Discovering Quantum Causal Models

Sally Shrapnel

Abstract

Costa and Shrapnel [2016] have recently proposed an interventionist theory of quantum causation. The formalism generalises the classical methods of Pearl [2000] and allows for the discovery of quantum causal structure via localised interventions. Classical causal structure is presented as a special case of this more general framework. I introduce the account and consider whether this formalism provides a causal explanation for the Bell correlations.

1 Introduction

It is notoriously difficult to apply causal theory to systems involving quantum phenomena. Entanglement correlations between space-like separated systems seem to defy causal explanation, and it seems near impossible to produce an account that avoids the difficulties posed by non-locality, contextuality, and the measurement problem. In the last thirty years or so, philosophers have advocated a number of fixes, arguing for non-local common causes (Suárez and San Pedro [2011]; Egg and Esfeld [2014]), non-screening off common causes (Butterfield [1992]) and 'uncommon' common causes (Hofer-Szabó *et al.* [2013]; Naeger [2015]). Others have argued for more exotic solutions such as retro-causation (Evans *et al.* [2013]) and super-determinism ('t Hooft [2009]).

All of these approaches share a common theme. Roughly speaking, one starts with quantum correlations (usually the Bell correlations), applies classical causal methodology, identifies a contradiction and then decides what has to go. For interventionist accounts, the dilemma is presented as a choice between relinquishing one of two assumptions: the Causal Markov Condition or faithfulness (no-fine-tuning).[1] The unfortunate consequence of giving up such assumptions, however, is that it becomes unclear why the resulting explanation ought to be thought of as *causal*.

Glymour [2006], Naeger [2015] and Wood and Spekkens [2015] have recently analysed quantum correlations using classical causal modelling methods. Glymour considers the problem to lie with the Causal Markov Condition, whereas Naeger takes aim at the other pillar of causal modelling and advocates that we simply accept unfaithful causal models. Wood and Spekkens suggest that rather than trying to modify one of the existing assumptions, one should reject the classical causal methodology *tout court* and instead focus on developing an explicitly quantum generalisation. We shall look more

[1]Roughly, the Causal Markov Condition states that the causal structure of a set of variables entails a set of conditional independence statements satisfied by the joint distribution over these variables. Faithfulness states that the joint probability distribution satisfies only these conditional independencies.

closely at these suggestions in Section 3.

When faced with the peculiarities of quantum correlations, a well worn alternative is to consider how quantum theory forces one to think of the world, in more general terms. Everett-style interpretations, de-Broglie-Bohm theory and Collapse models all fit within this kind of approach: the aim is not to address certain preconceptions we may have regarding causation, but rather certain preconceptions we may have regarding ontology. For adherents to such interpretations, particularly in light of Russell's eliminativist attack on causation (Russell [1912]), the desire for a specifically causal explanation for quantum correlations may perhaps seem a little unmotivated.[2] However, for those philosophers interested in interventionist conceptions of causation, the former approach certainly remains worth investigating. It is clear that actual experimental and theoretical practice in quantum physics utilises concepts of manipulation and control.[3] Prima facie, it seems an account of quantum interventionist causation ought to be possible. Thus, the question becomes whether one can in fact give a formal account of interventionist causation that correctly accounts for quantum correlations. Ideally, one which recovers classical interventionist causal structure in an appropriate limit. The explanatory power of classical interventionist causal methodology can then be seen as a special case of a more fundamental version, rather than something requiring direct modification.

The recent work of Costa and Shrapnel [2016] is an attempt to provide a consistent formalism that correctly captures quantum causal structure defined along such interventionist lines. Whether this account gives one a causal explanation of the Bell correlations depends, of course, on the manner in which one prefers to define causal relations. Nonetheless, understanding what quantum mechanics has to teach us about

[2]Although, Brown and Timpson [2014] discuss an Everettian interpretation of the Bell experiment that is local, using non-separability to retain the spirit of Reichenbach's common cause principle.

[3]See, for example, (Wiseman and Milburn [2010]).

causation has inspired much hair pulling on the part of both philosophers and physicists in the past: here, at the very least, is a fresh perspective on this old problem.[4]

On a more cautionary note, it is worth saying at the outset that this paper is *not* an attempt to present a particular metaphysical view of fundamental physics. There are many philosophers who believe that physics in general is inhospitable to the existence of causal relations. Problems include the fact that the variables of fundamental physics are not coarse-grained (Woodward [2007]), that physical theories are time symmetric (Farr and Reutlinger [2013], Russell [1912]) and that fundamental physics aims at universality. Roughly speaking, this Russellian view comes in two flavours: (i) those who think there is no room for causation in the fundamental physical ontology of the world, but concede that causal talk is used, and indeed useful, and (ii) those eliminativists who think fundamental physics leaves no room for the existence of causal facts *simpliciter*. I believe the second view can be challenged by a mature theory of interventionism and convincing counterarguments have already been presented. Frisch [2014] has recently produced a book-length defence of the utility of causal talk in the practice of fundamental physics, and Ismael [forthcoming] also takes a broadly anti-eliminativist line on causation and physics. Having said that, those looking for detailed arguments for, or against, the Russellian view will not find them here.

The paper is structured as follows: First, I present the interventionist account of causal explanation and discuss how the formalism (as it currently stands) fails to explain quantum phenomena. Second, I consider why one may wish to produce a causal account of quantum phenomena. Third, I introduce the quantum causal modelling framework of Costa and Shrapnel [2016]. Finally, I consider the implications this has for a causal understanding of the Bell correlations.

---

[4]A roughly similar approach is taken by Fenton-Glynn and Kroedel [2015]: they extend a particular philosophical theory of causation (the Lewisian counterfactual account) to provide a causal account of the Bell correlations. These authors conclude, contra the finding of this paper, that there is a direct causal relationship between space-like separated wings of a Bell experiment.

2 Classical causal models.

In virtue of *what*, exactly, is a causal model a representation of causal structure? For many interventionists this is primarily a pragmatic matter. Causal models enable one to identify effective strategies by distinguishing between probabilistic correlations that are due to causes and those that are merely accidental. Nancy Cartwright [1979] was arguably the first to clearly articulate the importance of this distinction in her paper 'Causal Laws and Effective Strategies'. She was arguing against the eliminativist element of Russell's argument: for her "causal laws cannot be done away with, for they are needed to ground the distinction between effective strategies and ineffective ones" (p420). For example, knowing that the mercury level and the onset of snow are correlated does not tell us whether turning the thermometer upside down will prevent snow. It is causal information that allows us to determine the right action in this situation.

In the last thirty years this basic idea has developed into a sophisticated account of causation known as interventionism. What distinguishes causal from merely correlational relations, is that the former can be modified via localised interventions. Roughly, C is a cause of E when manipulating C in the right way can bring about a change in E (or the probability distribution of E). The representational tool of choice for the computer scientists, statisticians and philosophers who utilise this theory is the causal model, a graphical structure that has found application across a wide range of disciplines.

For the interventionist, causal models are vehicles for learning about the manipulable elements of the world. It is therefore relevant to ask how one ought to think of the relations and relata of these models. Are they to be considered merely as agent-dependent projections (Price [2013])), inherently perspectival (Price [2007]) and ontologically deflationary (Price and Menzies [1993])? Or is it possible to maintain an objective, agent-independent account (Woodward [2003], [2007]) and suggest that causal models may have some (however thin) ontological significance (Woodward [2015])? While disparate positions have recently been brought somewhat closer (Ismael

[2015]), these kinds of metaphysical questions are far from settled.

