicability of the causal modeling framework to the puzzles of quantum correlations. With this motivation in place, we can then turn (in Sec. 5) to the recent result of Wood and Spekkens (2015), and see how attempts to explain quantum correlations using the causal modeling framework sketched above invariably violates the faithfulness condition.

8. More precisely, Spirtes, Glymour, and Scheines (2000, pp. 41-42) argue that for natural parametrizations (i.e., linear parametrizations) of any causal model, the faithfulness-violating settings of the parameters will be measure zero for any measure that's absolutely continuous with respect to the Lebesgue measure over the possible parameters. See Weinberger (2018) for a discussion of how to think about these measures over parameters and for a defusal of certain purported counterexamples to the faithfulness requirement.

3 Einstein/EPR and the Causal Markov Condition

In 1935, EPR argued that quantum mechanics (QM) is incomplete (Einstein, Podolsky, and Rosen 1935). The true essence of their argument was developed by Einstein in 1927 (see Howard (1985) and Harrigan and Spekkens (2010)) and is simpler than what was presented in EPR paper and can be presented as follows. To begin, suppose we have two spin-$\frac{1}{2}$ particles in the entangled singlet state, i.e., $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$.[9] According to QM, if I measure the spin of the first particle and see that it comes "↑" (or "↓"), then I immediately know that a measurement of the second particle will yield "↓" (or "↑") . Moreover, this will continue to be true even if the two particles are taken far enough away from each other so as to ensure that no signals, travelling at or below the speed of light, can get from one particle to the other within the time-frame of the measurements.

But this is in tension with the following observation. If we just *look* at the our representation of the state $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$, then we see nothing there that tells us whether a measurement of either of the particles will yield "↑" or "↓"; for both particles, the representation is symmetric between those two outcomes. Given this symmetry between the two outcomes, it's puzzling how we are able to *immediately* come to know that the state of the second particle is "↓" upon measuring the first particle to "↑". About what feature of the world have we acquired knowledge? If the answer is that we have come to know that the state of the second particle was "↓" all along, then we have conceded that the singlet state does not successfully represent all the features of the world in this experiment, because clearly it does not privilege "↓" over "↑". Consequently, quantum mechanics' representation of the world would be incomplete: there would be more properties of systems than quantum mechanics represents it as having.

One might argue that our measurement of the first particle *caused* the second particle to change from being in a state that was symmetric between "↑" and "↓" to being in the state "↓". However, this would imply that there are superluminal—indeed, instantaneous—influences between the two particles. After all, our ability to *immediately* know the state of the other particle upon measurement of one particle is unaffected by how far the particles are. To better understand this point, let's ask why we cannot construct an EPR-style argument with just a

9. As is standard, I have adopted Bohm's spin-version of the argument.

non-entangled superposition such as $\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$. One might think, following the argument in the previous paragraph, that this non-entangled superposition state is also symmetric between "↑" and "↓", but a measurement of a system in such a state returns either a definite "↑" or a definite "↓", suggesting that QM is incomplete. However, this argument is susceptible to the objection that the definite outcome of "↑" or "↓" was *caused* by the interaction of the system with the measuring apparatus, instead of revealing a preexisting property of the system. In contrast this objection isn't applicable in the EPR-style case since the first measurement is, if relativity is to be believed, causally disconnected from the first. So, it seem as if the only way we could come by knowledge of the second particle's state is if our measurement of the first particle revealed to us a preexisting property of the second particle. But this property isn't represented by the quantum formalism. Consequently, the quantum mechanical formalism must be incomplete. This was the essence of Einstein's argument.

Now let us phrase this in causal modeling terms.[10] We can think of Einstein as arguing that a certain kind of causal model is insufficient for explaining quantum phenomena. But which kind of causal model? The natural candidate is the causal model with the causal graph represented in Fig. 2. In this figure, A and B represent the possible values that the measurements on the two particles may yield, and they take values over $\{\uparrow, \downarrow\}$. S and T are variables that represent the measurement settings on the two wings of the experiment. In the development above, we considered only one possible measurement setting; hence, S and T only take the value of "measure z-spin". λ_A and λ_B are local variables that help fix, perhaps only probabilistically, the values that A and B take; these variables can be seen as encoding whatever local property the quantum state attributes to the particles.

Now if we apply the causal Markov condition to this graph, then we should expect the following independences:

(i) A is independent of B, λ_B, and T conditional on S and λ_A; and B is independent of A, λ_A, and S conditional on T and λ_B.

(ii) S, T, λ_A, and λ_B are all independent of each other.

10. Phrasing EPR-style arguments in causal modeling terms isn't new; see, e.g., Van Fraassen (1982) and Hausman (1999) for some classic treatments, and see, e.g., Suárez and San Pedro (2010) and Näger (2016) for more recent treatments. While I make no great claim to originality, my way of presenting this topic isn't quite the way other authors present it, and so I hope this presentation is of value. In any case, I present this material here so as to have a unified and coherent narrative, even if that runs the risk of repeating material that might be well-known to certain readers.

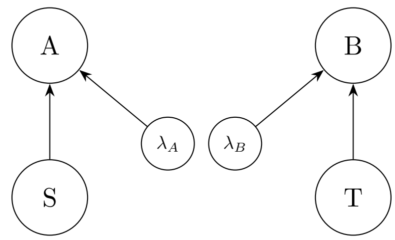

Figure 2: Einstein's argument rules out the possibility of this kind of causal model reproducing the predictions of quantum mechanics.

These independences can be easily obtained by applying the screening-off formulation of the CMC [Eq. (1)]. For instance, because λ_B is a non-descendant of λ_A, and λ_A has no parents, then Eq. (1) implies $P(\lambda_A|\lambda_B) = P(\lambda_A)$, which is equivalent to $P(\lambda_A\lambda_B) = P(\lambda_A)P(\lambda_B)$.

But why believe that this is the right graph and that the CMC is applicable to this graph? We can see the key premises of Einstein's *reductio* as motivations for the CMC applied to this graph. The first key premise of Einstein's argument is that the two wings of the experiment cannot be causally connected since the two wings can be separated arbitrarily far away, and there can be no signal that can travel between them during the time-frame of the experiments. This motivates both the structure of the causal graph and the causal Markov condition applied to it. In particular, it motivates the lack of any arrows from the one wing to the other, since there aren't any causal influences between the two wings. Further, it motivates the conditional independences specified in statement (i) above because the lack of causal influence between the two wings suggests that whatever happens on one side should be self-sufficient to causally explain what happens there.

The second key premise of Einstein's argument is that the entangled quantum state is indifferent between the two possible definite outcomes and that it is representationally complete. So if the quantum state is representationally complete, then whatever determines the outcomes of the quantum measurement on either wing should be indifferent between $\uparrow$ and $\downarrow$, and so λ_A and λ_B should assign equal probability to those two possible outcomes. And since they are causally disconnected, then they will be completely uncorrelated, which is what the CMC requires.[11]

11. The independence of λ_A and λ_B with S and T is trivial since S and T can only take one possible value.

Let now see, informally, why the causal graph of Fig. 2, combined with the CMC applied to that graph, is incompatible with what we see from quantum experiments. Quantum experiments show that A and B are always perfectly correlated. The causal Markov condition, in slogan form, asserts that there be no more correlations than what we expect from the structure of the graph. And since the two wings of the experiment are entirely independent of each other, the CMC tells us that A and B should be independent. And if they are not, as we see in the quantum experiments, then the CMC is violated.

