Operational equivalence and causal structure

Gábor Hofer-Szabó

*Institute of Philosophy, HUN-REN Research Center for the Humanities, Tóth Kálmán u. 4, Budapest, 1097, Hungary.

Corresponding author(s). E-mail(s): szabo.gabor@abtk.hu;

Abstract

In operational quantum mechanics two measurements are called operationally equivalent if they yield the same distribution of outcomes in every quantum state and hence are represented by the same operator. In this paper, I will show that the ontological models for quantum mechanics and, more generally, for any operational theory sensitively depend on which measurement we choose from the class of operationally equivalent measurements, or more precisely, which of the chosen measurements can be performed simultaneously. To this goal, I will take first three examples—a classical theory, the EPR-Bell scenario and the Popescu-Rochlich box; then realize each example by two operationally equivalent but different operational theories—one with a trivial and another with a non-trivial compatibility structure; and finally show that the ontological models for the different theories will be different with respect to their causal structure, contextuality, and fine-tuning.

Keywords: operational theory, operational equivalence, contextuality, causal models, fine-tuning

1 Introduction: a Bridgmanian perspective

On strict operationalism, concepts should be defined by empirical operations. In this tradition, going back to Percy Bridgman (1927) and the Vienna Circle (Schlick, 1930), two concepts which are defined by different operational procedures cannot be the same. Using Bridgman's example, length measured by a ruler and length measured by light signals are different concepts, and true science should use different names to discern them. As times passed, philosophy of science (and also Bridgman himself) has gradually moved away from strict operationalism and revealed various semantic, pragmatic and common sense criteria for identifying concepts with different operational basis

(Chang, 2019). In the case of physical magnitudes or observables, the standard way was to check whether the two measurements defining the two observables have the same outcome in their common domain. If the length of medium sized objects agree when measured by a ruler or measured by light signals, then—at least in this common domain—one is justified in using one length concept instead of two.

For this comparison, however, at least one of the two conditions needs to hold for each system:

(i) either both measurements should be able to be performed *simultaneously* on the system; or

(ii) we need to have a precise enough preparation procedure such that every preparation is an *eigenstate* for both measurements, that is a preparation in which both measurements have definite result.

This latter happens in classical physics where each pure (dispersion-free) state is an eigenstate for any measurements. Thus in classical physics, we can easily decide on whether two measurements measure the same observable: just prepare the system in an eigenstate, perform the one measurement, prepare the system again in the same eigenstate, perform the other measurement and compare the outcomes. However, if the preparation procedure is not as fine-grained as to yield definite results for all measurements, as is the case in quantum mechanics, we are left with option (i) to identify observables of different measurements: we need to measure them simultaneously and check whether the outcomes match in every single run.

But what if the two measurement procedures cannot be performed at the same time? From a strict operationalist position, we are not entitled to identify the two observables in this case. Still, in quantum mechanics this is what happens. Here, two measurements which yield the same outcome statistics in every quantum state are represented by the same operator and said to measure the same observable. Such measurements are called *operationally equivalent*. From the Bridgmanian perspective, the identification of the observables of operationally equivalent measurements is physically unjustified. The mere statistical match of outcomes of two measurements which cannot be performed at the same time on the same system does not guarantee that the two measurements would give the same outcome run-by-run and hence that they measure the same observable.

Operational equivalence is sometimes expressed in the form that observables are not associated with measurement procedures, as in Bridgman, but with operators. So instead of *one measurement–one observable* we have *one operator–one observable*. Let me refer to the first identification of observables as *Bridgmanian* and to the second as *standard* (standard in quantum mechanics). Schematically:

	Bridgmanian		*Standard*	
Operator:	$\mathbf{O}$		$\mathbf{O}$	
	$\swarrow$	$\searrow$	$\downarrow$	
Observable:	$\mathcal{O}_1$	$\mathcal{O}_2$	$\mathcal{O}$	
	$\downarrow$	$\downarrow$	$\swarrow$	$\searrow$
Measurement:	M_1	M_2	M_1	M_2

As an example, consider the following two measurement procedures for photon polarization. "The first, which we denote by M_1, constitutes a piece of polaroid oriented to pass light that is vertically polarized along the $\hat{z}$ axis, followed by a photodetector. The second, which we denote by M_2, constitutes a birefringent crystal oriented to separate light that is vertically polarized along the $\hat{z}$ axis from light that is horizontally polarized along this axis, followed by a photodetector in the vertically polarized output." (Spekkens, 2005, p. 2). The two measurements are operationally equivalent: they provide the same distribution of outcomes for photons in any quantum state. Consequently, they are represented by the same operator, $\boldsymbol{\sigma}_z$, in quantum mechanics.

Now, do M_1 and M_2 measure the same observable or they measure different observables?[1] According to the standard quantum mechanical approach, they measure the same observable $\mathcal{O}$, the polarization of the photon. According to the Bridgmanian approach, they measure different observables: M_1 measures $\mathcal{O}_1$, the polarization of the photon with respect to a piece of polaroid, and M_2 measures $\mathcal{O}_2$, the polarization of the photon with respect to a birefringent crystal.

Note that the above Bridgmanian conditions to identify $\mathcal{O}_1$ and $\mathcal{O}_2$ are not fulfilled now: (i) M_1 and M_2 cannot be simultaneously measured on the very same photon since we need to decide on whether we insert a polaroid or a birefringent crystal in the path of the photon. (ii) The outcome of M_1 and M_2 will not necessarily will be the same for two photons prepared in the same state if this state is not an eigenstate of $\boldsymbol{\sigma}_z$. Thus, from the Bridgmanian perspective, the two observable should not be identified.

Even though from the Bridgmanian perspective the standard position is unsatisfactory, it has its own rationale. If all that quantum mechanics can predict is the distribution of outcomes, and if there are no preparations which would discern two measurements with respect to the outcome distribution, then why would one like to discern the two measurements? They measure the same observables, like gas thermometer and alcohol thermometer measure the same temperature, and any difference in the concrete realization of the measurements is just of secondary importance.[2]

[1] Note that Spekkens, as a good operationalist, does not use the term "observable" in his 2005 paper.

[2] Interestingly, we can defend the standard position even from a Bridgmanian perspective *when modifying the concept of measurement*. If we take a measurement in quantum mechanics not to be a single measurement but a sequence of measurements and an outcome not to be a single outcome but a statistical distribution of outcomes, then we can apply criterion (ii) in arguing that two sequences of operationally equivalent measurements measure the same observable: they have the same distribution of outcomes in every state, therefore they measure the same observable.

In any event, the standard position remains consistent as long as we remain at the level of quantum theory. But at the moment when we try to extend the ontology by ontic (hidden) states, the identification of observables corresponding different measurements represented by the same operator becomes problematic. The Kochen-Specker theorems highlight just this fact. It is instructive to see how Kochen-Specker theorems are interpreted on the standard approach (Held 2022, Spekkens 2005, Hofer-Szabó, 2021a, b, 2022). On this account, the lesson of the Kochen-Specker theorems is that the value of certain observables associated with operators depends on the measurement with which it is measured or co-measured. This fact is commonly referred to as contextuality—or "ontological contextuality" (Redhead, 1989) if not only the value but also the observables themselves depend on which measurement they are measured by. But note that from the Bridgmanian perspective there is nothing contextual in this fact; it simply shows that we were too quick to identify observables measured by different measurements when we relied simply on the match of the outcome statistics.

In this paper, I will revisit the Bridgmanian view of operationalism and investigate how far we get when we do *not* identify observables associated with operationally equivalent measurements. To this goal, I will use the framework of operational theories and ontological models introduced by Rob Spekkens (2005). This framework is general enough to embrace classical, quantum, super-quantum theories, and to analyze contextuality, causal structure and many other important features across the different theories. The main claim of the paper can be formulated at the more specific level of quantum mechanics and at the more general level of operational theories. As for quantum mechanics, this claim reads as follows:

Ontological models for quantum mechanics are sensitive not only to the operators but also to the measurements realizing these operators; more specifically, to whether these measurements can be performed simultaneously or not. The ontological models for these different measurements realizing the same set of operators in quantum mechanics but having a different compatibility structure can be highly different with respect to the causal structure, contextuality, and fine-tuning.

This strong dependence of the properties of the ontological models on the realizing measurements, however, is not restricted to quantum mechanics. It is a general feature of any operational theory. To show this, in the paper I will construct for an operational theory another operational theory with a different compatibility structure such that the measurements of the two theories are operationally equivalent, still they admit different ontological models. Thus, the main general claim of our paper is the following:

Ontological models for general operational theories are sensitive not only to the operationally equivalent classes of measurements but also to the measurements themselves, more specifically, to the compatibility structure of these measurements. Ontological models for operationally equivalent theories with different compatibility structures can be highly different with respect to the causal structure, contextuality, and fine-tuning.