A number of authors have contributed in important ways to the development of causal modelling, notably Spirtes *et al.* [2000], Woodward [2003] and Pearl [2000]. We shall follow the physicists here and use Pearl's account to focus our discussion. The relata of causal models are typically classical random variables $X_1, .., X_n$. It is assumed that each variable can be associated with a range of 'values': properties that we can unambiguously reveal by measurement or direct observation. Such variables can be binary and used to represent the occurrence or otherwise of an event, can take on a finite range of values or have values that are continuous.[5] It is generally assumed that the properties that these values represent are non-contextual (in the sense of quantum contextuality) and exist prior to, and independently of the act of measurement or observation.[6] [7] Ultimately, these values represent the point of contact between the model and the world.

It is assumed that the joint probability distribution taken over the variables of the system is generated by its 'causal structure', with the structure formed by deterministic causal mechanisms acting between variables, plus some added noise. Such causal structure is represented via a directed acyclic graph (DAG), where the structure of the DAG is assumed isomorphic to the network of autonomous causal mechanisms. When one *passively* observes the system (collects data samples without setting variables to particular values), one is given a window into certain aspects of this structure; when one *intervenes* on the system (some of the data instances correspond to cases where

[5]As is the convention, capital letters ($X_1, X_2...$) will be associated with particular variables, and lower case letters with particular variable values ($x_i, x_j...$).

[6]Very roughly, contextuality means that the value of the observed property depends on the way it is observed, or on which other properties are observed together with it. See (Kochen and Specker [1967]) for an introduction to quantum contextuality and (Spekkens [2005]) for a more modern perspective.

[7]Clearly, this very intuitive starting point is deeply problematic in the quantum case.

particular variables in the network are set to specific values) typically one gains further information about the structure. It is the latter possibility of causal discovery via local interventions that ties the causal modelling framework to Cartwright's original insight. The ultimate goal of such models is to give the user a handle on the manipulable elements of the world: to provide a guide to future action. These ideas are made more precise via the Causal Bayesian Networks (CBN) formalism.

A CBN is an ordered triple $\langle V, G, P \rangle$. V is the set of variables, G a DAG and P a joint probability distribution over the variables V. The graph captures the causal relationships between the variables, with the nodes of the graph being the variables in V and the arrows between them representing causal mechanisms.[8] The CBN is defined by a list of conditional dependencies, one for each variable given its graphical parents ($P(X_i|Pa(X_i))$. These conditional distributions are sometimes known as the 'causal parameters' of the graph and are considered to be generated by the autonomous causal mechanisms acting between variables, plus some unmodelled noise. Parentless variables are associated with marginal distributions.

A CBN is associated with a joint probability distribution taken over all the model variables by applying the product rule to these local conditional distributions:

$$P(X_1, ..X_n) = \prod_i P(X_i|Pa(X_i)) \tag{1}$$

Sometimes equation 1 is called the Causal Markov Condition. However, there is no reason to suppose that such a factorisation would result in any specifically *causal* information, unless one commits to the existence of autonomous causal mechanisms and the possibility of discovery (and verification) of causal structure via local intervention. So, whilst the CMC is often represented mathematically as equation 1, the *causal*

[8]Some basic terminology is useful: a variable A is a 'parent' of B when there is a single arrow from A to B. In such a situation B is a 'child' of A. A is an 'ancestor' of B when there is a 'directed path' of several linked arrows from A to B, in such a case B is a 'descendant' of A.

information of a DAG is perhaps better encoded in the conditional probability:

$$P(x_1, \ldots, x_n | i_1, \ldots, i_N) = \prod_{j=1}^{n} P(x_j | pa_j, i_j), \tag{2}$$

where i_j represents the values of (external) intervention variables used to test the causal structure. Equation 2 reduces to equation 1 when $I_j = \text{idle}$ for all j. That is, for the observational, 'naturally generated' distribution.

Defining the assumptions required to discover causal structure when one *can't* condition on experimental interventions has been a core focus for the last two decades. In many situations, intervening is difficult for pragmatic or ethical reasons but one may nonetheless still wish to answer interventional queries. In such situations, two key assumptions underly the methods used to learn causal structure from observational data: the Causal Markov Condition (CMC) and faithfulness.

Recall, the Causal Markov Condition states that the causal structure of a set of variables entails a set of conditional independence statements satisfied by the joint distribution. When a causal structure is represented graphically via a directed acyclic graph (DAG), and satisfies the CMC, children are conditionally independent of their non-descendants, given their parents. One can think of it as a generalisation of Reichenbach's 'screening off' condition (Reichenbach [1956]).

Faithfulness, on the other hand, states that the joint probability distribution satisfies *only* these conditional independencies. Also known as 'no fine-tuning', faithfulness implies that the only conditional independencies in the distribution are those that hold for *any* set of causal parameters. The idea here is that one does not wish to allow for 'accidental' independencies that are created when causal paths cancel, or when certain symmetries exist for a subset of particular causal parameters.[9] An alternative way to understand the importance of faithfulness is to consider that for *any* distribution it is possible to construct a complete graph (where every node is connected to every other

[9]See (Eberhardt [2013]) for some good examples of unfaithful models, and (Zhang and Spirtes [2015]) for a taxonomy of various varieties and recent progress.

node) that is Markovian to the distribution, by fine-tuning the causal parameters. To avoid such possibilities one asks that the graph is both Markov *and* faithful to the distribution.[10]

Pearl has developed two well known algorithms that take as input a list of conditional independencies (found in a joint distribution over a given set of variables) and return a set of DAG's as output. The IC algorithm will return a DAG, under the assumptions of causal Markovianity, faithfulness and causal sufficiency (no unmeasured common causes). The IC* algorithm does not require the assumption of causal sufficiency, but will in general only return a partial ancestral graph (PAG): a DAG with any number of undirected edges. Including the possibility of latent variables (unmeasured common causes) significantly complicates the task of discovering causal structure from empirical sets containing purely observational data.[11]

Underlying Pearl's picture of causation is the further assumption that joint probability distributions that do *not* permit a Markovian, faithful representation are always a consequence of unmeasured common causes (latent variables). In principle, locating and measuring such variables ought to restore Markovianity to the model for some causal structure. Pearl sees this as a significant strength of the formalism:

> The Markov condition guides us in deciding when a set of parents $Pa(X_i)$ is considered complete, in the sense that it includes ALL the relevant immediate causes of X_i. It permits us to leave some of the causes out of

[10]Some authors ask that the graph be Markov and faithful to the distribution, others that the distribution be Markov and faithful to the graph. The distinction is irrelevant for what I have to say, and merely reflects the different directions in which one can apply the methods: one can start with empirical data and attempt to construct a graph, or start with a graph and check that the statistics match.

[11]In simultaneous and independent work, Peter Spirtes and Clarke Glymour developed equivalent algorithms to IC and IC*: the PC and FCI algorithms. See (Glymour [2016]) for a short and accessible history of the field.

> $Pa(X_i)$ (to be summarised as probabilities), but not if they also influence other variables modelled in the system. If a set $Pa(X_i)$ is too narrow, there will be disturbance terms that influence several variables simultaneously, and the Markov property will be lost. Such disturbances will be represented as latent variables. Once we acknowledge the existence of latent variables and represent their existence explicitly as nodes in a graph, the Markov property will be restored. (Pearl [2009], p. 44)

To retain the connection with autonomous mechanisms and intervention there is the implicit assumption that one could (in principle, if not in practice) always intervene on these latent variables in order to empirically verify the correct causal structure.

Let us take a step back, and look at two general features of this formalism. Firstly, what are the points of contact between such causal models and the world? Roughly speaking there are two: the data that underlies the variable 'values', and (in the case where interventions are possible) the data that we use to characterise the local interventions. It is worth noting here that both kinds of data are not explicitly included in the final model: it is the axioms of probability theory that connect the data instances to the final model.

The second important feature to note is that the twin assumptions of the causal Markov Condition and faithfulness provide a means for associating observational data to causal structure, but don't always result in the discovery of a unique DAG.[12] Crucially, the equivalence set of causal structures generated under these assumptions is deemed *causal* by virtue of the underlying interpretation: autonomous causal mechanisms exist that (i) generate the conditional distributions, and (ii) can be modified through external intervention. On this view, there is only one *correct* causal structure: the one that is, in principle, verifiable via local interventions (Korb [2006]).