More formally, applying the formulation of the CMC given in Eq. (2) to the graph, we get

$$P(ABST\lambda_A\lambda_B) = P(A|S\lambda_A)P(S)P(\lambda_A)P(B|T\lambda_B)P(T)P(\lambda_B). \tag{3}$$

Summing over S, T, λ_A, and λ_B, we get:

$$P(AB) = P(A)P(B), \tag{4}$$

i.e., A and B are uncorrelated. And this contradicts quantum experiments.

Faced with this contradiction, it seems as if one of the two key premises of the argument must be rejected. Given the success of relativity theory, Einstein and EPR think that it's untenable to reject the premise that the two wings of the experiment are causally disconnected. Thus, they reject the premise that the quantum state is representationally complete.

4 Bell and the Causal Markov Condition

The Einstein/EPR argument suggests a natural follow-up project. If QM is local but representationally incomplete, then perhaps we can come up with a local *completion* of QM, i.e., one that contains elements that represent those properties of systems that were left out by the quantum mechanical formalism (those elements about which we gained knowledge in an EPR-type set up so that we could immediately and with certainty come to know the outcome on the other wing) while still trafficking entirely in local influences. These elements are usually called *hidden variables*, and they would fix—perhaps only probabilistically—the definite outcomes seen in physical systems.

The language of causal models allows us to frame this project more precisely. I argued in the previous section that the kind of causal model that Einstein

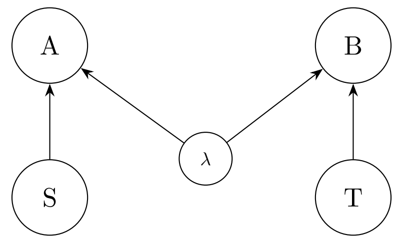

Figure 3: The violation of the Bell inequality rules out the possibility of this kind of causal model reproducing the predictions of quantum mechanics.

argued against is the one depicted in Fig. 2, along with CMC applied to it. I said that the lack of correlation between λ_A and λ_B in that diagram can justified by appeal to (a) the spatial separation between the two wings of the experiment and (b) the fact that the quantum state contains no preference for one outcome or the other. However, one might think that because entangled particles *started out* close together—perhaps they are electrons taken from the same Helium atom—we should not assume that the two wings of the experiment have variables entirely causally disconnected from each other. This thought leads us naturally to consider the causal diagram of Fig. 3, and ask whether such a causal diagram can reproduce the statistics seen in the kinds of quantum experiments considered by Einstein.

The famous results of Bell (Bell 1964; Bell 1975) showed that this question has a negative answer. In particular, Bell proved an inequality that the measurement statistics of *any* theory conforming to the schema of Fig. 3 must satisfy. However, these inequalities are *violated* by the measurement statistics predicted by QM. And these predictions have since been repeatedly empirically verified, by experiments performed with increasing carefulness.[12] Thus, Einstein's hope of providing a local completion of quantum mechanics is a dead end.

We will now see how the core assumption that Bell made in proving his theorem is closely connected to the causal Markov condition applied to the graph of Fig. 3. Much like the previous case, the CMC will be justified here by the physics of the experimental situation.

To begin let's unpack the causal graph of Fig. 3. Analogous to the EPR-style case, S and T are random variables representing the measurement settings on the left and right wings of the experiment respectively; A and B are random

12. See, e.g., Hensen et al. (2015), Giustina et al. (2013), and Shalm et al. (2015) for some recent carefully conducted experiments.

variables representing the measurement outcomes on either wing; and λ represents whatever common element that is shared between the two wings of the experiment and which may influence the outcomes of measurements on the two sides. Standard presentations of Bell's theorem usually take S and T as taking values over two possible measurement settings and A and B as taking values over two possible measurement outcomes. However, the set up is more general, and one can readily consider situations with a greater number of possible measurement settings and possible outcomes for each measurement and prove Bell-type theorems for such situations.

Note here that λ could be a high-dimensional variable, and need not just be a single number. As we are trying to come up with a completion of quantum mechanics after being moved by Einstein's argument, we take λ to represent that feature of the world about which the experimenter in one wing acquires knowledge of, by measuring their own system, which then allows them to predict, immediately and with certainty, the result in the other wing.

The key premise in Bell's theorem is what is often called *factorizability.*[13]. In our notation, this is the statement that:

$$\Pr(AB|ST\lambda) = \Pr(A|S\lambda)\Pr(B|T\lambda). \tag{5}$$

From this assumption, it is straightforward to derive the CHSH inequality, which is a simpler version of the Bell inequality.[14]

The factorizability condition is a straightforward consequence of the CMC applied to Fig. 3. To see this, first note that applying the screening-off formulation [Eq. (1)] of the CMC to S, T, and λ entails that the three variables are mutually independent. This entails that

$$\Pr(ST\lambda) = \Pr(S)\Pr(T)\Pr(\lambda). \tag{6}$$

Now, applying the factorization formulation of the CMC [Eq. (2)], we have:

$$\Pr(ABST\lambda) = \Pr(A|S\lambda)\Pr(B|T\lambda)\Pr(S)\Pr(T)\Pr(\lambda). \tag{7}$$

Writing the LHS as $\Pr(AB|ST\lambda)\Pr(ST\lambda)$, then cancelling $\Pr(S)\Pr(T)\Pr(\lambda)$ on both sides by using Eq. (6), we get the factorizability condition, Eq. (5).

How do we justify the CMC in this context? As in the EPR-style case, we can

13. See, e.g., Myrvold, Genovese, and Shimony (2021, Sec. 3.1) for more on this.
14. See, e.g., Brunner et al. (2014, p. 3) for details of the derivation.

justify it by appealing to the physical situation. The CMC in this context is the conjunction of the following independences, which can be obtained by employing the screening-off formulation of CMC [Eq. (1)]: $A \perp\!\!\!\perp BT|S\lambda$, $B \perp\!\!\!\perp AS|T\lambda$, $S \perp\!\!\!\perp BT\lambda$, $T \perp\!\!\!\perp AS\lambda$, and $\lambda \perp\!\!\!\perp ST$.[15]

The first two independences (i.e., $A \perp\!\!\!\perp BT|S\lambda$, $B \perp\!\!\!\perp AS|T\lambda$) can be justified by appealing to the fact that there are no direct influences between the two wings of the experiment. In the model of Fig. 3, this constraint is represented by the absence of causal pathways between the variables on the left wing and the variables on the right wing. In our experimental setup, we can enforce this constraint by the spacelike separation of the measurements on the two wings. This leads us to expect that the measurement settings and measurement outcomes in one wing of the experiment are independent from the settings and outcomes on the other wing, *except* for those correlations that can be accounted for by the shared features (which are encoded in λ) that might arise from past interactions between the systems.

The other independences (i.e., $S \perp\!\!\!\perp BT\lambda$, $T \perp\!\!\!\perp AS\lambda$, and $\lambda \perp\!\!\!\perp ST$) can be justified by the fact that the measurement settings on each wing are determined by processes that are disconnected from each other and from the history and physics of the system under study. We can enforce this by appropriately setting up our experiment. For instance, in a recent experiment, the measurement settings were set by photons arriving from parts of universe that have been causally disconnected for billions of years (Rauch et al. 2018). This ensures not only that the two measurement settings are entirely uncorrelated with each other, but also uncorrelated with any variable (such as λ) describing the system under study. Of course, we needn't go to such lengths to satisfy the independence condition. Choosing measurement settings randomly might suffice, say by coin tosses. As might whatever processes—biological, mental, personal—that cause experimenters choices of measurement setting, for we have no reason to believe that their processes are correlated with the hidden variable under study or with each other.[16] We might be less certain that processes governing coin tosses or experimenters making choices are uncorrelated with the system in question than that cosmologically separated photons are uncorrelated with the system in

15. Notation: $\alpha \perp\!\!\!\perp \beta\gamma\delta\ldots|\chi\psi\omega\ldots$ means that α is independent of any subset of $\{\beta,\gamma,\delta,\ldots\}$ conditional on *all* of $\chi\psi\omega\ldots$. If there's no third entry, then it's an unconditional independence between α and any subset of $\{\beta,\gamma,\delta,\ldots\}$.