More specifically, I will do the following. Operational theories come together with a set of measurements and a set of simultaneous measurements. For any operational theory, I will construct another theory with the following properties: a) the measurements of the new theory are operationally equivalent to the measurements of the old theory; b) in the new theory, there are no simultaneous measurements. I will call this procedure *trivialization*. With this procedure in hand, I will show the following:

1. The trivialization of an operational theory can be nicely represented graph theoretically as taking the line graph of the graph representing the original theory.
2. On the example of three non-disturbing (no-signaling) operational theories—a classical theory, the EPR-Bell scenario, and the Popescu-Rorhlich box, I will show how the most important features of the ontological models change when we replace an operational theory with a new, trivialized theory.
3. I will discern two different and logically independent concepts of contextuality, simultaneous contextuality and measurement contextuality, and show that the trivialization can alter the ontological models with respect to the former but not to the latter.

In the paper I will proceed as follows. After introducing the framework of operational theories (Sec. 2) and ontological models (Sec. 3), I define the procedure of trivialization (Sec. 4). Next, I compare the ontological model of three non-trivial (Sec. 5) and three corresponding trivial (Sec. 6) operational theories. I analyze the causal structure of the models (Sec. 7), show how trivialization leads to trivial causal graphs (Sec. 8), revisit the special case of quantum mechanics (Sec. 9) and conclude with a short discussion (Sec. 10).

2 Operational theories

The concept of measurements can be analyzed from several directions (Tal, 2020). Mathematical theories of measurement (Suppes, 1951) are concerned with the question of how to assign abstract terms to physical magnitudes. Operationalism and conventionalism, on the other hand, focus on the semantics of measurements, and defines the meaning of quantity-concepts in terms of operations (Bridgman, 1927) or 'coordinative definitions' (Reichenbach 1927). Realism addresses the metaphysical nature of measurable quantities and conceives of measurement results as approximations of the true values of physical quantities (Trout 1998). Information-theoretic and model-based accounts examine the epistemological aspects (informativity, coherence, consistency) of measurements (van Fraassen, 2008).

In this paper, I will take an operationalist perspective to measurements. This perspective has a long tradition in quantum theory starting with von Neumann (1932) and followed by Mackey (1957), Ludwig (1983), Busch et al. (2016), D'Ariano et al. (2017) and many others. Note, however, that, unlike many in the mathematical physics community, I will use the term 'measurement' in the original operationalist meaning referring to the physical procedure itself and not to its various mathematical representations, such as self-adjoint operators, PVMs, POVMs, effects, or whatever. My approach will follow the operational theories and ontological models framework of

Rob Spekkens (2005). In this and the next section, I introduce the main concepts of this framework.

An *operational theory* is a theory which specifies the probability of the outcomes of certain measurements performed on a physical system which was previously prepared in certain states. Let $\mathcal{P} = \{P_1, P_2, \dots\}$ be set of *preparations* of the system, $\mathcal{M} = \{M_1, M_2, \dots\}$ the set of *measurements* which can be performed on the system, and let $\mathcal{X} = \{X_1, X_2, \dots\}$ be the set of *outcomes*.[3] Let P, M, and X be random variables running over the preparations, measurements and outcomes, respectively, assigning to each event its index. (Thus, "$P = 1$" refers to the preparation P_1, "$M = 2$" refers to the measurement M_2, etc.) Using these random variables, an operational theory is simply a set of *conditional probabilities* of the outcomes given the various measurements and preparations, that is

$$p(X|M, P) \tag{1}$$

where P, M, and X run over the set $\mathcal{P}$, $\mathcal{M}$, and $\mathcal{X}$, respectively.

Two measurements M_1 and M_2 are *simultaneously measurable*, if they can be performed on the same system at the same time. Simultaneous measurability is an empirical question. Operationally, one identifies measurements by sets of laboratory instructions. The spin measurement of an electron, for example, is given by the detailed description of the path of the electron, the position of the Stern-Gerlach magnets and detectors, etc. As a consequence of this characterization of measurements by sets of laboratory instructions, two measurements M_1 and M_2 will be simultaneously measurable if and only if there is a measurement which can be identified by the *conjunction* of the sets of instructions characterizing M_1 and M_2. We call this measurement the *simultaneous measurement* of M_1 and M_2 and denote it by $M_1 \wedge M_2$ (which is again a measurement in $\mathcal{M}$). The random variable M will assign to $M_1 \wedge M_2$ the pair $(1, 2)$ and the outcomes of $M_1 \wedge M_2$ are taken from the set $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)}$. From the definition of simultaneous measurements it also follows that $M_1 \wedge M_2 \wedge M_3 \in \mathcal{M}$ implies $M_1 \wedge M_2 \in \mathcal{M}$. If M_1 and M_2 are not simultaneously measurable, we write $M_1 \wedge M_2 \notin \mathcal{M}$. If a measurement in an operational theory is not a simultaneous measurement of two or more other measurements, then we call it a *basic measurement*.

Note that $M_1 \wedge M_2$ and M_1 are simultaneous measurements since the conjunction of the sets of instructions characterizing $M_1 \wedge M_2$ and M_1 is just the set characterizing $M_1 \wedge M_2$. Similarly, a measurement and a certain *marginalization* (see below) of this measurement are simultaneous measurements since this latter measurement is just the measurement plus some extra instructions.

An important consequence of defining measurements by sets of instructions is that we do *not* identify two measurements just because they are operationally equivalent. Two measurements M_1 and M_2 are called *operationally equivalent* and denoted by $M_1 \sim M_2$ if they yield the same probability distribution of outcomes in every preparation[4] of the system, that is if

[3]Without loss of generality, we can assume that all (basic, see below) measurements have the same set of outcomes. If not, we just add null-outcomes to the outcome set of some measurements.

[4]Which are again identified by sets of laboratory instructions.

$$p(X|M_1, P) = p(X|M_2, P) \quad (2)$$

Note that two operationally equivalent measurements are different measurements if they are defined by different set of instructions.[5]

A maximal set of basic measurements which can be performed simultaneously on a system in an operational theory is called a *context*. M_1 and M_2 are in the same context if and only if $M_1 \wedge M_2 \in \mathcal{M}$. If $\{M_1, M_2, M_3\}$ is a context, then we call $M_1 \wedge M_2 \wedge M_3$ a *maximally simultaneous measurement* and $M_1 \wedge M_2$ a *non-maximally simultaneous measurement*. The set of all contexts is a *compatibility structure* of the theory. If in an operational theory there are no two measurements which can be simultaneously measured, then the compatibility structure is the empty set. We also refer to such operational theories as *trivial*.

We call two operational theories *operationally equivalent* (with respect to their measurement) if any measurement in the one theory is operationally equivalent with a measurement in the other theory or with a marginalization thereof (and the preparations are the same).

We call an operational theory *non-disturbing*[6] if no conditional probability depends on whether the measurements are performed alone or along with simultaneous measurements, that is:

$$p(X^{(i)}|M_i, P) = p(X^{(i)}|M_i \wedge M_j, P) \quad (3)$$

for any simultaneous measurement $M_i \wedge M_j \in \mathcal{M}$; otherwise, the operational theory is called *disturbing*. Obviously, trivial operational theories are non-disturbing.

Next, we introduce a graph theoretical representation of operational theories borrowed from the literature on the Kochen-Specker theorems (Kochen and Specker, 1967). In Figure 1, we depicted the graph of two Kochen-Specker theorems, the GHZ theorem (Greenberger et al., 1990) on the left and the Peres-Mermin square (Peres, 1990; Mermin, 1993) on the right. The vertices of the graph represent self-adjoint operators and a subset of vertices is connected by a (hyper)edge[7] if and only if the corresponding operators are pairwise commutating. In the GHZ graph one has 10 operators and 5 commuting subsets; in the Peres-Mermin graph one has 9 operators and 6 commuting subsets. The (hyper)graph of most of the Kochen-Specker theorems is *linear* which means that each pair of hyperedges intersects in at most one vertex.

In this paper, we take over this graphic representation and use it in the framework of the operational theories but with a different meaning. Vertices will represent here basic measurements and (hyper)edges will represent maximal sets of simultaneous measurements, that is contexts. In this interpretation, the above GHZ graph represents

[5] Operational equivalence can be introduced into an operational theory inductively and successively: one starts with a set of measurements and preparations and render measurements equivalent which provide the same outcome statistics in every preparations. This equivalence class is relative to the set of preparations; a new preparation procedure can break down operational equivalence if it discerns some measurements with respect to their outcome statistics.

[6] Or *no-signaling*, if the measurements are spacelike separated.

[7] A hyperedge can connect more than two vertices.

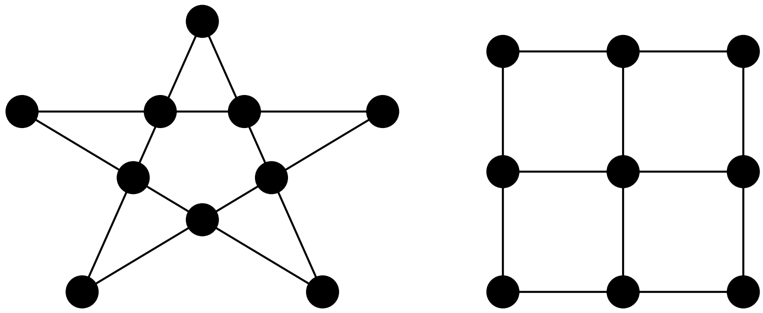

Fig. 1 The GHZ graph and Peres-Mermin graph

a non-trivial operational theory with 10 basic measurement arranged in 5 contexts and the Peres-Mermin graph represents a non-trivial theory with 9 basic measurement and 6 contexts. The (hyper)graph of both theory is linear: each basic measurement is featuring in exactly two contexts.