Given classical causal modelling methodology, we can now suggest some desiderata we may wish a quantum causal modelling framework to satisfy:

[12]See (Zhang and Spirtes [2015]) for some examples.

1. The formalism should allow for the *discovery* of causal structure from empirical data. At a minimum, such discovery should be possible using interventionist data (data instances where local events are under external control). It would be an advantage if causal structure could also be discovered in situations where interventionist information was incomplete.

2. *All* correlations between empirically derived data should be accounted for via notions of direct, indirect or common cause relations, i.e. there should be no 'unexplained' correlations. In situations where all correlations between empirically derived data can *not* be accounted for via direct, indirect or common cause relations, there should exist a method for extending the model to include possible unobserved, or 'latent', nodes in order to account for the correlations.

3. Classical causal models should be recovered as a limiting case of quantum ones.

3 What's the (quantum) problem?

In general, philosophers take Bell inequality violating correlations to be the focus point for discussions about quantum causation. Although the physical setup of the Bell experiments are relatively straightforward, a great deal of controversy surrounds the appropriate way to interpret the empirical results. These experiments were inspired by Bell's theorem, originally presented in his 1964 paper (Bell [1964]), and written in response to Einstein's 1935 thought experiment (Einstein *et al.* [1935]). Roughly, the EPR argument was presented as a reductio: if one assumes locality (space-like separated systems cannot influence each other), completeness (quantum mechanics is descriptively complete, and reality (as described below), then, coupled with the empirical results of quantum experiments, one arrives at a contradiction. The reality assumption was defined as follows:

> If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then

> there exists an element of physical reality corresponding to this physical quantity. (Einstein *et al.* [1935], p. 77)

For Einstein et. al., quantum mechanics refers to such elements of reality via the so called 'eigenvalue-eigenvector link': an eigenvalue of a quantum system prepared in the relevant eigenstate will result in an outcome that can be predicted with probability one. In the face of the trilemma, EPR advocate we give up completeness, in order to save locality and reality.

Bell's theorem (in its various forms, see Bell [1964], Bell [1971] and Bell [1976]) can be understood as refuting this move. The standard interpretation of Bell's various writings then (in the philosophical literature, at least) is that *any* theory that can account for (i.e. explain) the empirical predictions of quantum mechanics must be non-local. So, roughly, one faces a choice between giving up *locality* or giving up *explanation*.

Bell wrote many papers regarding the philosophical consequences of quantum theory, and much ink has been spilled over the various philosophical assumptions and implications of his work.[13] The recent fifty year celebrations of his famous 1964 paper have inspired another burst of academic activity, and this work suggests that it is not entirely straightforward to gauge exactly what Bell *meant* by his various characterisations of locality and local causality. Indeed, intelligent and careful analysis has resulted in (sometimes alarmingly) disparate positions (see, for example, Werner [2014a], [2014b]; Maudlin [2014b], [2014b]; Wiseman [2014], [2015]; Norsen [2015]). I suspect this reflects, to some extent, that Bell recognised the philosophical difficulties that confound a precise characterisation of both causation and causal explanation, and was honest about his various qualms. Broadly speaking, his work shows that the connections between the various concepts involved (e.g. instantaneous action-at-a-distance, locality, local causality and agency) are far from obvious.

For our purposes, it is relevant to consider some of Bell's comments in light of our

[13]For a comprehensive collection of Bell's papers, see 'Speakable and Unspeakable in Quantum mechanics' (Bell [1990]).

characterisation of interventionist causation in Section 1. Most pertinent are two of his later papers: 'The theory of local beables' and 'La Nouvelle Cuisine' (Bell [1976] and Bell [1990]). I take the view that these papers nicely capture the elements of interventionist thinking that often underly our intuitive notion of causation and causal explanation.[14] Bell's notion of 'local causality' (and, indirectly, causal explanation) is first introduced in 'The theory of local beables.' This paper is also where he generalises his approach to include stochastic theories, which is the setting on which we shall focus here.[15] In part, the aim of this paper was to make explicit some notions that Bell felt were already implicit in ordinary quantum mechanics. The job of his 'beables' was to recast quantum mechanics in terms of the physical properties with which we are already familiar, rather than the more abstract Hermitian operators associated with quantum observables. Bell suggests such beables should satisfy a number of desiderata, but most importantly beables should correspond to something 'physical', in order to distinguish variables that can be associated with 'real physical' values from those that pertain to abstracta. For Bell, the latter ought to be excluded *tout court* from any causal considerations.

As an example:

> The beables must include the setting of switches and knobs on experimental equipment, the currents in coils and the reading of instruments. (Bell [1976], p. 57)

Of particular concern are *local* beables: variables that can be assigned to a particular space-time region. Local causality is then defined with respect to such local beables.

[14]The details I present here are simply those that most closely align with interventionist causation, for a more detailed analysis of these papers see, for example, those papers cited in the previous paragraph and (Brown and Timpson [2014]).

[15]See (Brown and Timpson [2014]) and (Wiseman and Cavalcanti [2015]) for a discussion on the relationship between the deterministic background assumptions of (Einstein *et al.* [1935]); (Bell [1964]) and this later work of Bell's.

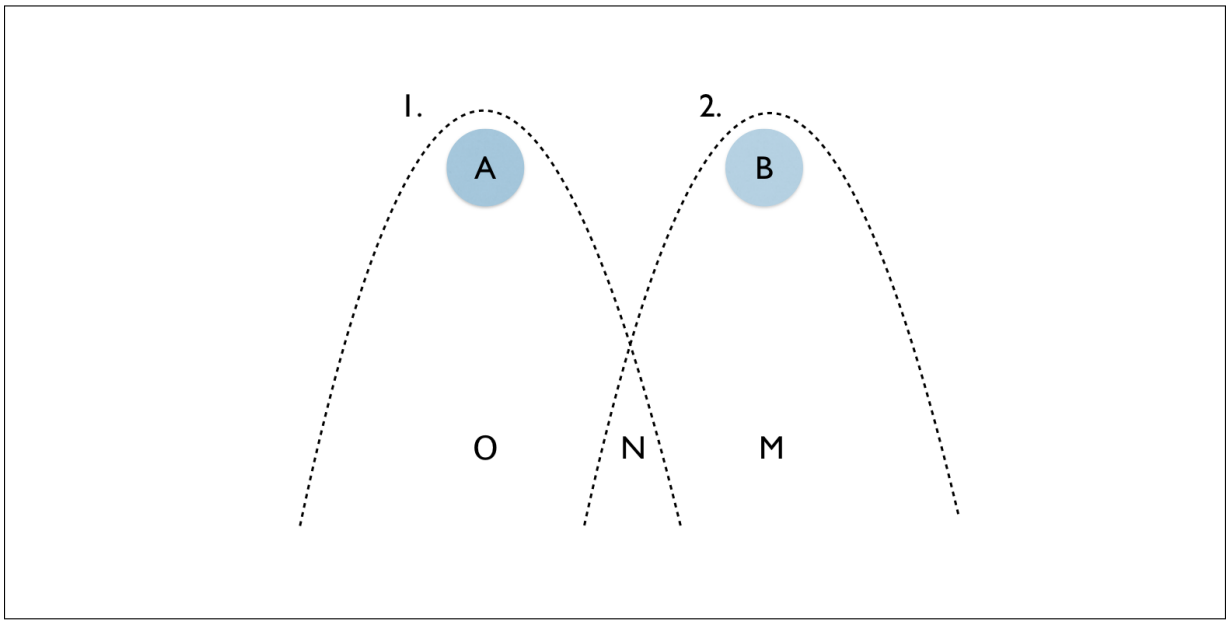

Figure 1: **Bell correlations** A and B are beables associated with space-like separated regions 1 and 2, N denotes a complete specification of all the beables belonging to the overlap of the backward light cone of the two regions, O and M refer to the beables in the remainder of the two light cones for the two regions associated to A and B respectively.