16. This is why it is sometimes said there's a "free will" assumption in Bell's theorem. But this is a misnomer; we need no sophisticated thesis about free will, only an assumption about the probabilistic independence of certain processes.

question. But we can still be *very* certain, for it is hard to articulate what sort of physics would lead to a violation of these independences outside of implausible stories involving conspiracies.[17] But this also makes it hard to articulate clear objections against such physics, for we don't have a clear target for criticism. That said, we will see in the next section what theories violating these independences will look like within the framework of causal models, and we'll see that there's a broad methodological objection to such independence-violating theories within that framework.

Given all this, the experimental violation of the Bell inequality amounts to a crisis for causal explanations of quantum mechanical phenomena. The CMC applied to Fig. 3 seems to encode what we are enforcing in our experiments using known physics, namely, locality and the independence of measurement settings from the systems under study. Thus, it seems as though if we want a causal explanation of Bell inequality violations, we must be willing to abandon one of these assumptions, even if it conflicts with physics we think we know. This is what leads to a standard articulation of the puzzle posed by quantum correlations, which is that we are faced with a dilemma: Either abandon locality or accept that there are conspiratorial coincidences plaguing our experiments.

However, as we shall see in the next section, this is a false dilemma. We have reason to believe that even if we abandon locality, there might still be a conspiratorial aspect required in the resulting explanation of quantum correlations, threatening causal explanations of these correlations.

5 Wood and Spekkens and Faithfulness

Faced with the violation of Bell inequalities, we infer, then, that if we want a causal explanation of quantum correlations, then we must edit the causal model of Fig. 3. However, given the physical arguments for that causal model, these edits would require the violation of some cherished physical principle or other.

How might we edit the causal model to account for Bell-inequality-violating correlations? Suppose, as the Bohmian and spontaneous collapse theories do, that we allow for there to be superluminal influences.[18] This allows arrows from (say) the left wing of the experiment to the right wing of the experiment, as in Fig. 4. Such arrows clearly violate the empirically well-supported principle

17. See Shimony, Horne, and Clauser (1976) for an articulation what a conspiracy intended to establish certain correlations might look like.
18. See, e.g., Maudlin (2019) for an overview of these interpretations.

arising from the special theory of relativity that physical influences cannot travel at superluminal speeds. However, one might conclude, as some have done, that empirical Bell inequality violations just show us that this principle, well-supported though it might be, is simply false.[19]

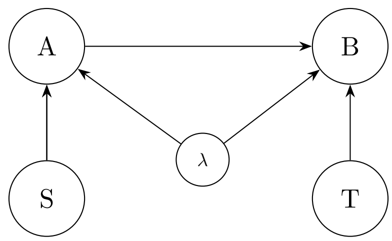

Figure 4: A causal diagram allowing for superluminal influences. These kinds of causal models have to be fine-tuned (i.e., violate faithfulness) if they are to preserve the no-signalling criterion.

A different strategy to edit the causal model of Fig. 3 is the one adopted by superdeterminists, who embrace the idea that the measurement settings on the two sides of the experiments are not truly freely chosen, i.e., the measurement settings are such that they induce the relevant correlations between the measurement outcomes despite every effort to ensure free selection of measurement settings. Superdeterminist theories can be represented by a causal diagram, such as Fig. 5, which contains influences between λ and the measurement settings S and T, along with influences from λ to A and B.

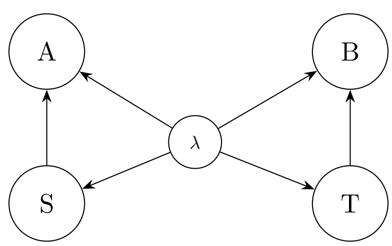

Figure 5: A causal diagram allowing for superdeterminism, which prevent free choice of measurement settings. These kinds of causal models also have to be fine-tuned (i.e., violate faithfulness) if they are to preserve the no-signalling criterion.

I won't here examine the plausibility of such models on their own terms. Instead, I want to point out that such edits don't provide a fully satisfactory

19. See Albert and Galchen (2009) and Maudlin (2014) for discussions that take this line.

causal explanation of quantum correlations because they're *fine-tuned*; specifically they violate *faithfulness* (Sec. 2). To see this, let's start with the models that allow nonlocal influences, such as Fig. 4. The main idea is simple. Such models will have to reproduce the no-signaling condition—i.e., the measurement outcomes on one side have to be probabilistically independent of the measuring settings on the other. However, if one admits superluminal influences, then one has to fine-tune the physics just so as to prohibit the use of these influences to signal. This fine-tuning is why faithfulness is violated in such models.

Specifically, from Fig. 4, we can see that the probabilistic independence of the measurement settings on one wing and the measurement outcomes (i.e., the no-signaling condition) on the other wing will be unfaithful to the causal structure. This is because there is a causal pathway from S to B, via A. If the distribution over the variables were faithful to this causal graph, then there would be no independence between S and B. For, recall, that the faithfulness condition says that there ought to be no more independences than specified by the CMC. However, given the no-signaling theorem, we must have an independence between S and B, because otherwise we could manipulate S to influence B, sending a signal. Such an independence is also empirically verified. Hence, models with nonlocal influences must be unfaithful to their causal structure.

Turning now to superdeterminist theories, the main idea is clear here as well. Such theories require fine-tuning, because microphysics would have to be in an extraordinarily specific state so that just these certain measurement settings occur in all these different and varied types of entanglement experiments and data. This need for fine-tuning is reflected in the failure of faithfulness for superdeterministic causal models. Consider Fig. 5. In experiments, we clearly manage to make S be independent of T; the statistics of measurement settings are clearly such that $P(ST) = P(S)P(T)$. However, this is not an independence that is delivered by the CMC applied to this graph. The CMC applied to this graph only tells us that S and T are independent *conditional on* λ, while what we observe is an unconditional independence. And so because we have more independences than delivered by the CMC, any empirically adequate superdeterministic model must violate faithfulness.

Let me emphasize at this point that I don't intend to suggest that superdeterministic theories are in any sense on a par with theories that require nonlocal influences. For one, theories such as Bohmian mechanics or spontaneous collapse theories are much more precisely specified than any superdeterministic proposal so far, and so enjoy theoretical virtues that have nothing to do with fine-tuning.

For another, even on the count of fine-tuning, it is likely that a theory such as Bohmian mechanics fares much better than a generic superdeterministic theory. For, the Bohmian can argue that the Born rule probability, though needing to be specially selected, is relatively straightforwardly specifiable and enjoys certain natural properties within quantum theory.[20] Meanwhile, it is unclear if any superdeterministic theory can provide a simple, physically plausible specification of the joint distribution over the variables in an experiment.[21]

So it seems if we try to get an empirically adequate causal model of quantum correlations—even one that violates well-established physical principles such as relativistic locality or independence of systems and experimental settings—we have to fine-tune the model in some way or other. But could that be just a symptom of not having considered sufficiently sophisticated causal models? After all, the diagrams sketched above are but some in a vast space of possible models. The answer is no. Wood and Spekkens (2015) prove a theorem that shows that no faithful classical causal model can reproduce quantum correlations. This shows us that even if we are willing to consider novel physics so as to explain quantum correlations, we will not be able to provide a full-blooded causal explanation of these correlations, because they will necessarily violate faithfulness.[22]

6 Where do we go from here?

Taken together, the arguments above lead us to question whether we can causally explain quantum correlations. It looks as if Bell-type correlations do not really fit with the empirically and methodologically well-motivated framework of causal models. We want the probabilistic relations between the variables in our theories to mirror the structure of the causal relations between them (i.e., to satisfy the CMC and faithfulness), but the arguments of EPR, Bell, and Wood & Spekkens tell us that we can't achieve this.