3 Ontological models

The role of an *ontological model* (hidden variable model) is to account for the conditional probabilities of an operational theory in terms of underlying *ontic states* (hidden variables, elements of reality, beables, real states) of the measured system. Let the set of ontic states be $\mathcal{L}=\{\Lambda_1, \Lambda_2, \dots\}$ and let the random variable over $\mathcal{L}$ be Λ. An ontological model specifies a *probability distribution* over the ontic states associated with each preparation:

$$p(\Lambda|P) \tag{4}$$

and a set of *response functions* that is a set of conditional probabilities associated with every measurement and every ontic state:

$$p(X|M,\Lambda) \tag{5}$$

again with the obvious normalizations. Assuming the independence of the probability distributions from the measurements, called *no-conspiracy*:

$$p(\Lambda|M,P) = p(\Lambda|P) \tag{6}$$

and the independence of the response functions from the preparations in which the ontic states are featuring, called *l-sufficiency*:

$$p(X|M,P,\Lambda) = p(X|M,\Lambda) \tag{7}$$

and using the theorem of total probability, one can recover the operational theory from the ontological model in terms of the probability distributions and response functions:

$$p(X|M,P) = \sum_{\Lambda} p(X|M,\Lambda)\, p(\Lambda|P) \tag{8}$$

An ontological model is called *outcome-deterministic (value-definite)* if

$$p(X|M, \Lambda) \in \{0, 1\} \tag{9}$$

otherwise it is called *outcome-indeterministic*.

Next, we define two different and logically independent concepts of noncontextuality (see Hofer-Szabó, 2021a, b, 2022). First, an ontological model is called *simultaneous noncontextual* if every ontic state determines the probability of the outcomes of every measurement independently of what other measurements are simultaneously performed, that is

$$p(X^{(i)}|M_i, \Lambda) = p(X^{(i)}|M_i \wedge M_j, \Lambda) \tag{10}$$

for any simultaneous measurement $M_i \wedge M_j \in \mathcal{M}$; otherwise the model is called *simultaneous contextual*. Simultaneous noncontextuality is a kind of inference to the best explanation for why an operational theory is non-disturbing: if the ontological model for an operational theory is noncontextual in the sense of (10), then—assuming no-conspiracy (6) and l-sufficiency (7)—one can show that the operational theory is non-disturbing (3).

Second, an ontological model is called *measurement noncontextual* if any two *operationally equivalent* measurements, that is $M_i, M_j \in M$ which have the same probability distribution of outcomes in every preparation

$$p(X|M_i, P) = p(X|M_j, P) \tag{11}$$

also have the same probability distribution of outcomes in every ontic state

$$p(X|M_i, \Lambda) = p(X|M_j, \Lambda) \tag{12}$$

Otherwise the model is called *measurement contextual*. Measurement noncontextuality is again a kind of inference to the best explanation; in this case the explanation of operational equivalence: (12)—together with no-conspiracy (6) and l-sufficiency (7)—implies (11) (Hofer-Szabó, 2021a, Lin 2021).

In quantum mechanics where operationally equivalent measurements $M_1 \sim M_2$ are represented by the same operator $\mathbf{O}$, measurement noncontextuality is just the requirement that the response functions of an ontological model should depend only on the operator and not on which specific measurement is realizing the operator, that is

$$p(X|M_1, \Lambda) = p(X|M_2, \Lambda) = p(X|\mathbf{O}, \Lambda)$$

Note, that trivial operational theories are trivially simultaneously noncontextual (since there are no simultaneous measurements) but they still can be measurement contextual. Also note that although simultaneous noncontextuality and measurement noncontextuality are different and logically independent notions, in case of non-disturbing theories measurement noncontextuality implies simultaneous noncontextuality: if M_j does not disturb M_i, then (11) holds for M_i and $M_i \wedge M_j$ (with $X = X^{(i)}$), but then, due to measurement noncontextuality, also (12), which is just simultaneous noncontextuality (10).

4 Trivialization

With the framework of operational theories and ontological models in hand, we can now formulate the main claim of our paper more precisely. This claim was the following: ontological models for quantum mechanics, and generally, for any operational theory sensitively depend on which measurement we choose from the class of operationally equivalent measurements. To show this dependence, I will investigate operational theories in pairs such that the two theories have operationally equivalent measurements but the first operational theory *does have* and the second operational theory *does not have* simultaneous measurements. In other words, the first theory has a non-trivial compatibility structure and the second theory has a trivial one. More precisely, I will provide a construction, which I call *trivialization*, assigning to any non-trivial operational theory a trivial theory. This construction will yield us pairs of operational theories which then can be compared with respect to the ontological models they admit and with respect to such properties as the causal structure, contextuality, fine-tuning, etc. We will see how sensitively the ontological models depend on whether the operational model is trivial or not.

In this section, I will only outline the procedure of trivialization and show some of its graph theoretical properties. In the next two sections, I will compare—on a classical, a quantum, and a super-quantum mechanical example—the ontological models of three non-trivial and three corresponding trivial theories. In Section 9, I return to the quantum mechanical example in order to highlight that operational theories realizing the same set of operators by different measurements can be vastly different.

Let us now turn to the trivialization. Consider an operational theory with a non-trivial compatibility structure that is a theory which comprises both basic and simultaneous measurements. *Trivialization* then consists in the following procedure:

Replace some (or perhaps all) of the measurements in the non-trivial operational theory with new, operationally equivalent measurements such that in the resulting operational theory there are no two measurements which can be performed simultaneously.

An example might help. Consider a non-disturbing operational theory with the following set of measurements:

$$\mathcal{M} = \{M_1, M_2, M_3, M_4, M_5,\ M_1 \wedge M_2,\ M_1 \wedge M_2 \wedge M_3,\ M_1 \wedge M_4\}$$

The theory has five basic measurements, M_1, M_2, M_3, M_4, M_5; one non-maximal simultaneous measurement $M_1 \wedge M_2$; and two maximally simultaneous measurements, $M_1 \wedge M_2 \wedge M_3$ and $M_1 \wedge M_4$ corresponding to the contexts $\{M_1, M_2, M_3\}$ and $\{M_1, M_4\}$, respectively.

The trivialization of $\mathcal{M}$ consist in the the following steps. We keep the basic measurement M_5 and replace $M_1 \wedge M_2 \wedge M_3$ and $M_1 \wedge M_4$ with two new measurements M_{123} and M_{14} with outcome sets $\mathcal{X}^{(123)}$ and $\mathcal{X}^{(14)}$, respectively. Note that M_{123} and M_{14} are completely new measurements and not the conjunctions $M_1 \wedge M_2 \wedge M_3$ and $M_1 \wedge M_4$. The simple reason we denote them by multiple indices is to relate them to these conjunctions. These new measurements are operationally equivalent to the old maximally simultaneous measurements:[8]

$$M_{123} \sim M_1 \wedge M_2 \wedge M_3 \quad \text{and} \quad M_{14} \sim M_1 \wedge M_4$$

This means that for every preparation:

$$\begin{aligned} p(f(X^{(1)} \wedge X^{(2)} \wedge X^{(3)})|M_{123}, P) &= p(X^{(1)} \wedge X^{(2)} \wedge X^{(3)}|M_1 \wedge M_2 \wedge M_3, P) \\ p(g(X^{(1)} \wedge X^{(4)})|M_{14}, P) &= p(X^{(1)} \wedge X^{(4)}|M_1 \wedge M_4, P) \end{aligned}$$

where f is a bijection mapping the outcomes $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)}$ of $M_1 \wedge M_2 \wedge M_3$ to the outcomes $\mathcal{X}^{(123)}$ of M_{123}, and g is another bijection mapping the outcomes $\mathcal{X}^{(1)} \times \mathcal{X}^{(4)}$ of $M_1 \wedge M_4$ to the outcomes $\mathcal{X}^{(14)}$ of M_{14}.