Prima facie, the idea of beables aligns fairly nicely with Pearl's causal relata: classical random variables.

Bell goes on to define local causality in terms of these beables by introducing his famous factorisability condition. Consider two space-like separated beables, A and B associated with space-like separated regions 1 and 2. Let N denote a complete specification of all the beables belonging to the overlap of the backward light cone of 1 and 2 (Fig.1). Let O and M refer to the beables in the remainder of the two light cones for A and B respectively. Then, if one assumes that the joint probability $(A, B|O, M, N)$ factorises into $(A|O, N)(B|M, N)$, one can use expectation values to derive an inequality (a version of the CHSH inequality), which is violated by quantum mechanics.[16] For Bell, this factorisation property "says simply that correlations between A and B can only arise because of common causes N". For many, factorisation here *just is* Bell's notion of local causality.

The overall idea then, is that even by adding in putative hidden beables that may exist in the joint past of beables A and B, one *still* cannot explain their correlation in

[16]An example of this is considered below.

locally causal terms. Thus the 'incompleteness' escape from non-locality suggested by EPR is blocked.

In the final section of 'The theory of local beables', Bell considers that despite the suggestion that nature is causally non-local, we nonetheless cannot *use* such non-locality to signal faster than light. By separating beables into two classes, those apt for human manipulation (e.g. settings) and those that are not manipulable (e.g. the outcomes), he shows that in "this *human* sense relativistic quantum mechanics is locally causal" [p64]. That is, according to quantum theory we are forbidden from manipulating one variable to induce changes in a different, space-like separated variable. Of course, this 'human sense' of causation is pretty close to what interventionists use to characterise causal relations. Although, as we saw in section 2, it is not simply a matter of changing one variable *here*, and seeing if another variable changes *there*, but rather causal relations are thought to be relative to a number of other specific assumptions.

In the closing paragraphs of 'The theory of local beables' Bell reminds us of one further assumption required to derive the inequality: marginal independence of the settings.

> It has been assumed here that the settings of instruments are in some sense free variables - say at the whim of the experimenters – or in any case not determined in the overlap of the backward light cones. Indeed, without such freedom I would not know how to formulate *any* idea of local causality, even the modest human one.

For the interventionist, such 'free variables' correspond to intervention variables. The key point is not that they are somehow uncorrelated with *anything*, but rather that they are not directly correlated with any variables in the model other than their causal target. Woodward's four criteria for an ideal intervention (Woodward [2003]) and Pearl's 'surgical interventions' (Pearl [2000]) neatly capture this requirement.

It should, by now, be fairly obvious that several of Bell's 1976 assumptions fit nicely with the interventionist framework of Section 1. Interestingly, his concerns regarding the *a priori* distinction between 'controllable' and 'uncontrollable' variables and also his

worry that causation may require a notion of agency are still live debates in the interventionist literature *per se* (for example, see (Price [2013]); (Woodward [2015]); (Ismael [2015])). It is intriguing that such concerns, at least in the context of interventionist causation, seem to be pertinent to characterisations of causation, rather than due to the peculiarities of quantum mechanics.

Also relevant for our current purposes, is the final caveat Bell adds regarding his own particular characterisation of causal explanation:

> Of course, the assumptions leading to [the inequality] can be challenged. Equation 22 [factorisation] may not embody *your* idea of local causality. You may feel that only the 'human' version of the last section is sensible and may see some way to make it more precise.

The causal modelling formalism of Costa and Shrapnel [2016] is an attempt to pursue this route. The more precise 'human' version of causation is a generalisation of interventionist causation, that allows for the use of representational devices that go beyond the more familiar classical random variables. Whether this requires one to abandon the notion of 'beables' will be addressed in due course.

We next look at a specific example of the CHSH inequality to bring the association between Bell's characterisation of local causality and modern interventionist causal methods into sharper focus. Two parties are able to perform local measurements on a physical system received from a distant source. The same source is used for the systems received by both parties, and the two systems are emitted simultaneously. The parties randomly choose from one out of two possible measurements A and B (the 'settings').[17] These measurements reveal one out of two possible *outcomes*, labelled X and Y. A latent variable λ represents the physical conditions in the shared past of the two systems. A first pass at a plausible DAG structure might be that of Figure 2. The two wings are set up to be space-like separated, so the omitted edges between (i) the

[17]Some authors consider the three setting case, others consider the two setting CHSH inequality. The differences are irrelevant for what I say here.

outcomes, (ii) the settings, and (iii) each setting and its opposite outcome are (in the first place) assumed due to relativistic constraints.

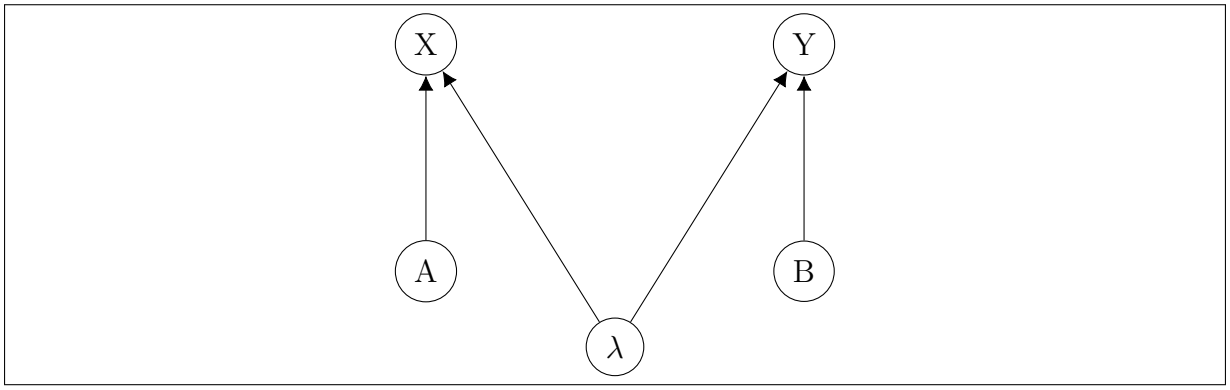

Figure 2: A possible DAG for the Bell experiment. A and B are random variables representing the settings at each wing, and X and Y are the random variables representing the corresponding outcomes. λ captures the physical conditions occurring in the joint past of the two systems (initiated from the source).

For the interventionist, the aim is to produce a causal DAG that can explain the correlations between the variables. There are two possible avenues to pursue. One can postulate a plausible graph structure (as in figure 2) and check if the experimental data does in fact display the statistical patterns implied by the graph. That is, that the distribution is Markov and faithful to the graph. Alternatively, one can start with the empirical data and use the statistical patterns to construct the appropriate graph.

Glymour [2006] was the first to analyse the Bell correlations in terms of the CMC and faithfulness. He argues that constructing a causal graph that satisfies the CMC not only forces one to accept fine-tuning of the causal parameters, but also to accept that the fine-tuned causal influences in question can violate relativity. Broadly speaking, Glymour's position is to advocate a re-think of the Causal Markov condition.

Wood and Spekkens [2015] have also recently analysed the Bell correlations using classical causal modelling methodology. For these authors, a key message is that *any* causal model that reproduces the observed statistics of the Bell correlations must be unfaithful (equivalently, fine-tuned). Using only marginal independence of the settings and no-signalling as input to Pearl's discovery algorithms, Wood and Spekkens [2015] (henceforth WS) show that even if one allows for the existence of latent variables, one must permit fine-tuning of the causal parameters to both explain Bell inequality

violations and still observe the no-signalling conditional independencies. In this case, applying the principle of faithfulness to the observed independencies *implies* the lack of superluminal causal influence that is the hall-mark of Bell's local causality.