So, if that hope is dashed, then where do we go? One strand in the literature either took the results of EPR and Bell as undermining the CMC[23] or as evidence

20. See, e.g., Goldstein and Struyve (2007).

21. For a recent philosophical critique of superdeterministic theories, see, e.g., Baas and Le Bihan (2023).

22. Cavalcanti (2018) strengthens the Wood-Spekkens result by unifying it with Kochen-Specker theorems.

23. See, e.g., Van Fraassen (1982), who takes Bell's theorem to refute Reichenbach's principle of common cause (an important justification for the CMC), and hence to refute an argument in favor of epistemic realism.

for the inapplicability of the CMC in quantum contexts.[24] Such responses are plausibly in line pragmatist/anti-realist views of quantum mechanics.[25] For, if one does not take the quantum state as representing the physical state of the world, then the physics-based arguments for adopting the CMC in the EPR/Bell-type cases (arguments given Secs. 3 and 4) are far less persuasive.

Another line of response, more popular among those with realist views about quantum mechanics, has been to hold on to the CMC, and instead to try and explain why violations of faithfulness are unproblematic according to one's preferred view of quantum mechanics (such as Bohmian mechanics, collapse, or retrocausality).[26]

I won't here evaluate the plausibility of such responses. Instead, I want to suggest a different response. A response which has the following commitments:

1. It is realist in its view of quantum mechanics (unlike the Van Fraassen style response).

2. It rejects the blanket applicability of the standard framework of causal modeling in the context of quantum mechanics (unlike the views that argue why faithfulness violations aren't problematic).

3. But it explains why the apparatus of causal modeling works well in most classical contexts, failing only in quantum contexts.

4. And it still offers a way to causally explain quantum correlations, but it has to be *quantum* causal.

The view I'm proposing, which has the above commitments, is the combination of Everettian quantum mechanics (EQM) and a framework for quantum causal reasoning, recently developed by Allen et al. (2017). This view is realist because EQM is a realist view about quantum mechanics. It rejects the applicability of the standard framework of causal models to quantum phenomena because it represents subsystems using density matrices instead of random variables, which is what the standard framework uses. Further, it explains the usual validity of classical causal modeling in a quantum world via appeal to decoherence: usually there is significant decoherence which means we can get

24. See, e.g., Hausman and Woodward (1999), who argue that quantum correlations aren't so much counterexamples to the CMC, but contexts in which the CMC isn't applicable.

25. A connection made explicitly in the case of Van Fraassen, but not in the case of Hausman or Woodward.

26. See, e.g., Egg and Esfeld (2014), Näger (2016), and Evans (2021).

away with treating our systems using random variables. However, if there is controlled entanglement present then we cannot make this assumption, and standard causal modeling is invalid. But this doesn't mean we need to abandon hope of causally explaining quantum correlations. We just need a framework of causal modeling that can accommodate the way in which quantum mechanics represents subsystems—namely, via density matrices—while still encoding what we think of as important to causal explanations. I argue that the framework developed by Allen et al. (2017) fits this bill. Further, because this framework is compatible with unitary quantum mechanics, it is also compatible with Everettian quantum mechanics. Taken together, I argue that we get a satisfying picture which allows us to causally explain quantum correlations, while avoiding the pitfalls that previous attempts to do so have fallen into.

7 The Everettian explanation of Bell violations

The core idea behind Everettian quantum mechanics (EQM) is that we just need to take seriously the idea that all systems—including measurement devices and humans—are constituted by quantum mechanical parts and thus can enter into, and be part of, superpositions. Thus, when a quantum mechanical system in superposition interacts with another quantum system with many uncontrolled degrees of freedom—such as a measurement device or the system's environment—the latter quantum system enters into a superposition too, governed by the Schrödinger equation. Within the different branches in this larger superposition there emerge multiple quasiclassical *worlds*, stable states of affairs exhibiting approximately classical behavior.[27]

Thus, when a measurement is performed on a quantum system, there obtain multiple branches, with each branch corresponding to a definite outcome. A branch comes equipped with a *branch weight*—which is equal to the mod-squared amplitude associated with that branch. Branch weights are a measure over branches and they determine the probabilities of measurement outcomes.[28]

The key point here that is relevant to the applicability of causal modeling is

27. See (Wallace 2012) for a detailed development and defense.

28. Establishing the connection between this measure and the observed probabilities is by far the most contentious question concerning the Everett interpretation (see, e.g., some of the papers in (Saunders et al. 2010) for a good discussion). I avoid engaging with this issue here, and assume that *some* account of how branch weights ground observed probabilities is successful. Needless to say, if one thinks that EQM is unsuccessful at explaining observed probabilities, then it won't be a viable interpretation, and the rest of what I say here will be largely moot.

that this kind of branching which grounds probabilistically distributed stable outcomes only obtains when the system is *decohered*. It is only then that we have license, in a quantum world, to describe the system using random variables. And the assumption that we can describe systems using random variables is clearly central to the causal modeling framework. So what we see here is that this central assumption of the causal modeling framework is not uniformly permitted by quantum physics. (Note that this point is technically independent of EQM. However, it fits particularly neatly with EQM, for decoherence is crucial within EQM for explaining the emergence of randomness and stability.)

Now, this observation by itself, isn't enough to tell us that the causal modeling framework will fail in explaining quantum correlations. After all, in EPR/Bell-type experiments, the experimental results are really measurement statistics, and so it seems there is enough decoherence to license the use of random variables. In technical terms, what this means is that, post-measurement, in each wing of a Bell-type experiment, we can associate to the system a decohered density matrix that is almost exactly diagonal in the measurement basis, and which has diagonal entries equal to the Born probabilities of various possible outcomes. This diagonal density matrix is what grounds the random variable description of the measurement outcomes in EQM.

The place where the EQM picture starts to deviate from the picture presupposed by causal modeling is that the latter picture assumes that the *joint* distribution over all the variables is always well-defined. We can see this in the way in which constraints like the CMC and faithfulness are formulated: they are ways in which the joint probability distribution over the variables in a causal model are constrained by the structure of the causal graph. However, in a quantum world, because a random variable description is only licensed by a decohered density matrix, we can't always assume a global decohered density matrix is well-defined which licenses describing the systems with a joint probability distribution.

The upshot of this in the context of Bell-type experiments is that even though we have local decohered structures, these structures don't *immediately* ramify up to a global one because there are no instantaneous interactions. Thus, there won't be a globally decohered density matrix simply because we have local decohered density matrices. Consequently, we won't have the license to use random variables to represent the *joint* state of the two wings. This means that we cannot talk about the correlations between the two wings of experiments until and unless the two wings interact in a way that allows for a globally decohered structure to be established—which typically happens when the future light-cones of the two experiments intersect—that then allows us to talk about the correlations between the two sets of outcomes. These are the correlations that violate the Bell inequality. That is, the overwhelming branch weight will be on branches with values of the correlation that violate the Bell inequality. Consequently, observers in such branches will be unable to explain their observations with a causal model, because of the arguments considered in the previous sections. Thus, on an Everettian picture, the violation of the Bell inequality signals the absence of a globally decohered structure at earlier times, not nonlocality or finetuning.