The measurements M_5, M_{123} and M_{14} are basic measurements and the operational theory is *trivial* since no measurements can be simultaneously measured. Since the theory is non-disturbing, the other five old measurements, M_1, M_2, M_3, M_4 and $M_1 \wedge M_2$ will operationally equivalent to certain coarse-graining or marginalization of the new basic measurements, M_{123} and M_{14}. To see this, let us introduce the following notation. Let $f(X^{(1)})$ denote the *union* of those outcomes of $M_1 \wedge M_2 \wedge M_3$ which are assigned to the outcome $X^{(1)}$ of M_1 by the bijection f. Let $f(X^{(2)})$ and $f(X^{(3)})$ be defined in a similar way. Furthermore, let $f(X^{(1)} \wedge X^{(2)})$ denote the *union* of those outcomes of $M_1 \wedge M_2 \wedge M_3$ which are assigned to the outcome $X^{(1)} \wedge X^{(2)}$ of $M_1 \wedge M_2$ by the bijection f. Finally, let $g(X^{(1)})$ denote the *union* of those outcomes of $M_1 \wedge M_4$ which are assigned to the outcome $X^{(1)}$ of M_1 and let $g(X^{(4)})$ be similarly defined. Then, the five old measurements will be operationally equivalent to the following marginalization of the new basic measurements:

$$\begin{aligned} p(f(X^{(1)})|M_{123}, P) = p(g(X^{(1)})|M_{14}, P) &= p(X^{(1)}|M_1, P) \\ p(f(X^{(2)})|M_{123}, P) &= p(X^{(2)}|M_2, P) \\ p(f(X^{(3)})|M_{123}, P) &= p(X^{(3)}|M_3, P) \\ p(g(X^{(4)})|M_{14}, P) &= p(X^{(4)}|M_4, P) \\ p(f(X^{(1)} \wedge X^{(2)})|M_{123}, P) &= p(X^{(1)} \wedge X^{(2)}|M_1 \wedge M_2, P) \end{aligned}$$

For the operational equivalence of these marginalizations, we will use the short hand

$$M_{123}^{(1)} \sim M_{14}^{(1)} \sim M_1, \quad M_{123}^{(2)} \sim M_2, \quad M_{123}^{(3)} \sim M_3, \quad M_{14}^{(4)} \sim M_4, \quad M_{123}^{(12)} \sim M_1 \wedge M_2$$

where $M_{123}^{(1)}$ denotes the measurement that we first perform the measurement M_{123}, and then coarse-grain the outcomes into blocks such that each block corresponds—due to the bijection f— to an outcome $X^{(1)}$ of M_1.

[8]Obviously, it is a theoretical-experimental question whether such new measurements exist. In the quantum mechanical scenario which we use in the paper they do, and we will explicitly show them in Section 9.

To sum up, the new operational theory will be the following

$$\mathcal{M}' = \{M_{123}, M_{14}, M_5\}$$

The two operational theories $\mathcal{M}$ and $\mathcal{M}'$ are operationally equivalent, $\mathcal{M} \sim \mathcal{M}'$, since any measurement in $\mathcal{M}$ is operationally equivalent to a measurement or a specific marginalization of a measurement in $\mathcal{M}'$ and vica versa. They are, however, different since they contain different measurements (except for the common M_5). $\mathcal{M}$ is non-trivial but $\mathcal{M}'$ is trivial, it contains no simultaneous measurements.

Note again that even though M_{14} is indexed by two indices, it is just as a basic measurement in the new operational theory as M_1 and M_4 was in the old theory. The only reason why we use these multiple indices is to be able to relate M_{14} to M_1 and M_4 simply by marginalization and operational equivalence. This notation, however, should not blur the fact that M_{14} can be a simple measurement.[9]

Now, let us turn to the graphic representation of the trivialization. As we saw, trivialization results in an operational theory with trivial compatibility structure. If we represent this new operational theory by a graph, this graph will have only vertices but no edges. The first two graphs in Figure 2 show the graph of our above mini operational theory and the trivialized new theory.

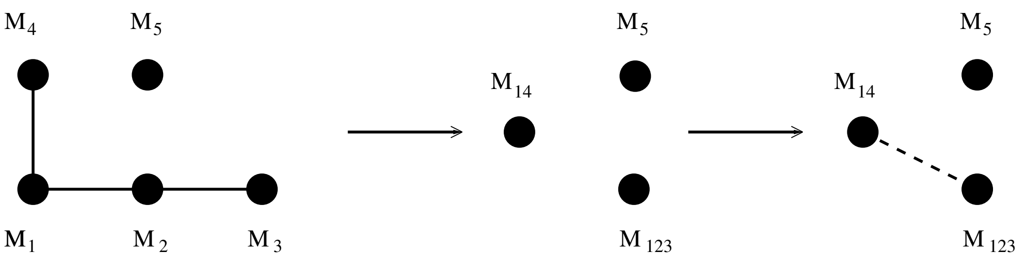

Fig. 2 The graph and line graph of the toy operational theory

We could stop at this point but then some information would be lost, namely, that certain marginalizations of the new measurements are operationally equivalent. To preserve this information, we add (hyper)edges to the graph of the new theory *with the following meaning*: we draw an (hyper)edge between a set of vertices in the trivial theory, if the corresponding basic measurements have operationally equivalent marginalizations.[10] For example, the graph of our mini operational will get an edge (see the third graph in Figure 2) because the appropriate marginalization M_{123} and M_{14} are operationally equivalent to one another (both being operationally equivalent to M_1). Note that the (hyper)edges in the graph of the non-trivial and trivial theories

[9] Just to stress this fact again, one need not think of M_{14} as being two measurements performed on two different subsystems, as in the usual spin measurement scenarios in quantum mechanics. M_{14} can also be a measurement on a localized system.

[10] Or equivalently, if the contexts conforming to the maximally simultaneous measurements in the non-trivial theory (which are represented by the vertices in the trivial theory) had at least one common basic measurement.

mean different things: in the non-trivial operational theory they meant simultaneous measurability, while in the trivial operational theory they mean having operationally equivalent marginalizations. To express this difference, we use continuous lines in the non-trivial operational theories and broken lines in the trivial ones.

This construction can be nicely represented graph theoretically by simply taking the line graphs of the (hyper)graph of the non-trivial operational theory. A *line graph* $L(G)$ is constructed from a graph G such that for each (hyper)edge in G we make a vertex in $L(G)$ and for every two (hyper)edges in G that have a vertex in common, we make an edge between their corresponding vertices in $L(G)$. The line graphs of the GHZ graph and Peres-Mermin graph, for example, are depicted in Figure 3. The

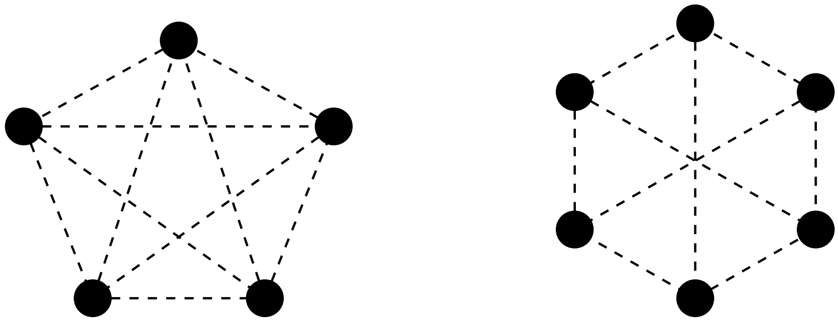

Fig. 3 The line graphs of the GHZ graph and Peres-Mermin graph

number of the vertices and edges flip in both line graphs: the line graph of the GHZ graph contains 5 vertices and 10 edges, the line graph of the Peres-Mermin graph contains 6 vertices and 9 edges. Since both the GHZ graph and Peres-Mermin graph are linear, their line graphs contain only edges but no hyperedges.

To sum up, in the graph G of the non-trivial operational theory, vertices represent the old basic measurements and (hyper)edges represented contexts that is sets of simultaneous measurements. In the line graph $L(G)$ of the trivialized theory, vertices represent the new basic measurements but—since there are no simultaneous measurements— the (hyper)edges mean something else: they connect vertices representing measurements which have operationally equivalent marginalizations.

5 Three operational theories with non-trivial compatibility structure

In this Section, we consider three non-disturbing and non-trivial operational theories, all of the form

$$\mathcal{M} = \{A_0, A_1, B_0, B_1,\ A_0 \wedge B_0,\ A_0 \wedge B_1,\ A_1 \wedge B_0,\ A_1 \wedge B_1\}$$

Each theory has the same four basic measurements A_0, A_1, B_0, B_1 such that A_0, A_1 have binary outcomes X_0, X_1 and B_0, B_1 have binary outcomes Y_0, Y_1. In all three theories, any A-measurement is simultaneously measurable with any B-measurement but neither the two A-measurements nor the two B-measurements can be simultaneously measured. In short, the compatibility structure of all three theories will be

$$\big\{\{A_0, B_0\}, \{A_0, B_1\}, \{A_1, B_0\}, \{A_1, B_1\}\big\}$$

Consequently, the graph (and line graph, see next section) depicted in Figure 4 is the same for all three operational theories. Since the graph is linear, the line graph contains only edges and no hyperedges.