A very nice feature of this paper is that the authors are able to show that several popular interpretations of quantum mechanics postulate fine-tuning in order to explain the Bell correlations: de Broglie-Bohm theory, superdeterminism and retrocausal interpretations all require fine-tuning of one sort or another. Additionally, it seems clear that in such cases there are no known local interventions that can break the fine-tuning and discover the 'correct' causal structure. Thus, such interpretations cannot be considered to give a causal interpretation of the Bell experiments in the interventionist sense.[18] Finally, Wood and Spekkens note that their no-go result only holds in the framework of *classical* causal modelling. The take home message, therefore, is that one ought to *reject* the framework of classical causal modelling, in particular the use of random variables for common causes and the use of classical probability theory, in favour of an explicitly *quantum* generalisation.

On the strength of such work, Naeger [2015] advocates we give up faithfulness in order to provide a causal explanation for the Bell correlations. In contrast to Glymour [2006], Naeger suggests that rather than giving up on the CMC in the face of the Bell correlations, the better approach is to accept fine-tuned causal models. Recall, however, in Pearl's formalism, faithfulness only carries weight as a restriction on Markovian models: if we give up Markovianity we lose the value of faithfulness into the bargain. If we give up faithfulness wholesale, we lose the possibility of causal discovery. It seems neither approach affords us a causal explanation worthy of the title.

All three of these modern takes on causal explanation take the position that the data gathered in the Bell experiments ought to be thought of as *observational*: we lack the right kind of interventionist access and so must fall back on the usual CMC and

[18]Of course, if new experimental evidence comes to light that violates no-signalling, then these results will need to be reconsidered. For many, including myself, this seems rather unlikely.

faithfulness constraints in order to tell a causal story about the correlations. The causal modelling framework of Costa and Shrapnel [2016] takes a quite different route. By starting with plausible notions of *interventions* and *causal mechanisms*, that are inspired by the way physicists represent manipulation of quantum systems, they build a theory of quantum interventionist causation from the ground up. Their models satisfy the desiderata of Section 1., and as classical causal structure is recovered in a suitable limit, one is *not* forced into relinquishing classical interventionism, with all its intuitive appeal.

4 Quantum causal models: why bother?

At this point it is worth pausing to consider why we should wish to characterise quantum causal models. Apart from foundational concerns, is there likely to be any pragmatic advantage to the overall project? After all, the philosophical value of causal modelling is almost always defended on pragmatic grounds. And the situation in the special sciences seems, *prima facie*, quite different to the situation in physics. In the special sciences we typically wish to use causal discovery because we have incomplete knowledge of the relevant laws (if indeed such laws exist). For example, we wish to know whether smoking causes cancer, whether inflation causes economic depression, whether increased carbon emission causes climate change. In contrast, for quantum systems it seems we do in fact posses the required knowledge: given initial states, the Hamiltonian for the system and the Schrodinger equation, can we not calculate exactly how things will behave? It would seem that any information gained via causal modelling would, in some sense, be redundant.[19]

There is an important misconception that lies at the heart of this objection. That is, the idealisation of a closed system, for which we can give a perfect Hamiltonian representation and evolve initial states according to global law. It is important to remember that for many real world applications, such an idealised situation is fiendishly

[19]Thanks to an anonymous referee for identifying this issue.

difficult to achieve. We typically assume that there are no initial correlations between the environment and the system of interest, and furthermore that the environment does not act as a 'memory' for the duration of the experiment under consideration. Perfectly isolated experiments where one knows (and can control) *all* the causally relevant degrees of freedom (i.e. one does not have to contend with potentially causally relevant noise) is certainly the exception, rather than the rule. Open quantum systems research recognises this fact, and it is no accident that the causal modelling formalism presented here uses the language of this field. Condensed matter physics, quantum information and quantum technology all utilise systems that comprise many degrees of freedom, with many and varied interactions. Identifying when a system has a Markovian, causal representation is often a difficult task. So in a sense, discovery of causal relations in the special sciences and in physics is perhaps not so different after all. In both situations, causal modelling is useful when we have a multitude of interacting components with only limited access and control.

The second point of note is simply pragmatic. As engineered quantum networks become increasingly complex, the task of characterising the order in which quantum events take place will become increasingly important (and difficult).[20] Current tomographic methods do not provide a method for determining such an order: if one does not know *a priori* whether one has measurements on two halves of a bipartite system, for example, or on a single system at two distinct times, then one does not know whether to apply state or process tomography. By contrast, the Quantum Causal Modelling framework works in the absence of such knowledge.

[20]Classical distributed computing currently uses specific protocols to keep track of the order between local operations (e.g. Lamport time-stamps and vector clock methods). We currently have no analagous methods for characterising the order of quantum operations.

5 Markov Quantum Causal Models

The notion of quantum intervention used to define the Quantum Causal Models (QCMs) of Costa and Shrapnel [2016] matches the way certain physicists represent manipulation and control of quantum systems. Thus the central claim of interventionist causation is retained: the quantum causal models help identify the manipulable elements of the world. From here, they build up a theory of quantum causal structure in much the same way as Pearl builds a notion of classical causal structure. The mathematical structures of this theory are not the joint and conditional probability distributions of Pearl's formalism, but rather those utilised by quantum open systems research.

The QCMs are defined by sets of possible spatio-temporally local quantum operations (graphically represented as nodes) and sets of quantum *channels* that represent the causal influences acting between the nodes (graphically represented as edges). Causal structure can be *discovered* using a general notion of intervention which is defined as a choice of a quantum 'instrument' (a particular set of quantum operations). Using these objects one can define direct, indirect and common cause relations. One can also show that by defining a quantum Markov condition, it is possible to identify when a causal graph is incomplete, in the sense that there is an unmodelled (latent) common cause. Extending the model to include such nodes restores the Markov property to the causal graph, and an analogue definition of faithfulness is also possible. Finally, it is also possible to recover classical causal structures, as a limiting case where all local operations are diagonal in a fixed basis.

These definitions are all fairly technical, and to the uninitiated rather daunting. Rather than simply reproduce the results of the paper here, I shall try to focus on those differences between the classical and quantum causal modelling formalisms that are likely to be of most interest to philosophers (but see Appendix A for technical definitions and Shrapnel [2016] for a more detailed philosophical presentation).

It is most instructive to compare each of the components of the QCM formalism to its classical counterpart. For classical models, each variable (graphically, node) represents a possibility space, defined by a set of possible values or events. For the

quantum models, the nodes are also associated to a possibility space, although in this case defined by a set of completely positive maps (CP maps).[21] Each map can be given a classical label and for a single run of an experiment a single map is assumed to represent a quantum event occurring in the spatio-temporal region associated to the relevant node. These maps provide a possibility space for the many ways a quantum state may change as a result of a local intervention.

In the classical case, interventions can be 'ideal', where a single value is chosen with probability one, or 'generalised', where the intervention rather induces the value to respond in a probabilistic manner. Such 'probabilistic choosing' more accurately represents intervening in most experimental situations, where perfect control is merely an idealisation.[22] In the QCM framework, interventions correspond to this more general kind: an intervention chooses only probabilistically among a set of CP maps (in quantum information language, an intervention corresponds to a 'quantum instrument'). Such quantum interventions are associated to completely positive trace preserving maps (CPTP maps), formed by summing the individual CP maps. The trace-preserving property is due to the fact that we assume that with certainty at least one of the events represented by a particular CP map will occur.