This should also make it clear, we typically get failures of classical causal modeling when it contains entanglement that is *accessible* or *controllable*. To see this, suppose we are trying to build a causal model of some situation that is distant from quantum mechanics. Say the relevant variables are the amount of precipitation in a forest and the population of beavers in that forest. While it is very likely true that the quantum degrees of freedom *constituting* or *grounding* these variables are entangled with each other in all sorts of complex ways, that entanglement is not accessible or controllable. This then means we can continue using the standard framework of classical causal models, because in cases like this we have a globally decohered structure that goes along with locally decohered structures. However, this is not the case for systems whose entanglement is maintained with some care, such as in Bell-type experiments. In such cases, we may have enough decoherence to treat the individual systems as classical random variables but not enough for the joint system to be so treatable. So, outside of such cases, we may safely employ the classical causal modeling framework.

Incidentally, notice here that this explanation of why standard causal models don't apply to Bell-type experiments is not quite the same as an explanation that is sometimes given for why EQM is able to avoid the bite of Bell's theorem, which is that while Bell's theorem assumes that experiments have definite outcomes, EQM tells us that experiments don't have definite outcomes. *Prima facie*, this explanation might not be very convincing since Bell's theorem only concerns relations between random variables and one might think that EQM certainly allows description of systems in terms of random variables, on pain of empirical inadequacy if it did not. There's nothing in that about definite outcomes. In contrast, my way of explicating why Bell's theorem doesn't apply in an Everettian world—namely, that global random variables can't always be defined—*does* explicitly deny a premise of Bell's theorem, namely that joint probabilities are always well-defined.[29]

Let me consider one potential objection against my argument. One might deny my premise that it is only a globally decohered structure that would license the use of random variables. After all, random variables are cheap. You can just define one whenever you want. As soon as the measurements are performed in the two wings, the laws of physics entail that in the future, EPR/Bell correlations will be observed. Consequently, we could define the correlations as obtaining even before the future light cones of the two experiments intersect. It's only that the grounds for these correlations are in the future. But this doesn't logically prevent us from defining a joint distribution before they obtain.

But this objection doesn't much affect the main thrust of my argument, because these kinds of "cheaply defined" random variables cannot be taken as *representing* the physical goings-on. And my key point is that the causal models framework assumes that systems can be *represented* by random variables, and that is the assumption that is violated in Bell-type experiments.

To better understand the importance of random variables *representing* the physical goings-on for successful causal explanation, let us go back to EPR. EPR's argument posed a problem for causal explanation only because we thought that a measurement in one wing *immediately* gave us knowledge of a measurement outcome on the other wing, suggesting the need to posit nonlocal influences or hidden common causes. That is, if we had a random variable that encoded these correlations, then such a random variable creates a problem for a causal explanation only if we think of it as representing the physical goings on at the time of measurements. This then lets us see that if EQM is right, then even if we mathematically define a random variable at the time of measurements that encodes correlations, such a random variable only represents future correlations.

We can see this point in a slightly different way as well. On the Everettian picture, when a measurement is performed, all outcomes with nonzero amplitudes obtain. Hence, when we perform a measurement on one wing, we only obtain knowledge about which outcome we will share our branch with in the future. This is one way in which my no-global-random-variables response connects with the no-definite-outcomes response of the Everettian.

Nevertheless, this doesn't mean that we can just build a regular causal model for the Everettian story of what is going on in Bell-type experiments. For instance, given the account above, one might think that a causal model of

29. This isn't to say that the no-definite-outcomes response doesn't succeed, or to say that it entirely distinct from the no-global-random-variable response. See below.

the sort in Fig. 6 might be made to work. Here, the variable C is assumed to represent the values that the correlations between A and B take. However, this can't be empirically adequate for the same reasons (namely, Bell's theorem) that the causal model of Fig. 3 isn't empirically adequate: within the standard causal modeling framework, the Bell-violating correlations between A and B are established before we get to C, and so we would need nonlocal influences or finetuning.

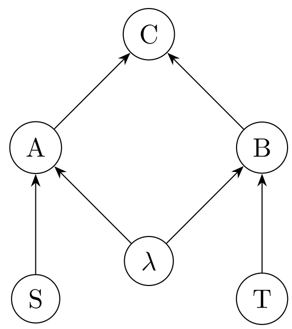

Figure 6: A causal diagram that one *might* think captures the Everettian story of Bell correlations. However, this can't quite work because it still assumes that correlations between A and B are always well-defined.

So far we have a picture that satisfies commitments 1, 2, and 3. The Everettian story about Bell correlations is realist, it rejects the blanket applicability of classial causal modeling since it only licenses the use of random variables in the presence of decohered structures, and it explains why classical causal modeling usually works well: because of the widespread presence of decoherence. So that leaves the question of how we might satisfy commitment 4, and we turn to that question next.

8 Quantum causality and Everett

So, can we causally explain quantum correlations? If one can develop a systematic theory of causal explanation where the relata of causal relations are density matrices instead of random variables, analogous to the powerful systematic theory of classical causal explanation given by Spirtes, Glymour, and Scheines (2000) and Pearl (2000), then that would be a strong start towards causally explaining quantum mechanical phenomena as well.

I will now argue that at least one recently developed framework, due to Allen et al. (2017), of *quantum* causal models fits the ticket. It is a framework for causal reasoning about quantum phenomena that takes as its starting point the idea that what we should represent systems with reduced density matrices.

Not only that, I will that even though it was developed independently of the Everettian worldview, it fits well with it because it is compatible with unitary quantum mechanics, and Everettian quantum mechanics is, at its core, just unitary quantum mechanics. Thus, once we have this framework in our toolkit, we have a story about Bell correlations that simultaneously satisfied commitments 1-4. Furthermore, given that this framework was developed with unitary compatibility as a core assumption, it is unclear how rivals to Everett can quite so easily help themselves to such a framework.

Let me emphasize that Allen et al. (2017) do not take their framework to be Everettian. On my reading, they use the "church of the larger Hilbert space" as a just a tool to derive their framework much like how Pearl (2000) uses the deterministic structural equations framework as a tool to derive his framework of probabilistic causal models. I, however, want to use the unitary-compatibility of the Allen et al. (2017) framework as allowing it to be interpreted in an Everettian way.

The best way to introduce this particular framework of quantum causal models for our purposes is by analogy with the deterministic structural equation modelling (SEM) framework and its relation with probabilistic classical causal models (see, e.g., Hitchcock 2020). Imagine there are *deterministic* functions that govern the behavior underlying the variables in a classical causal model. These relations are arranged in a directed acyclic graph (DAG). For a given SEM graph, suppose we consider a proper subgraph (which is also a DAG). Then we can get probabilities on the variables of the subgraph by averaging over those variables which don't show up in this subgraph. This is similar to how we can see probabilistic behavior in statistical mechanical systems at an emergent level even if the underlying dynamics is deterministic. Mathematically, if we have a deterministic function that is of the form $Y = f(\lambda, X)$, then we can get a distribution $P(Y|X) = \sum_\lambda P(\lambda)f(\lambda, X)$, where $P(\lambda)$ is some distribution (which could arise due to a deterministic process) over the "hidden variable" λ. In Fig. 7, I show in diagrammatic form a deterministic SEM model and the probabilistic causal model derived from it. (The representation I employ here of the SEM and the probabilistic model derived from it is really the dual of their representation as a DAG in the sense that the lines represent variables and the boxes represent the transformations. In what follows, including in the quantum case, I'm going to stick to this dual representation, and draw boxes to represent the transformations on the systems and use wires to denote inputs to the transformations. Doing so unfortunately breaks the visual analogy with the DAGs, but allows for greater conceptual clarity.)