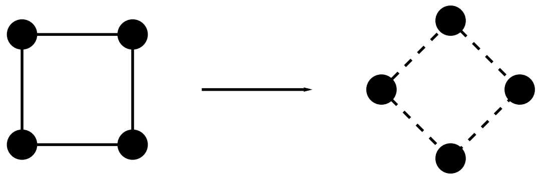

Fig. 4 The graph and line graph of the three operational theories

Let A be a random variable over the measurements $\{A_0, A_1\}$ and B a random variable over the measurements over the measurements $\{B_0, B_1\}$. Similarly, let X and Y be random variables over the outcomes $\{X_0, X_1\}$ and $\{Y_0, Y_1\}$, respectively, such that $A, B, X, Y = 0, 1$. The operational theories differ in the preparations. Each theory has only one preparation: the first one P_{CL}, the second P_{EPR}, and the third P_{PR}. We refer to the operational theories as a *classical operational theory*, the *EPR-Bell situation*, and the *Popescu-Rorhlich (PR) box* (Popescu and Rohrlich, 1994), respectively.

The three operational theories can be characterized by the following conditional probabilities:

$$p(X|A, P) = p(Y|B, P) = \frac{1}{2} \tag{13}$$

$$p(X, Y|A, B, P_{\text{CL}}) = \begin{cases} \frac{1}{2} & \text{if } X \oplus Y = 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$p(X, Y|A, B, P_{\text{EPR}}) = \begin{cases} \frac{3}{8} & \text{if } X \oplus Y = 0 \text{ and } A \cdot B = 0 \\ \frac{1}{8} & \text{if } X \oplus Y = 1 \text{ and } A \cdot B = 0 \\ \frac{1}{2} & \text{if } X \oplus Y = 0 \text{ and } A \cdot B = 1 \\ 0 & \text{if } X \oplus Y = 1 \text{ and } A \cdot B = 1 \end{cases} \tag{15}$$

$$p(X, Y|A, B, P_{\text{PR}}) = \begin{cases} \frac{1}{2} & \text{if } X \oplus Y = A \cdot B \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where P is a variable over $\mathcal{P} = \{P_{\text{CL}}, P_{\text{EPR}}, P_{\text{PR}}\}$ and $\oplus$ is the sum modulo 2. We come back to the quantum mechanical representation of the EPR-Bell situation in Section 9.

All three operational theories are non-disturbing:

$$p(X|A,P) = p(X|A,B,P) = \frac{1}{2} \tag{17}$$
$$p(Y|B,P) = p(Y|A,B,P) = \frac{1}{2} \tag{18}$$

but they they represent three different classes of theories. The first is a classical theory, the second is a quantum mechanical, the third is a super-quantum mechanical theory. This difference is manifested in the satisfaction/violation of the CHSH inequality (Clauser, Horne, Shimony, and Holt, 1969). Namely, the CHSH expression

$$\text{CHSH}_P = \langle A_0, B_0\rangle_P + \langle A_0, B_1\rangle_P + \langle A_1, B_0\rangle_P - \langle A_1, B_1\rangle_P \tag{19}$$

where

$$\langle A, B\rangle_P = p(X \oplus Y = 0|A,B,P) - p(X \oplus Y = 1|A,B,P)$$

is 2 for the classical theory, satisfying the CHSH inequality, $|\text{CHSH}_P| \leqslant 2$; it is 2.5 for the EPR-Bell situation, violating (not maximally) the CHSH inequality; and 4 for the PR box which is beyond the Tsirelson bound $2\sqrt{2}$.

Next, we construct an ontological model for each operational theory. The exact probabilistic specification of the models in terms of distributions and response functions is given in the Appendix. From our perspective, however, it will be more instructive to look at the *bundle diagrams* (see Abramsky et al., 2017; Abramsky & Brandenburger, 2011) of the models depicted in Figure 5.

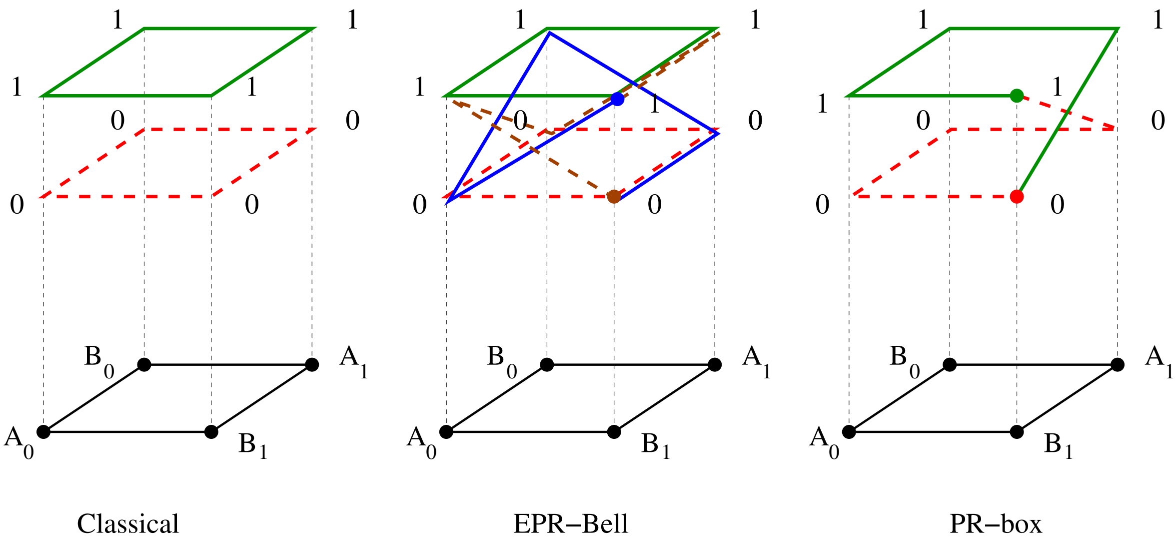

Fig. 5 Bundle diagrams of the ontological models for the operational theories with non-trivial compatibility structure

First, look at the "cuboid" of the classical model on the left. The quadrangle at the bottom is the base space of the bundle, actually the graph of the operational theory "laid down". It consists of four vertices representing the four measurements A_0, A_1, B_0, B_1 such that two measurements are connected if and only if they are in the same context. The vertical broken lines are the fibers of the bundle. The two vertices on a given fiber at different heights denoted by 0 and 1 represent the outcomes of the corresponding measurements: $X = 0, 1$ for $A = 0, 1$ and $Y = 0, 1$ for $B = 0, 1$. Now, there are two quadrangles in the figure, one connecting the upper vertices of the adjacent fibers and one connecting the lower vertices. Each quadrangle represents the response functions of the model for a given ontic state. The green and continuous upper quadrangle represents the ontic state Λ_1. In this ontic state the outcome of each measurement is 1. (The outcome of A_0, A_1 is X_1 and the outcome of B_0, B_1 is Y_1.) The red and broken lower quadrangle represents the ontic state Λ_0 for which each outcome is 0. The model is outcome-deterministic. It also fixes the outcomes of the simultaneous measurements in the different contexts such that no outcome of any measurement in any ontic state depends on whether a simultaneous measurement is also performed. Thus, the model is simultaneous noncontextual. Moreover, the model is also measurement noncontextual: in both ontic state the outcome of any two operationally equivalent measurements is the same. Setting the probability of both ontic states to $\frac{1}{2}$, the operational theory can be recovered.

Let us now go over to the bundle diagram of the PR-box on the right of Figure 5. Again, we have two ontic states Λ_1 and Λ_0 but the green and red lines do not close up now. They are discontinuous at the fibre of B_1. To avoid ambiguity with respect to the outcome of B_1, we put a dot at the one end of both discontinuous lines. This dot indicates the outcome of B_1 if measured *alone* and not together with A_0 or A_1 (when the outcome of B_1 is indicated by the value of the appropriate segment of the green or red lines connecting the fibre of B_1 with the fibre of A_0 or A_1). Thus, the model is outcome deterministic. However, it is simultaneous contextual:

$$\delta_{Y,\Lambda} = p(Y|B_1, \Lambda) \neq p(Y|A_1 \wedge B_1, \Lambda) = \delta_{Y \oplus 1,\Lambda} \tag{20}$$

That is performing the measurement B_1 in the "green" ontic state, Λ_1, together with A_1, the outcome of B_1 will be Y_0, while performing B_1 together with A_0, the outcome will be Y_1; and *vice versa* for the "red" ontic state, Λ_0. Since simultaneous contextuality implies measurement contextuality for non-disturbing theories, the model for the PR-box will also be measurement contextual. Indeed,

$$p(Y|B_1, P_{\text{PR}}) = p(X|A_1 \wedge B_1, P_{\text{PR}}) \tag{21}$$

despite the fact that inequality (20) holds. We can recover the PR-box theory again by setting the probability of both ontic states to $\frac{1}{2}$.

Finally, the bundle diagram in the middle of Figure 5 represents an ontological model for the EPR-Bell scenario. Here we have four ontic states portrayed by lines of different color and style. The "green" and "red" ontic states are outcome deterministic and noncontextual in both senses. The "blue" and "brown" ontic states, however, are outcome deterministic but simultaneous and hence measurement contextual: their lines do not close on the fibre of B_1. This means that in these ontic states the outcome of B_1 will be different when measured alone and when co-measured with A_0 or A_1. The dots at one end of the lines indicate the outcomes the outcome of B_1 when measured alone. By setting the probability of the two noncontextual ontic states to $\frac{3}{8}$ and the probability of the two contextual ontic states to $\frac{1}{8}$, the probabilities of the EPR-Bell scenario can be recovered (see Appendix).