The story already seems to involve quite a departure from the classical causal modelling paradigm. In the QCM formalism we assume that all quantum events are associated with a change or process (CP maps are used to characterise evolution of a quantum state). It is not obvious how one ought to relate this to the classical idea of values or events determined by the kind of 'passive measurement' assumed in Pearl's formalism. It seems plausible however, to consider that even classical measurements can be considered as resulting in a process, or change in the state of the system being measured. Advances in the philosophy of measurement theory suggest our epistemic access to measurement outcomes is a rather convoluted and model-based affair. We

[21]Note, individual CP maps need not be trace-preserving.

[22]Korb [2006] calls such interventions 'imperfect', Pearl [2000] calls them 'generalised'.

ought not to think of measurements as simply providing an outcome that is directly isomorphic to an underlying physical state. As Tal [2013] reminds us, "measurement consists of two levels: (i) a concrete process involving interactions between an object of interest, an instrument, and the environment; and (ii) a theoretical and/or statistical model of that process, where "model" denotes an abstract and local representation constructed from simplifying assumptions." Associating events with possible processes rather than static values is perhaps not as counterintuitive as at first it may seem.[23]

Recall, for classical causal models, the edges of the causal graph correspond to the *causal mechanisms* responsible for determining the statistical correlations that exist between events at different nodes. In the quantum causal models, we assume the functional relationships between the nodes are determined by quantum systems passing between different spatio-temporal regions (nodes), possibly interacting with an unknown environment. In accordance with classical models, we call such functional relationships *quantum causal mechanisms* and depict them graphically via edges. Such connecting mechanisms are also represented via CP maps that sum to CPTP maps (although in this case there is a slightly different representation, amounting to a partial transpose).

In the classical case, autonomous causal mechanisms are considered to be responsible for the probabilistic correlations of a causal network, and it is the autonomy of the mechanisms that allows for the possibility of so-called 'surgical' interventions. This leads us to consider the relationship between interventions and mechanisms in the

[23]This is somewhat related to the question of whether there is either a 'realist' or 'operationalist' stance in lurking in the background of the QCM formalism. In terms of standard conceptions of the terms, the formalism most naturally fits within an operationalist paradigm. The quantum interventions are most easily understood as being associated with choices we make in the lab, and we do not make use of the eigenvector-eignavalue link. Having said that, one is also free to understand the CP maps as attaching to single quantum events, and so understood from the perspective of a process ontology, it may well be possible to consider the QCMs as affording a kind of realist interpretation.

quantum case. A deterministic quantum mechanism in this formalism would correspond to a unitary map, relating the output space of one node to the input space of another. Recall in the classical case any unmodelled noise is assumed to be uncorrelated, thus ensuring the autonomy of the mechanisms and underpinning the relevance of the Causal Markov condition. In the quantum case, external noise leads to the use of the more general CPTP map, rather than a unitary map to represent the mechanism.[24]

Using these basic ingredients, Costa and Shrapnel [2016] are able to build an interventionist theory very close in spirit to the classical interventionist theory of Pearl. Discovery of causal structure from empirical data is made possible via a quantum Markov condition and a requirement of faithfulness that is analogous to the classical version. Importantly, it is also shown that classical causal structure (with the usual screening off properties) can be recovered by assuming local operations to be diagonal in a fixed basis. There is, however, nothing in the paper to suggest what drives such a limiting case:

> Whether this condition is enforced by decoherence (Zurek [2014]), collapse models (Ghirardi [1985]), 'fuzzy measurements' (Peres [1992]), or in other ways, will not be discussed here. Rather, it will be shown that, provided the transition to classicality takes place, Markov quantum causal models reduce to classical ones.

6 The Bell experiment

The formalism of Costa and Shrapnel [2016] is designed to define causal relations between multiple quantum systems of arbitrary dimensions. As such, when applied to the Bell correlations, the results seem almost trivial. The experimental statistics will decompose according to the Quantum Causal Markov Condition to satisfy a simple

[24]One can also use CPTP maps to describe irreducible noise such as that arising in dynamical collapse models. We do not need to commit to that particular interpretation here.

common cause structure. There is no causal relation between the wings of the experiment, but rather the source acts as a common cause. Whilst it is tempting to argue then, that Costa and Shrapnel have developed somewhat of a sledgehammer to crack a nut, without a formalism that generalises to multiple systems of arbitrary dimensions, one would be hard-pressed to claim that it is a complete causal theory that can provide adequate causal explanations.

7 Bell's objections

Now we have a basic understanding of the formal structure of the quantum causal models, we can review some philosophical implications. For many, the key question is whether the models do in fact provide a causal explanation for the Bell correlations. Recall the historical trajectory:

1. EPR suggest that quantum correlations force a choice between non-locality (causal influence between space-like separated events) and completeness,

2. Bell shows that if causal influence is defined along classical interventionist lines, adding further hidden variables does not circumvent the need for non-locality: that is, one cannot save locality by assuming incompleteness,

3. Wood and Spekkens show that even if one allows for non-local influences, the resulting causal explanation is flawed (due to the presence of fine-tuning). One cannot save causal explanation by allowing for non-local causal influences.

The quantum causal modelling framework presented in (Costa and Shrapnel [2016]) assumes it is the characterisation of causation and causal explanation that is at fault, and seeks to provide an alternative. The alternative presented is complete, in the sense that all empirically derived statistics can be explained by common cause, direct cause or indirect cause relations. If such an explanation is not possible, then this signals the existence of hidden causes and an extended model that does correctly characterise the empirical statistics can be formed. The formalism recovers the classical modelling formalism as a limiting case, thus we do not need to entirely give up on the

interventionist account, with all its intuitive appeal. Rather, it is seen as an approximation of something deeper.

Regarding locality, direct causal influence in these models is always consistent with relativistic constraints by virtue of its consistency with quantum mechanics. Thus, empirically derived statistics can always be explained using such models without postulating non-local causal influence (direct causal mechanisms acting between space-like separated regions).

From the perspective of the Bell literature, there are some obvious objections. Recall, Bell was uncomfortable with including a 'human' element within an account of causation: correctly capturing possible signalling relations did not seem, for Bell, to be an apt characterisation of causation. We saw in Section 1 that this charge has also been levelled at interventionist accounts of causation: whether interventionist causation reduces to agency is a hotly debated question. The interesting point however, is that this problem does not seem to be peculiar to the *quantum* characterisation. It seems to me that one could argue against the need for agency along exactly the same lines as the standard interventionist response (Woodward [2003]). That is, we observe certain regularities in experimental situations (where we have an element of control) that allow us to infer particular causal relations between naturally occurring events (which we do not control).

The question of how the QCM's relate to Bell's notion of beables is less clear. Certainly, the models are connected to empirical evidence gathered from "the setting of switches and knobs on experimental equipment, the currents in coils and the reading of instruments." However, if Bell's aim was to use beables in order to expunge the use of *any* mathematical representational devices that move beyond classical random variables, then clearly the QCM's fail in this respect.

As I see it, if one believes that interventionism is the correct way to think about causation, then the empirical results of quantum experiments force one of three choices: (i) dub the quantum world as mysteriously acausal, (ii) *abandon* the interventionist account as the preferred account of causation and look for an alternative causal theory

that *can* explain quantum correlations, or (iii) *generalise* the interventionist account so that it can account for *both* classical and quantum causal relations. The QCM's of Costa and Shrapnel [2016] represent an attempt at the third path.

Funding

This work was supported by an Australian Research Council Centre of Excellence for Quantum Engineered Systems grant (CE 110001013), and by the Templeton World Charity Foundation (TWCF 0064/AB38).

Acknowledgements

I am indebted to Fabio Costa, Gerard Milburn, Christopher Timpson and three anonymous referees for their help and advice.

Sally Shrapnel

† *School of Historical and Philosophical Inquiry;*

∗ *Centre for Engineered Quantum Systems, School of Mathematics and Physics*

University of Queensland,

St Lucia, Qld, 4072

s.shrapnel@uq.edu.au

References

Bell, J. S. [1964]: 'On the Einstein-Podolsky-Rosen paradox', *Physics*, **1**(3), pp. 195–200.