Now, let's turn to the quantum version of such an analysis. Here, the analogue of the deterministic functions of variables is unitary transformations on a Hilbert space. The corresponding analogue of probabilistic model is obtained by tracing over the subsystems of the Hilbert space which we are discarding. A map obtained by taking a unitary map and discarding its action on auxiliary subsystems is called a *quantum channel*, which, mathematically is represented by so-called *completely positive* maps. These maps take density matrices—which are obtained by tracing over subsystems of a quantum wavefunction—to other density matrices.

Quantum channels can also be thought of as the quantum generalization of the classical channels of Shannon's information theory (hence the name "channel"). Classical channels are represented by conditional probability functions: a channel $X \to Y$ is associated to the distribution $P(Y|X)$, which represents the probability of receiving a message $Y = y$ given that the message $X = x$ was sent. Thus, while in the case of classical causal diagrams, the causal links between variables are associated with conditional probabilities, in the quantum case, the causal links between subsystem are associated with quantum channels $\mathcal{E}_{X \to Y}$.[30]

Thus, in analogy with the classical causal model framework, the inputs are subsystems of the global Hilbert space and the transformations are quantum channels, which are to be thought of as being obtained by tracing over auxiliary quantum subsystems of the Hilbert space. This is depicted in Fig. 8, where a quantum channel is derived from a unitary transformation by tracing over auxiliary degrees of freedom. Allen et al. (2017) show how we can obtain a picture of DAGs from the starting points of the brief sketch I presented here. They also present a generalization of the causal Markov condition. I refer the reader to Allen et al. (2017) for more details.

A key question now is: How to model interventions in quantum causal models? After all, interventionist ideas are crucial in understanding classical causal models (Woodward 2003). In the context of classical causal models, the

30. Note that Allen et al. (2017) use a particular representation of quantum channels, called the Choi-Jamiolkowski representation, to develop their theory. On this representation, we can associate a density matrix $\rho_{A|B}$ with the channel itself, which makes the analogy with conditional probability functions sharper. But I do not present it this way here to avoid complicating the presentation.

Figure 7: A deterministic structural equation model (SEM) and a classical probabilisitic causal model obtained from it. (a) A very simple deterministic structural equation: $Y = f(\lambda, X)$. (b) A very simple classical channel obtained from this deterministic SEM by averaging over λ.

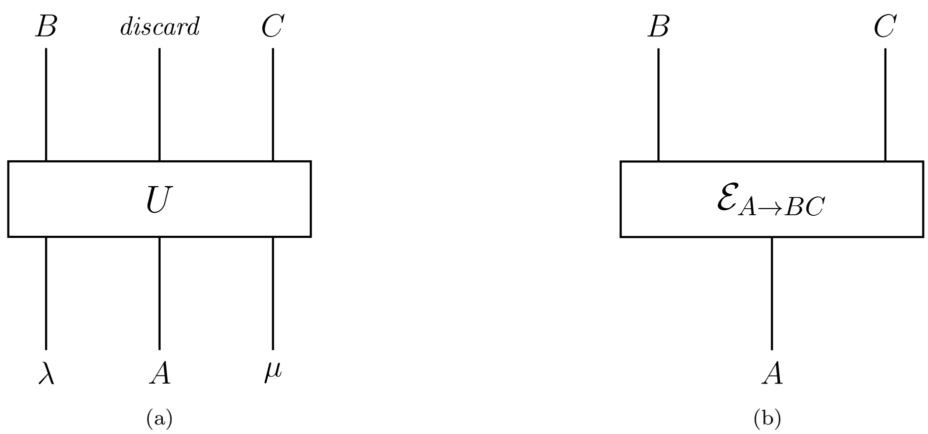

Figure 8: A deterministic quantum map and a quantum causal model obtained from it. (a) A very simple deterministic quantum transformation: $|\psi_{\text{out}}\rangle = U(|\psi_{\lambda A\mu}\rangle$. (b) A very simple quantum channel obtained from the by tracing over an auxiliary system [depicted in (a) by *discard*]: $\rho_{BC} \equiv \text{Tr}_{\text{aux}} |\psi_{\text{out}}\rangle \langle\psi_{\text{out}}| \equiv \mathcal{E}_{A\to BC}(\rho_A)$, where $\rho_A = \text{Tr}_{\lambda\mu}(|\psi_{\lambda A\mu}\rangle \langle\psi_{\lambda A\mu}|)$.

idea of an intervention is modeled, mathematically, using the *do*-calculus. We intervene on a particular variable X by setting the variable equal to particular value: written $do(X = x)$. What this does is that it breaks the arrows between

X and its parents. And then we are interested in the ways in which the act of intervening on X changes its descendants. In the quantum mechanical case, how do we represent the *do* operation? We represent it by a *quantum instrument*: a collection of completely positive maps $\mathcal{E}_k$, where k is to be thought of as a classically readable result of applying the quantum instrument on the system. Thus, if the quantum instrument reads out k, then the quantum state of the system is now $\mathcal{E}_k(\rho)/\mathrm{Tr}(\mathcal{E}_k(\rho))$, assuming that the system was in state ρ before the interaction with the quantum instrument. The denominator—i.e., $\mathrm{Tr}(\mathcal{E}_k(\rho))$—represents the probability that the instrument will record k.

Why are quantum instruments the right framework with which to understand interventions in quantum causal models? What we want an intervention on a system to do, for the purposes of causal explanation, is to fix the system in a particular definite state. In classical physics, what one can do is, in effect, *erase* the previous state of the system, and *rewrite* a new state onto the system. However, this is not possible in general for quantum systems, because the transformations have to be *linear*.[31] The closest thing one can do to deliberately pushing the system on to a desired state is operate on the system with a kind of measuring device, which, with some probability, puts the system into one of many different possible states. This operation is what is mathematically modeled using the framework of quantum instruments I just sketched.[32]

Having done all this set up, the presentation of what is going on in a Bell-violation experiment is really quite straightforward, and can be represented directly in Fig. 9. The two particles of the entangled pair correspond to two subsystems which are carted off to two far away regions, though they maintain their entanglement (this is the part where control of entanglement becomes essential). Then, they are brought into interaction with the measurement devices at the two ends by a unitary process, producing post measurement systems. The unitary interaction with the measurement devices is consistent with the Everettian framework: there is no collapse, and hence no globally unique outcome. These post-measurement systems are then brought together where they interact unitarily with a device that measures the correlations between the two post-measurement systems, producing a quantum system,

31. If a linear map takes all elements of a vector space to a single vector, then that map has to be the zero map. If $Ax = Ay$ for all x, y, then, by linearity, $A(x - y) = 0$ for all x, y. Thus $A = 0$.

32. Note, however, that the framework of quantum instruments is powerful enough to represent the transformation by which one uses a SWAP operation to swa in a quantum system (prepared in an arbitrary quantum state) and swap out the current state. So the "measurements" of generic quantum measurements can be quite complex and allow for delicate local interventions.

which represents as a probabilistic mixture—i.e., a diagonal density matrix—the correlations. According to the Everett interpretation, this diagonal density matrix will correspond to different observers on different branches of the wavefunction seeing different sets of outcomes. But observers seeing Bell-violating statistics will have the overwhelming weight.