To sum up, we constructed three (among the many) outcome-deterministic ontological models for the three operational theories such that the model for the classical theory is noncontextual (in both senses) and the models for other two theories are contextual (again, in both senses). This is in tune with the satisfaction and violation of the CHSH inequality for the different theories.

6 Three operational theories with trivial compatibility structure

The three operational theories in the previous Section were non-trivial, they had a non-trivial compatibility structure. Let us now "trivialize" them in the way outlined in Section 4 and investigate the ontological models for these trivialized theories. Trivialization consists in replacing each simultaneous measurement

$$A_0 \wedge B_0, \quad A_0 \wedge B_1, \quad A_1 \wedge B_0, \quad A_1 \wedge B_1$$

with an operationally equivalent new basic measurement:

$$\begin{aligned} C_{00} &\sim A_0 \wedge B_0, \qquad & C_{01} &\sim A_0 \wedge B_1 \\ C_{10} &\sim A_1 \wedge B_0, \qquad & C_{11} &\sim A_1 \wedge B_1 \end{aligned}$$

Note that C_{00}, C_{01}, C_{10} and C_{11} cannot be measured simultaneously.

Let the outcome space of each of the measurements in $\mathcal{C} = \{C_{00}, C_{01}, C_{10}, C_{11}\}$ be the same $\mathcal{Z} = \{Z_{00}, Z_{01}, Z_{10}, Z_{11}\}$. Let C be a random variable over $\mathcal{C}$ assigning to every measurement its index pair. Similarly, let Z be a random variable over $\mathcal{Z}$ assigning to every outcome its index pair. Both C and Z can be expressed as a Cartesian product: $C = C_1 \times C_2$ and $Z = Z_1 \times Z_2$ where C_1 and Z_1 assign to every measurement or outcome its first index and C_2 and Z_2 assign the second index.

Since the old operational theory is non-disturbing, the following marginalizations of the new basic measurements are operationally equivalent to the old basic measurements:[11]

$$\begin{aligned} p(Z_1|C_{00}, P) = p(Z_1|C_{01}, P) &= p(X|A_0, P) \\ p(Z_1|C_{10}, P) = p(Z_1|C_{11}, P) &= p(X|A_1, P) \end{aligned}$$

[11]Note that Z_1 is the *union* of those two outcomes of C_{00}, C_{01}, C_{10} and C_{11} which has the same first index. In the terminology introduced in Section 4, Z_1 is the *union* of those outcomes which are assigned to the outcome X of A_0 by a bijection f_0 or to the outcome X of A_1 by a bijection f_1 and also to the outcome Y of B_0 by a bijection g_0 or to the outcome Y of B_1 by a bijection g_1. To keep the notation simple, we drop these bijections and write simply Z_1 instead of $f_0(X)$, $f_1(X)$, $g_0(Y)$ and $g_1(Y)$.

$$p(Z_2|C_{00}, P) = p(Z_2|C_{01}, P) \; = \; p(Y|B_0, P)$$
$$p(Z_2|C_{01}, P) = p(Z_2|C_{11}, P) \; = \; p(Y|B_1, P)$$

or using the short hand introduced in Section 4:

$$C_{00}^{(1)} \sim C_{01}^{(1)} \sim A_0, \qquad C_{10}^{(1)} \sim C_{11}^{(1)} \sim A_1$$
$$C_{00}^{(2)} \sim C_{10}^{(2)} \sim B_0, \qquad C_{01}^{(2)} \sim C_{11}^{(2)} \sim B_1$$

That is, the old basic measurements A_0, A_1, B_0, B_1 can be recovered as marginalizations of the new basic measurements. Thus, the new trivialized operational theories will be:

$$\mathcal{M}' = \{C_{00}, C_{01}, C_{10}, C_{11}\}$$

$\mathcal{M}'$ is operationally equivalent to $\mathcal{M}$ introduced in the previous section, $\mathcal{M} \sim \mathcal{M}'$, or more precisely:

$$\mathcal{M}_{CL} \sim \mathcal{M}'_{CL}, \qquad \mathcal{M}_{EPR} \sim \mathcal{M}'_{EPR}, \qquad \mathcal{M}_{PR} \sim \mathcal{M}'_{PR}$$

The line graph of $\mathcal{M}$' is depicted on the right side of Figure 4.

The three trivial operational theories can be characterized by the following conditional probabilities:

$$p(Z|C, P_{\text{CL}}) \; = \begin{cases} \frac{1}{2} & \text{if } Z_1 \oplus Z_2 = 0 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

$$p(Z|C, P_{\text{EPR}}) \; = \begin{cases} \frac{3}{8} & \text{if } Z_1 \oplus Z_2 = 0 \text{ and } C_1 \cdot C_2 = 0 \\ \frac{1}{8} & \text{if } Z_1 \oplus Z_2 = 1 \text{ and } C_1 \cdot C_2 = 0 \\ \frac{1}{2} & \text{if } Z_1 \oplus Z_2 = 0 \text{ and } C_1 \cdot C_2 = 1 \\ 0 & \text{if } Z_1 \oplus Z_2 = 1 \text{ and } C_1 \cdot C_2 = 1 \end{cases} \tag{23}$$

$$p(Z|C, P_{\text{PR}}) \; = \begin{cases} \frac{1}{2} & \text{if } Z_1 \oplus Z_2 = C_1 \cdot C_2 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Observe that the probabilistic description of the trivial operational theories is formally analogous with that of the non-trivial theories of the previous section: we obtain equations (14)-(16) from (22)-(24) by simply replacing C_1, C_2, Z_1, Z_2 with A, B, X, Y, respectively. The measurements and outcomes, however, are different in the two theories.

All three operational theories are non-disturbing in a trivial sense: there are no simultaneous measurements. Therefore, the CHSH inequalities cannot be defined. Again, one can construct an ontological model for each operational theory. The distribution of ontic states is the same as in the models for the non-trivial theories. The response functions are obtained from those of the non-trivial theory by simply replacing A, B, X, Y with C_1, C_2, Z_1, Z_2. All this is specified in the Appendix and visualized in Figure 6. As can be seen, the lines representing the outcomes of

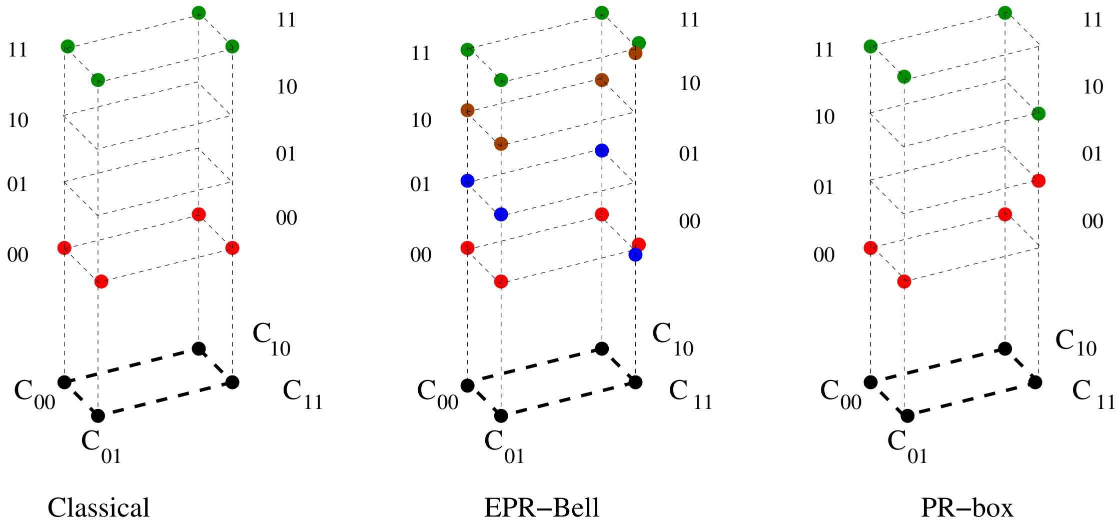

Fig. 6 Bundle diagrams of the ontological models for the operational theories with trivial compatibility structure

simultaneous measurements have disappeared. Each basic measurement has a definite outcome in every ontic state denoted by a dot at the appropriate height on the fibre corresponding to the measurement. In the classical theory and in the PR box there are two ontic states ("green" and "red"), in the EPR-Bell scenario there are four ontic states ("green", "red", "blue" and "brown"). All three models are outcome-deterministic and simultaneously non-contextual since there are no simultaneous measurement. But the non-classical (EPR and PR) models are measurement contextual. Certain marginalizations of the measurements, for example, $C^{(2)}_{01}$ and $C^{(2)}_{11}$ are operationally equivalent. Still, both the "blue" and "brown" ontic states in the EPR model and "green" and "red" ontic states in the PR model assign different outcomes to them. This shows that measurement noncontextuality is a stronger concept than simultaneous noncontextuality.