Bell, J. S. [1971]: 'Introduction to the hidden variable question', in *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, pp. 40–47.

Bell, J. S. [1976]: 'The theory of local beables', in *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, pp. 57–65.

Bell, J. S. [1990]: 'La Nouvelle Cuisine', in *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, pp. 232–248.

Brown, H. R. and Timpson, C. G. [2014]: 'Bell on Bell's theorem: The changing face of nonlocality', To appear in '50 Years of Bell's Theorem', Mary Bell and Shan Gao (eds.), CUP.

Butterfield, J. [1992]: 'David Lewis Meets John Bell', *Philosophy of Science*, **59(1)**.

Cartwright, N. [1979]: 'Causal Laws and Effective Strategies.', *Nous*, **13 (4)**.

Costa, F. and Shrapnel, S. [2016]: 'Quantum causal modelling', *New Journal of Physics*, **18**.

Eberhardt, F. [2013]: 'Direct Causes and the Trouble with Soft Interventions', *Erkenntis*, **79(4)**, pp. 755–777.

Egg, M. and Esfeld, M. [2014]: 'Non-local Common Cause Explanations for EPR', *European Journal for Philosophy of Science*, **4**.

Einstein, A., Podolsky, B. and Rosen, N. [1935]: 'Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?', *Phys. Rev.*, **47**, pp. 777–780.

Evans, P. W., Price, H. and Wharton, K. B. [2013]: 'New Slant on the EPR-Bell Experiment', *British Journal for the Philosophy of Science*, **64**, pp. 297–324.

Farr, M. and Reutlinger, A. [2013]: 'A Relic of a Bygone Age? Causation, Time Symmetry and the Directionality Argument.', *Erkenntnis*, **78(2)**, pp. 215–235.

Fenton-Glynn, L. and Kroedel, T. 2015. Relativity, Quantum Entanglement, Counterfactuals, and Causation *The Birthish Journal for the Philosophy of Science* 66(1) 45-67.

Frisch, M. [2014]: *Causal Reasoning in Physics*, Cambridge University Press.

Glymour, C. [2006]: 'Markov properties and quantum experiments', in *Physical Theory and its Interpretation*, Springer, pp. 117–126.

Glymour, Clark. 2016. Clark Glymour's responses to the contributions to the Synthese special issue *Causation, probability, and truth: the philosophy of Clark Glymour Synthese* 193 (4):1251-128.

Hofer-Szabó, G., Rédei, M. and Szabó, L. [2013]: *The Principle of the Common Cause*, Cambridge University Press.

Ismael, J. [2015]: 'How do causes depend on us? The many faces of perspectivalism', *Synthese*, **191(1)**, pp. 245–267.

Ismael, J. [forthcoming]: 'Against Globalism About Laws', in B. van Fraassen and I. Peschard (*eds*), *Experimentation and the Philosophy of Science*, Chicago: University of Chicago Press.

Kochen, S. and Specker, E. [1967]: 'The Problem of Hidden Variables in Quantum Mechanics', *Journal of Mathematics and Mechanics.*, **17**.

Kevin B. Korb and Erik Nyberg. [2006]. 'The Power of Intervention' *Minds and Machines*, 16(3): 289–302.

Maudlin, Tim. 2014b. Reply to Werner *arXiv:408.1828.*

Maudlin, T. [2014b]: 'What Bell did', *Journal of Physics A: Mathematical and Theoretical*, **47**(42), pp. 424010.

Näger, P. M. 2015. The causal problem of entanglement *Synthese*, 1-29

Norsen, Travis. 2015. 'Are there really two different Bell's theorems?', *arXiv preprint arXiv:1503.05017.*

Oreshkov, O., Costa, F. and Brukner, C. [2012]: 'Quantum correlations with no causal order.', *Nature Communications*, **3**, pp. 1092.

Pearl, J. [2000]: *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 1st edition.

Pearl, J. [2009]: *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 2nd edition.

A. Peres. 1992. 'Emergence of local realism in fuzzy observations of correlated quantum systems' *Found. Phys.* 22: 819–828.

G. Ghirardi, A. Rimini, and T. Weber, 'A model for a unified quantum description of macroscopic and microscopic systems' In *Quantum Probability and Applications II*, L. Accardi and W. von Waldenfels, eds., vol. 1136 of *Lecture Notes in Mathematics*, Springer Berlin Heidelberg, 223–232.

Price, H. [2007]: 'Causal Perspectivalism', in H. Price and R. Corry (*eds*), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford University Press, pp. 250–292.

Price, H. [2013]: 'Causation, Intervention and Agency - Woodward on Menzies and Price', in H. C. Beebee, Helen and H. Price (*eds*), *Making a Difference.*, Oxford University Press.

Price, H. and Menzies, P. [1993]: 'Causation as a secondary quality', *Journal for the Philosophy of Science*, **44**, pp. 187 – 203.

Reichenbach, H. [1956]: *The Direction of Time*, University of Los Angeles Press.

Russell, B. [1912]: 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, **13**, pp. 1–26.

Shrapnel, S. [2016]: *Using interventions to discover quantum causal structure*, Ph.D. thesis, University of Queensland. *https://espace.library.uq.edu.au/view/UQ:411093*

Shrapnel, Sally and Costa, Fabio. 2017. 'Updating the Born rule'. *arXiv:1702.01845*

Spekkens, R. W. [2005]: 'Contextuality for preparations, transformations, and unsharp measurements', *Phys. Rev. A*, **71**(5), pp 052108-25,

Spirtes, P., Glymour, C. and Scheines, R. [2000]: *Causation, Prediction, and Search*, Cambridge, Mass.: MIT Press.

Suárez, M. and San Pedro, I. [2011]: 'Causal Markov, Robustness and the Quantum Correlations', in M. Suárez (*ed.*), *Probabilities, Causes and Propensities in Physics*, Dordrecht: Springer, pp. 173–193.

't Hooft, Gerard. [2009]. 'Entangled quantum states in a local deterministic theory.' arXiv: 0908.3408

Tal, E. [2013]: 'Old and New Problems in Philosophy of Measurement', *Philosophy Compass*, **8**(12), pp. 1159–1173.

Werner, R. F. [2014a]: 'Comment on 'What Bell did'', *Journal of Physics A: Mathematical and Theoretical*, **47**(42), pp. 424011.

Werner, R. F. [2014b]: 'What Maudlin replied to', *arXiv:1411.2120*.

Wiseman, H. and Milburn, G. [2010]. *Quantum Measurment and Control*, Cambridge University press.

Wiseman, H. M. [2014]: 'The Two Bell's Theorems of John Bell', *Math. Theor.*, **47**, pp. 424001.

Wiseman, H. M. and Cavalcanti, E. G. [2015]: 'Causarum Investigatio and the Two Bell's Theorems of John Bell', *arXiv:1503.06413*.

Wiseman, H. M. and Rieffel, E. G. [2015]: 'Reply to Norsen's paper 'Are there really two different Bell's theorems?"', *arXiv:1503.06978*.

Wood, C. J. and Spekkens, R. W. [2015]: 'The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning', *New Journal of Physics*, **17**(3), pp. 033002.

Woodward, J. [2003]: *Making Things Happen*, 'Oxford: Oxford University Press.

Woodward, J. [2007]: 'Causation with a Human Face', in H. Price and R. Corry (*eds*), *Causation, Physics, and the Constitution of Reality*, Oxford: Oxford University Press, pp. 66–105.

Woodward, J. [2015]: 'Methodology, Ontology, and Interventionism', *Synthese*, **192(11)**, pp. 3577–3599.

Zhang, J. and Spirtes, P. [2015]: 'The Three Faces of Faithfulness', *Synthese.*, **1**, pp. 17.

Zurek, W. [2014]: 'Quantum Darwinism, classical reality, and the randomness of quantum jumps', *Physics Today*, **67**.