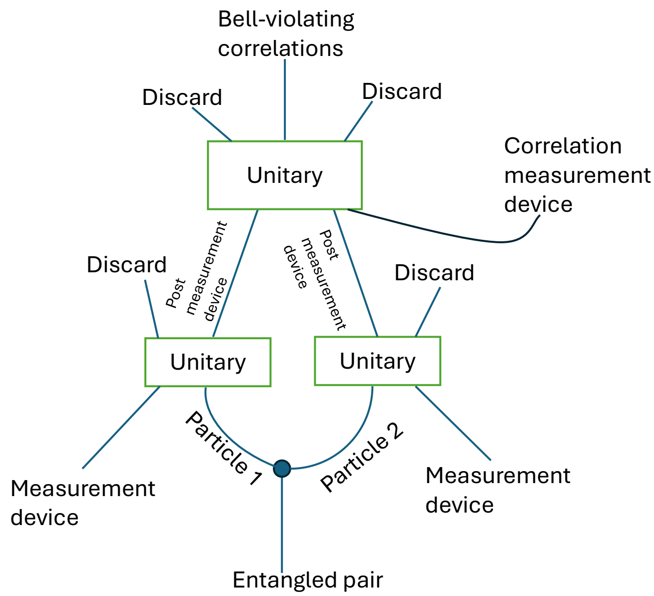

Figure 9: A quantum causal diagram showing how quantum causal diagrams can provide a representation for the Everettian story of what is happening in the experiments showing violations of Bell inequalities.

One point that becomes clear when we use the quantum causal framework to represent such an experiment is that intervening on one side of the experiment has no causal effect on the other side of the experiment. This is because we can only intervene using a quantum instrument which does not allow you select to a state to "collapse" to, which is made particularly clear within the Everettian framework, which has no collapses. Note however, that this condition is *not* transparent either in the quantum formalism itself or in attempts to represent quantum processes in the framework of classical causal models. Within the "textbook" formulation, a measurement on one end instantaneously *collapses* the entangled state, and thus causing an instantaneous definite state to be the correct state on the other end.

Another point becomes clear as well. The quantum causal framework doesn't require any fine-tuning in explaining EPR/Bell-style correlations. This is made precise by formulating the faithfulness condition within this quantum causal framework. Barrett, Lorenz, and Oreshkov (2021) define faithfulness for quantum causal models roughly as follows: an assignment of an initial state and a collection of quantum channels (which results in an assignment of density matrices to subsystems) is faithful to the causal structure (represented by the DAG), insofar as that assignment allows for signaling between any two systems connected by a channel. What does "allow for signaling" mean? Essentially, it means that one can intervene (using a quantum instrument) on one of the systems and effect a change in the probability distribution of potential measurements at the other end. Thus, an assignment of channels/density matrices is unfaithful if, despite a causal pathway between two systems, we cannot signal between them.

From this we can see that there is no worry about faithfulness-violations in the Everettian/quantum causal way of thinking about Bell correlations. Wherever there are causal connections between systems, those connections can be used to signal, but none of these signals are problematic in any way. That is, signaling between the causally connected systems in Fig. 9 doesn't violate the no-signaling theorem or require invoking physics in tension with relativity. In particular, that diagram has no connections between the left wing and right wing until the worldlines of the two experimenters come into contact.

9 Conclusion

So after rehearsing the ways in which classical causal explanations struggle to deal with Bell correlations, I have argued that the combination of Everettian quantum mechanics and a recently developed framework of quantum causal models offers an attractive package using which we can offer realistic causal explanation for Bell correlations that is both local and non-fine-tuned.

To close, let us deal with one final question. Are quantum causal explanations really causal explanations? After all, we took the classical causal modeling framework seriously because we thought it embedded the core assumptions of what good causal explanations consist in. And if these core assumptions no longer hold when it comes to quantum causal models, then to what extent can I say that the quantum causal explanations are actually causal explanations?

I concede that quantum causal models and the assumptions built into them haven't been as well-tested as classical causal models have been. So if these assumptions turn out to be more shaky and less philosophically defensible than one might have initially thought, then it would be questionable whether we have satisfactorily causally explained Bell correlations. I leave it to future work to further unpack and defend the assumptions employed in quantum causal modeling, and to articulate to what extent they can still be reasonably called "causal".[33]

That said, part of what I'm challenging when it comes to the causal modeling framework is not so much its particular assumptions—such as the Markov condition or faithfulness—that embody the principles we expect causal explanations to satisfy, but rather the background *physical* presuppositions that that framework employs. Indeed, the new quantum causal framework is trying to keep the older principles of causal explanations around—i.e., it is trying to retain notions of intervention and trying to formulate appropriate versions of the Markov condition and faithfulness for the quantum context. Instead, what my arguments have done is to show that there was always a *physical* assumption (about the representative aptness of random variables) built into the causal modeling framework and that *physical* assumption is *defeated* in the context of quantum mechanics. This shows us the need to develop a version of our causal modeling that accords with our knowledge of the physical world, while retaining well-tested principles of causal explanations.[34]

Acknowledgments. Special thanks to David Wallace and Jim Woodward for valuable discussions and comments on drafts. Thanks to Nick Huggett and Gal Ben-Porath for useful discussions. Thanks also to those who came to my poster on this topic at the PSA 2022 and Foundations of Physics 2023; in particular to Neil Dewar, John Dougherty, Alex Franklin, and Katie Robertson who raised valuable questions.

References

Albert, David Z, and Rivka Galchen. 2009. "A quantum threat to special relativity." *Scientific American* 300 (3): 32–39.

33. For extant work, see (Lorenz 2022) for a discussion of the motivations behind the framework I've considered here. See the work of Shrapnel (2014, 2019) for a discussion of quantum causal explanations that focuses on a different framework (Costa and Shrapnel 2016).

34. On this broader point, see (Andersen 2017; Papineau 2022; Weinberger, Williams, and Woodward 2024) for approaches that try to recover classical causal modeling from assumptions about the physical world. On such approaches, it shouldn't be surprising if quantum physics requires different kinds of causal models.

Allen, John-Mark A., Jonathan Barrett, Dominic C. Horsman, Ciarán M. Lee, and Robert W. Spekkens. 2017. "Quantum Common Causes and Quantum Causal Models." *Phys. Rev. X* 7 (3): 031021. https://doi.org/10.1103/PhysRevX.7.031021.

Andersen, Holly. 2017. "Patterns, Information, and Causation." *Journal of Philosophy* 114 (11): 592–622. https://doi.org/10.5840/jphil20171141142.

Baas, Augustin, and Baptiste Le Bihan. 2023. "What Does the World Look Like according to Superdeterminism?" *The British Journal for the Philosophy of Science* 74 (3): 555–572. https://doi.org/10.1086/714815.

Barrett, Jonathan, Robin Lorenz, and Ognyan Oreshkov. 2021. "Cyclic quantum causal models." *Nature Communications* 12 (1): 1–15. https://doi.org/10.1038/s41467-020-20456-x.

Bell, J. S. 1964. "On the Einstein Podolsky Rosen paradox." *Physics Physique Fizika* 1 (3): 195–200. https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195.

Bell, John S. 1975. "The theory of local beables." *Epistemological Letters* 9:11–24. Reprinted in: Bell, John S., Abner Shimony, Michael A. Horne, and John F. Clauser. "An exchange on local beables." *Dialectica* (1985): 85-110. https://www.jstor.org/stable/42970534.