7 The causal structure of the ontological models

Let us turn now to the causal structure of the ontological models. Since these models provide information only about the probabilistic relations of the events and not about their spatiotemporal or other relations, the reconstruction of the causal structure will rely solely on these probabilistic information. The machinery to deduce causal relations from probabilistic relations is known as *causal discovery algorithms* and was introduced in (Pearl, 2009; Spirtes, Glymour, Scheines, 2001). These algorithms do not make use of the full probabilistic setting, they use only the conditional and unconditional independence relations to construct a causal graph. A causal graph is a directed acyclic graph (DAG),[12] where the vertices represent random variables and the directed edges represent causal relevance between these variables. For a variable X, the set of vertices that have directed edges in X is called the parents of X, denoted by $Par(X)$, and the set of vertices that are endpoints of a directed paths from X is called the descendants of X, denoted by $Des(X)$. A set V of random variables (on a classical probability space) is said to satisfy the *Causal Markov Condition* relative to a causal graph G if for any $X \in V$ and $Y \notin Des(X)$:

$$p(X|Par(X), Y) = p(X|Par(X))$$

That is, conditioning on its parents any random variable will be probabilistically independent from any of its non-descendants.

Now, causal discovery algorithms take as input a set of conditional and unconditional independence relations among random variables and provide a causal graph G as output which returns these independence relations if the Causal Markov Condition is applied to the graph.[13] Here we do not enter into the details of these algorithms; rather we simply list the independence relations of the ontological models of the non-trivial and trivial operational theory and the corresponding causal graphs.[14]

Let us start with the causal structure of the ontological models of the *non-trivial* operational theories, $\mathcal{M}$, introduced in Section 5. The conditional independence relations in the ontological models of our three non-trivial theories are the following:

$$\begin{aligned} p(X|A,B) &= p(X|A) && (25)\\ p(Y|A,B) &= p(Y|B) && (26)\\ p(X|A,Y,\Lambda) &= p(X|A,\Lambda) && (27)\\ p(Y|X,B,\Lambda) &= p(Y|B,\Lambda) && (28)\\ p(X|A,B,\Lambda) &= p(X|A,\Lambda) && (29)\\ p(Y|A,B,\Lambda) &\stackrel{\text{(CL)}}{=} p(Y|B,\Lambda) && (30) \end{aligned}$$

The first two relations are just the non-disturbance equations (17)-(18), the subsequent relations follow from the appropriate response functions (44)-(46), (49)-(51), and (54)-(56) of the models specified in the Appendix. The first five conditional independence relations (25)-(29) hold for all the three models but the last relation (30) holds only for the classical model.

[12]Note that these causal graphs are different from the graphs and line graphs used in the previous Sections representing compatibility structure and common marginalization.

[13]More precisely, the independence relations are returned if all those graphical criteria are applied to the graph which can be derived from the Causal Markov Condition plus the semi-graphoid axioms. These criteria are captured by the so-called d-separation criterion (see Pearl, 2009, Ch. 1).

[14]For the application of the causal discovery algorithm for the EPR-Bell scenario, see (Suarez, 2007; Suarez and SanPedro, 2009; Wood and Spekkens, 2015). Also note that the independence relations also include the ontic states. Thus, the causal discovery algorithms are not *discovery* algorithms in the sense that they are based solely on the empirically accessible probabilities.

The causal graphs which return the independences for the three models are depicted in Figure 7. These graphs are *minimal* in the sense that no subgraph can

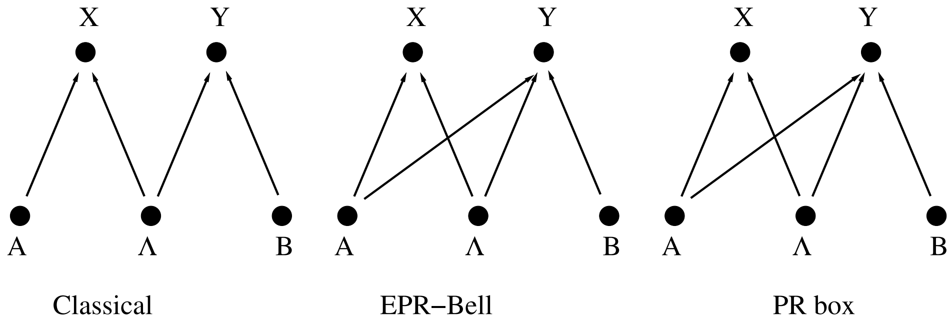

Fig. 7 Causal structure of the ontological models with non-trivial compatibility structure

return all the independence relations. Applying the Causal Markov Condition to the graphs, one obtains also an extra unconditional independence relation among the exogenous variables (that is variables which have no parents):

$$p(A, B, \Lambda) = p(A)p(B)p(\Lambda) \tag{31}$$

These relations are not specified in the model but are consistent with it. They are a special case of the *no-conspiracy* condition (6).

Observe that there is an edge in the graph of the non-classical models connecting A and Y. This edge represents the causal influence responsible for simultaneous contextuality: the value of Y causally depends not only on the value of X and Λ but also on the value of A. If A and Y are spacelike separated, this edge represents a non-local causal influence. Note again, however, that in constructing the graphs, we relied only on the probabilistic features of the models and not on the spatiotemporal localizations of the events—in strong contrast to the usual EPR-Bell analysis.

A further difference between the classical and non-classical models concerns fine-tuning. To see this, first recall that any joint probability distribution of the random variables which is compatible with the corresponding causal graphs in Figure 7 is of the form

$$p(X, Y, A, B, \Lambda) \overset{\text{(CL)}}{=} p(X|A, \Lambda)p(Y|B, \Lambda)p(A)p(B)p(\Lambda) \tag{32}$$

for the classical model and of the form

$$p(X, Y, A, B, \Lambda) \overset{\text{(EPR, PR)}}{=} p(X|A, \Lambda)p(Y|A, B, \Lambda)p(A)p(B)p(\Lambda) \tag{33}$$

for the non-classical models. In both equations, the conditional probabilities (the response functions) are called *causal parameters* and the unconditional probabilities are called *statistical parameters* (where $p(\Lambda)$ is just a short hand for $p(\Lambda|P)$). By manipulating these parameters, one obtains all the joint distributions compatible with the causal graphs. Since causal discovery algorithms are sensitive only to the independence relations and not to the full joint probability distribution, the question arises, whether these independence relations are robust enough against the perturbation of the causal-statistical parameters, that is whether they continue to hold when these parameters are not those specified in the Appendix but take on arbitrary values. If so, the graph is said to be *faithful*, if not, it is said to be *fine-tuned*.

Now, for the classical model all the conditional independences (25)-(30) can be derived from the joint probability distribution equation (32) plus the theorem of total probability. This means that the conditional independences hold for any choice of the parameters. Thus, the classical model is faithful. The crucial step in the derivation of the conditional independences is factorization

$$p(X, Y|A, B, \Lambda) \overset{\text{(CL)}}{=} p(X|A, \Lambda)p(Y|B, \Lambda)$$

By summing up for the different variables, one recovers the different conditional independences (25)-(30). In the non-classical models, however, one has

$$p(X, Y|A, B, \Lambda) \overset{\text{(EPR, PR)}}{=} p(X|A, \Lambda)p(Y|A, B, \Lambda)$$

instead of the factorization and hence summing up does *not* recover (30), (28) and the non-disturbance (26). And indeed, for a non-zero measure of the parameters, these conditional independences will fail to hold. Therefore, the non-classical models are fine-tuned.

These facts are in tune with Cavalcanti's (2018) theorem on bipartite Bell scenarios stating that every causal model for a non-disturbing operational theory violating the CHSH inequality requires fine-tuning. Cavalcanti's result highlights a deep connection between simultaneous contextuality of the model and fine tuning of the corresponding graph. At the end of his paper, he asks whether also *measurement* noncontextuality can be understood as arising from the no-fine-tuning condition.

To answer Cavalcanti's question, let us now turn to the causal structure of the ontological models of the *trivial theory*. In these models there are no conditional independence relations, except among the exogenous variables:

$$p(C, \Lambda) = p(C)p(\Lambda) \tag{34}$$

which is again consistent with the models. The causal graph which is compatible with (34) is depicted in Figure 8. Note, that the graph is the same for all three models. The four measurements cannot be simultaneously performed, therefore the models are (trivially) simultaneously noncontextual. The models are also faithful since any choice of the parameters in the joint probability distribution equation

$$p(Z, C, \Lambda) = p(Z|C, \Lambda)p(C)p(\Lambda) \tag{35}$$

compatible with the graph in Figure 8 will return the same independence relations, that is (34). Thus, if Cavalcanti's question is whether measurement contextual ontological models for an operational theory are also fine-tuned, then the answer is no. Both the models of the non-trivial and trivial non-classical operational theories are measurement contextual, still the causal graphs are fine-tuned for the former and

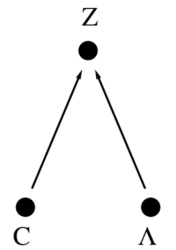

Fig. 8 Causal structure of the ontological models with trivial compatibility structure

faithful for the latter. Fine-tuning relates to simultaneous contextuality but not to measurement contextuality.