8 Appendix

The following definitions are taken from (Costa and Shrapnel [2016]). For a more comprehensive list, with illustrative examples, see both (Costa and Shrapnel [2016]) and (Shrapnel [2016]).

Definition 1. Quantum event

A **quantum event** *is represented by a completely positive (CP) map $\mathcal{M} : A_I \to A_O$, where input and output spaces are the spaces of linear operators over input and output Hilbert spaces, $A_I \equiv (\mathcal{H}^{A_I})$, $A_O \equiv (\mathcal{H}^{A_O})$, respectively. Note, a CP map can be represented as a matrix by using the Choi-Jamiołkowski isomorphism:*

$$M^{A_I A_O} = \sum_{jl} |l\rangle\langle j|^{A_I} \otimes \left[\mathcal{M}(|j\rangle\langle l|)^{A_O}\right]^T, \tag{3}$$

$$\mathcal{M}(\rho)^{A_O} = \left[\mathrm{tr}_{A_I}\left(\rho^{A_I} \otimes \mathbb{1}^{A_O} \cdot M^{A_I A_O}\right)\right]^T, \tag{4}$$

where $\{|j\rangle\}_{j=1}^{d_{A_I}}$ is an orthonormal basis in $\mathcal{H}^{A_I}$ and ^{T} denotes transposition in that basis.

Definition 2. Local region

The space of potential events is called a **local region** *and is identified with the set of CP maps between an input (A_I) and an output (A_O) space, which is isomorphic to the space $A_I \otimes A_O$. Input and output spaces can be identified with the past and the future space-like boundaries of the space-time region where the event takes place.*

Definition 3. *A* **quantum causal mechanism** *maps the output space of a local region to the input space of another one. The analogue of a* deterministic *mechanism is a unitary. External noise can be described by an interaction with an environment which is then traced out: the most general definition of a mechanism is a CPTP (completely positive trace preserving) map.*

Definition 4. *A* **quantum intervention** *represents the collection of all possible events that can be observed given a specific choice of probing the system. An intervention is*

defined as a choice of quantum instrument. *For a local region $A_I \otimes A_O$, an instrument is a set of CP maps that sum up to a CPTP map:*

$$\begin{aligned} \mathcal{J} &= \left\{ M_x^{A_I A_O} \right\}_{x=1}^{m}, \quad M_x^{A_I A_O} \geq 0, \\ \mathrm{tr}_{A_O} \sum_{a=1}^{m} M_x^{A_I A_O} &= \mathbb{1}^{A_I}. \end{aligned} \tag{5}$$

(The trace-preserving condition for the sum guarantees that probabilities sum up to 1*.)*

Definition 5. Quantum Process Rule

To describe an experiment consider a set of local regions $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^{n}$, representing n disjoint space-time regions, each bounded by two space-like surfaces. In an individual run of an experiment, instruments $\mathcal{J}_1^{L_1}, \ldots, \mathcal{J}_n^{L_n}$ are implemented in these local regions and the corresponding events recorded. The events are described by CP maps $M_1^{L_1}, \ldots, M_n^{L_n}$. It is possible to prove that the probability for such events to occur is given by the Quantum Process Rule (Oreshkov et al. *2012, Shrapnel and Costa 2017):*

$$P(M_1^{L_1}, \ldots, M_n^{L_n} | \mathcal{J}_1^{L_1}, \ldots, \mathcal{J}_n^{L_n}) = \mathrm{tr}\left[M_1^{L_1} \otimes \cdots \otimes M_n^{L_n} \cdot W^{L_1 \ldots L_n} \right], \tag{6}$$

where $W^{L_1 \ldots L_n} \geq 0$ is called the process matrix *and represents the information about the outside world available in the local regions.*[25]

Definition 6. *Given a set of local regions $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^{n}$ and a process matrix $W^{L_1 \ldots L_n}$, a region L_h represents a* ***direct cause*** *for a region $L_k \neq L_h$ if, for* any possible *set of instruments $\left\{ \mathcal{J}_j^{L_j} \right\}_{j \neq k}$, there exist instruments $\mathcal{J}_k^{L_k}$ and $\widetilde{\mathcal{J}_h}^{L_h}$ such that*

$$\begin{aligned} &P(M_k^{L_k} | \mathcal{J}_1^{L_1}, \ldots, \mathcal{J}_k^{L_k}, \ldots, \mathcal{J}_h^{L_h}, \ldots, \mathcal{J}_n^{L_n}) \\ &\qquad \neq P(M_k^{L_k} | \mathcal{J}_1^{L_1}, \ldots, \mathcal{J}_k^{L_k}, \ldots, \widetilde{\mathcal{J}_h}^{L_h}, \ldots, \mathcal{J}_n^{L_n}). \end{aligned} \tag{7}$$

Definition 7 (MQCM)**.** *Given a set of local regions $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^{n}$, a* **Markov quantum causal model (MQCM)** *is a pair $\langle G, W \rangle$, where*

[25] This rule is also known as the generalised Born rule.

1. $G = \langle \mathcal{L}, \mathcal{E} \rangle$ *is a DAG that has the local regions as vertices;*

2. *to each edge* $e \in \mathcal{E}$ *is associated a space* S_e *such that* $O_j = \bigotimes_{e \in \mathrm{OUT}_j} S_e$, $j = 1, \ldots, n$, *where*

$$\mathrm{OUT}_j := \{e \in \mathcal{E} | e = (L_j, L_k)\} \tag{8}$$

is the set of edges departing from the vertex L_j*;*

3. W *is a process matrix of the form*

$$W^{L_1 \ldots L_n} = \bigotimes_{j=1}^{n} T_j^{PS_j I_j} \otimes \mathbb{1}^{O_D}, \tag{9}$$

where $O_D := \bigotimes_{k \in \mathcal{D}} O_k$ *is the output space of the regions with no outgoing edges,* $\mathcal{D} := \{k | \mathrm{OUT}_k = \varnothing\}$*;* $PS_j := \bigotimes_{e \in \mathrm{IN}_j} S_e$ *is the* parent space *associated with region* L_j*, with*

$$\mathrm{IN}_j := \{e \in \mathcal{E} | e = (L_k, L_j)\} \tag{10}$$

the set of incoming edges to L_j*; and*

$$T_j^{PS_j I_j} \geq 0, \quad \mathrm{tr}_{I_j} T_j^{PS_j I_j} = \mathbb{1}^{PS_j}, \quad j = 1, \ldots, n. \tag{11}$$

Definition 8. Latent regions and extended models

A local region in which the events are not observed will be called **latent**. *A* **Markovian causal explanation** *for a process matrix* $W^{L_1 \ldots L_n}$ *is an MQCM* $\left\langle G, \widetilde{W} \right\rangle$*, where* G *is a DAG with vertices containing* $L_1, \ldots, L_n$*, and possibly latent regions* $\widetilde{\mathcal{L}} = \left\{ \tilde{L}_1, \ldots, \tilde{L}_m \right\}$*, such that*

$$W = \mathrm{tr}_{\widetilde{\mathcal{L}}} \left[C_1^{\tilde{L}_1} \otimes \cdots \otimes C_m^{\tilde{L}_m} \cdot \widetilde{W}^{L_1 \ldots L_n \tilde{L}_1 \ldots \tilde{L}_m} \right] \tag{12}$$

for some CPTP maps $C_1^{\tilde{L}_1}, \ldots, C_m^{\tilde{L}_m}$*, where* $\mathrm{tr}_{\widetilde{\mathcal{L}}}$ *denotes the partial trace over all the regions in* $\widetilde{\mathcal{L}}$*.* W *is called* reduced process matrix *and provides a full description of the physical situation for the observed regions, once the instruments in the latent regions are*

fixed; $\widetilde{W}$ *is an* extension *of* W.