Brunner, Nicolas, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. 2014. "Bell nonlocality." *Rev. Mod. Phys.* 86 (2): 419–478. https://doi.org/10.1103/RevModPhys.86.419.

Cavalcanti, Eric G. 2018. "Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning." *Phys. Rev. X* 8 (2): 021018. https://doi.org/10.1103/PhysRevX.8.021018.

Costa, Fabio, and Sally Shrapnel. 2016. "Quantum causal modelling." *New Journal of Physics* 18 (6): 063032.

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

Einstein, A., B. Podolsky, and N. Rosen. 1935. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Phys. Rev.* 47 (10): 777–780. https://doi.org/10.1103/PhysRev.47.777.

Evans, Peter W. 2021. "A Sideways Look at Faithfulness for Quantum Correlations." *Journal of Philosophy* 118 (1): 28–42. https://doi.org/10.5840/jphil202111812.

Friederich, Simon. 2023. "Fine-Tuning." In *The Stanford Encyclopedia of Philosophy,* Winter 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University.

Giustina, Marissa, Alexandra Mech, Sven Ramelow, Bernhard Wittmann, Johannes Kofler, Jörn Beyer, Adriana Lita, Brice Calkins, Thomas Gerrits, Sae Woo Nam, et al. 2013. "Bell violation using entangled photons without the fair-sampling assumption." *Nature* 497 (7448): 227–230. https://doi.org/10.1038/nature12012.

Goldstein, Sheldon, and Ward Struyve. 2007. "On the uniqueness of quantum equilibrium in Bohmian mechanics." *Journal of Statistical Physics* 128:1197–1209. https://doi.org/10.1007/s10955-007-9354-5.

Harrigan, Nicholas, and Robert W Spekkens. 2010. "Einstein, incompleteness, and the epistemic view of quantum states." *Foundations of Physics* 40 (2): 125–157. https://doi.org/10.1007/s10701-009-9347-0.

Hausman, Daniel M. 1999. "Lessons From Quantum Mechanics." *Synthese* 121 (1-2): 79–92. https://doi.org/10.1023/a:1005281714660.

Hausman, Daniel M., and James Woodward. 1999. "Independence, Invariance and the Causal Markov Condition." *The British Journal for the Philosophy of Science* 50 (4): 521–583. https://doi.org/10.1093/bjps/50.4.521.

———. 2004. "Modularity and the Causal Markov Condition: A Restatement." *The British Journal for the Philosophy of Science* 55 (1): 147–161. https://doi.org/10.1093/bjps/55.1.147.

Healey, Richard. 2023. "Quantum-Bayesian and Pragmatist Views of Quantum Theory." In *The Stanford Encyclopedia of Philosophy,* Winter 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University.

Hensen, Bas, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenberg, Raymond FL Vermeulen, Raymond N Schouten, Carlos Abellán, et al. 2015. "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres." *Nature* 526 (7575): 682–686. https://doi.org/10.1038/nature15759.

Hitchcock, Christopher. 2020. "Causal Models." In *The Stanford Encyclopedia of Philosophy,* Summer 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Howard, Don. 1985. "Einstein on locality and separability." *Studies in History and Philosophy of Science Part A* 16 (3): 171–201.

Lorenz, Robin. 2022. "Quantum Causal Models: The Merits of the Spirit of Reichenbach?s Principle for Understanding Quantum Causal Structure." *Synthese* 200 (5): 1–27. https://doi.org/10.1007/s11229-022-03887-5.

Malinsky, Daniel, and David Danks. 2018. "Causal Discovery Algorithms: A Practical Guide." *Philosophy Compass* 13 (1): e12470. https://doi.org/10.1111/phc3.12470.

Maudlin, Tim. 2014. "What Bell did." *Journal of Physics A: Mathematical and Theoretical* 47 (42): 424010. https://doi.org/10.1088/1751-8113/47/42/424010.

———. 2019. *Philosophy of Physics: Quantum Theory.* Princeton University Press.

Myrvold, Wayne, Marco Genovese, and Abner Shimony. 2021. "Bell's Theorem." In *The Stanford Encyclopedia of Philosophy,* Fall 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Näger, Paul. 2016. "The Causal Problem of Entanglement." *Synthese* 193 (4): 1127–1155. https://doi.org/10.1007/s11229-015-0668-6.

Papineau, David. 2022. "The Statistical Nature of Causation." *The Monist* 105, no. 2 (March): 247–275. ISSN: 0026-9662. https://doi.org/10.1093/monist/onab034. eprint: https://academic.oup.com/monist/article-pdf/105/2/247/42897362/onab034.pdf.

Pearl, Judea. 2000. *Causality.* Cambridge University Press.

Rauch, Dominik, Johannes Handsteiner, Armin Hochrainer, Jason Gallicchio, Andrew S. Friedman, Calvin Leung, Bo Liu, et al. 2018. "Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars." *Phys. Rev. Lett.* 121 (8): 080403. https://doi.org/10.1103/PhysRevLett.121.080403.

Reichenbach, Hans. 1956. *The Direction of Time.* Edited by Maria Reichenbach. Mineola, N.Y.: Dover Publications.

Saunders, Simon, Jonathan Barrett, Adrian Kent, and David Wallace. 2010. *Many worlds?: Everett, quantum theory, & reality.* Oxford University Press.

Shalm, Lynden K., Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, et al. 2015. "Strong Loophole-Free Test of Local Realism." *Phys. Rev. Lett.* 115 (25): 250402. https://doi.org/10.1103/PhysRevLett.115.250402.

Shimony, A., M.A. Horne, and J.F. Clauser. 1976. "Comment on "Theory of Local Beables"." *Epistemological Letters* 13:1–8. Reprinted in: Bell, John S., Abner Shimony, Michael A. Horne, and John F. Clauser. "An exchange on local beables." *Dialectica* (1985): 85-110. https://www.jstor.org/stable/42970534.

Shrapnel, Sally. 2014. "Quantum Causal Explanation: Or, Why Birds Fly South." *European Journal for Philosophy of Science* 4 (3): 409–423. https://doi.org/10.1007/s13194-014-0094-5.

———. 2019. "Discovering Quantum Causal Models." *The British Journal for the Philosophy of Science* 70 (1): 1–25. https://doi.org/10.1093/bjps/axx044.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search.* MIT Press: Cambridge.

Suárez, Mauricio, and Iñaki San Pedro. 2010. "Causal Markov, Robustness and the Quantum Correlations." In *Probabilities, Causes and Propensities in Physics,* edited by Mauricio Suárez, 173–193. Springer. https://doi.org/10.1007/978-1-4020-9904-5_8.

Van Fraassen, Bas C. 1982. "The Charybdis of Realism: Epistemological Implications of Bell's Inequality." *Synthese* 52 (1): 25–38. https://doi.org/10.1007/bf00485253.

Wallace, David. 2012. *The emergent multiverse: Quantum theory according to the Everett interpretation.* Oxford University Press.

Weinberger, Naftali. 2018. "Faithfulness, Coordination and Causal Coincidences." *Erkenntnis* 83 (2): 113–133. https://doi.org/10.1007/s10670-017-9882-6.

Weinberger, Naftali, Porter Williams, and James Woodward. 2024. "The Worldly Infrastructure of Causation." *The British Journal for the Philosophy of Science,* https://doi.org/10.1086/730698.

Wood, Christopher J, and Robert W Spekkens. 2015. "The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning." *New Journal of Physics* 17, no. 3 (March): 033002. https://doi.org/10.1088/1367-2630/17/3/033002.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.