To sum up, the causal graph of the model of the classical and non-classical non-trivial operational theories are different; the graphs of the non-classical models are fine-tuned and contain a directed edge representing the causal connection responsible for simultaneous contextuality. This difference between the graphs collapses upon trivializing the theories; the graph of all three models will be the same: trivial and faithful.

8 How trivialization leads to trivial causal graphs?

In this section, I return to the concept of trivialization and show how it leads to trivial causal graphs. At the beginning of Section 6, we replaced the non-trivial operational theory

$$\mathcal{M} = \{A_0, A_1, B_0, B_1,\ A_0 \wedge B_0,\ A_0 \wedge B_1,\ A_1 \wedge B_0,\ A_1 \wedge B_1\}$$

with an operationally equivalent but *different* theory:

$$\mathcal{M}' = \{C_{00}, C_{01}, C_{10}, C_{11}\}$$

Why are the causal graphs different for $\mathcal{M}$ and $\mathcal{M}'$?

As stated in the previous section, causal graphs arise from the conditional independences of an ontological model via the causal discovery algorithms. Let us pick one such independence from the non-trivial theory, for example, (29) with $A = 0$:

$$p(X|A_0 \wedge B_0, \Lambda) = p(X|A_0 \wedge B_1, \Lambda) = p(X|A_0, \Lambda) \tag{36}$$

What would be the "analogue" of (36) in the trivial theory?

One has two options—unfortunately neither yielding a conditional independence. The first option is to replace $A_0 \wedge B_0$ by C_{00}, $A_0 \wedge B_1$ by C_{01} and A_0 by $C^{(1)}_{00}$ or $C^{(1)}_{01}$. Thus, we obtain:

$$p(Z_1|C_{00}, \Lambda) = p(Z_1|C_{01}, \Lambda) = p(Z_1|C^{(1)}_{00}, \Lambda) = p(Z_1|C^{(1)}_{01}, \Lambda) \tag{37}$$

But (37) contains no conjunction, only marginalization. Consequently, (37) is not a conditional independence. The other option is to replace, say, A_0 by $C^{(1)}_{00}$, B_0 by $C^{(2)}_{10}$

and B_1 by $C_{11}^{(2)}$. We obtain

$$p(Z_1|C_{00}^{(1)} \wedge C_{10}^{(2)}, \Lambda) = p(Z_1|C_{00}^{(1)} \wedge C_{11}^{(2)}, \Lambda) = p(Z_1|C_{00}^{(1)}, \Lambda) \tag{38}$$

The problem with (38) is that it is ill-defined: the conjunctions $C_{00}^{(1)} \wedge C_{10}^{(2)}$ and $C_{00}^{(1)} \wedge C_{11}^{(2)}$ are not defined since C_{00}, C_{10} and C_{11} cannot be simultaneously measured. And this is as it should be since $\mathcal{M}'$ is a trivial theory.

In sum, trivialization removes conditional independences because in the trivialized theory there will be no simultaneous measurements. Consequently, trivialization leads to trivial causal graphs.

An anonymous referee suggested the following recovery of the conditional independence (36). First, allow for an operational theory to contain measurements which are defined not only by *conjunctions* but also by *disjunctions* of instructions. That is, allow for measurements like, for example, $C_{00} \vee C_{10}$. Now, consider the following operational theory

$$\mathcal{M}'' = \{C_{00} \vee C_{01}, C_{11} \vee C_{01}, C_{11} \vee C_{10}, C_{00} \vee C_{10},\, C_{00}, C_{01}, C_{10}, C_{11}\}$$

Since C_{00} is a conjunction of $C_{00} \vee C_{01}$ and $C_{00} \vee C_{10}$ (and similarly for C_{01}, C_{10}, C_{11}), $\mathcal{M}''$ will be a nontrivial theory. Then the conditional independence (36) will look like this:

$$p(Z_1|(C_{00} \vee C_{01}) \wedge (C_{00} \vee C_{10}), \Lambda) = p(Z_1|(C_{00} \vee C_{01}) \wedge (C_{11} \vee C_{01}), \Lambda) = p(Z_1|C_{00} \vee C_{01}, \Lambda) \tag{39}$$

Equation (39) indeed recovers (36); so embracing disjunctions seems to bring back non-trivial theories and non-trivial causal structure. However, the operational theory $\mathcal{M}''$ is different from $\mathcal{M}'$. Namely, an operational theory is defined by a set of measurements (and preparations). If one extends the set of measurements either by conjunctions or by disjunctions, one gets *another* operation theory. This extended theory can well have a causal structure which is different from that of the original theory. But even if $\mathcal{M}''$ has a non-trivial causal structure, this does not invalidate our central claim that the causal structure of $\mathcal{M}'$ is trivial while the causal structure of $\mathcal{M}$ which is operationally equivalent to $\mathcal{M}'$ is non-trivial.

To be clear, I have no problem with allowing for *disjunctive* measurements ("measure polarization along axis x *or* measure polarization along axis y") in an operational theory, as in the above example. What is important, however, is to keep in mind that in the definition of simultaneous contextuality (10) *only conjunctions* pop up. Therefore, the dividing line draws between theories in which there are only basic measurements—be them disjunctive or not—and theories in which there are also conjunctions thereof. The causal structure of the former theories will always be trivial, whereas that of the latter can be nontrivial. And this difference depends only on the presence of conjunctions and not of disjunctions in the operational theory.

It is worths reflecting for a moment on the difference between trivial and non-trivial theories from a general Bridgmanian perspective. If we trivialize a theory, we change the empirical content. The new basic measurements will not be the same as the old maximally simultaneous measurements and the old basic measurements will not be the same of the marginalization of the new basic measurements. As a special consequence of this general fact, measurements sitting in two different contexts in the non-trivial theory (as A_0 in our above example) will "multiply realized" in the trivial theory by two different marginalized measurements (by $C_{00}^{(1)}$ and $C_{01}^{(1)}$ in our example). This results in the disappearing of those conditional independences in which the original measurement was featuring and consequently in the a radical change of the causal structure based on these conditional independences. From a general point of view, this is understandable: operational theories with different empirical content can have different causal explanation. A causal explanation relies not only on the outcome statistics of the measurements but also on their compatibility structure. Focusing only on operationally equivalent measurement classes, this information about simultaneous measurability gets lost.

9 Quantum mechanics

Quantum mechanics, at least in the minimalist interpretation, is an operational theory in a special linear algebraic representation. Therefore, it is instructive to see how quantum mechanics represents the EPR-Bell scenario and how this representation relates to the Bridgmanian and the standard identification of observables. The probabilities of the both the non-trivial operational theory (13) and (15) and the trivial operational theory (23) are generated quantum mechanically as follows:

$$\langle \Psi_s | (\mathbf{X^A} \otimes \mathbf{I}) \Psi_s \rangle = p(X|A, P_{\text{EPR}}) \tag{40}$$

$$\langle \Psi_s | (\mathbf{I} \otimes \mathbf{Y^B}) \Psi_s \rangle = p(Y|B, P_{\text{EPR}}) \tag{41}$$

$$\langle \Psi_s | (\mathbf{X^A} \otimes \mathbf{Y^B}) \Psi_s \rangle = p(X, Y|A, B, P_{\text{EPR}}) \tag{42}$$

where $|\Psi_s\rangle$ is the singlet state representing the preparation P_{EPR} in the Hilbert space $H_2 \otimes H_2$; $\mathbf{I}$ is the unit operator in H_2; and $\mathbf{X^A}$ and $\mathbf{Y^B}$ scroll over eight projections

$$\mathbf{X_0}^{\mathbf{A_0}}, \mathbf{X_1}^{\mathbf{A_0}}, \mathbf{X_0}^{\mathbf{A_1}}, \mathbf{X_1}^{\mathbf{A_1}}$$
$$\mathbf{Y_0}^{\mathbf{B_0}}, \mathbf{Y_1}^{\mathbf{B_0}}, \mathbf{Y_0}^{\mathbf{B_1}}, \mathbf{Y_1}^{\mathbf{B_1}}$$

corresponding to eight unit vectors $|X^A\rangle$ and $|Y^B\rangle$ in H_2 such that

$$|\langle X^A | Y^B \rangle|^2 = \begin{cases} \frac{3}{4} & \text{if } X \oplus Y = 0 \text{ and } A \cdot B = 0 \\ \frac{1}{4} & \text{if } X \oplus Y = 1 \text{ and } A \cdot B = 0 \\ 1 & \text{if } X \oplus Y = 0 \text{ and } A \cdot B = 1 \\ 0 & \text{if } X \oplus Y = 1 \text{ and } A \cdot B = 1 \end{cases}$$

The operators representing the four measurements are:

$$\mathbf{A_0} = \mathbf{X_0}^{\mathbf{A_0}} - \mathbf{X_1}^{\mathbf{A_0}}, \qquad \mathbf{A_1} = \mathbf{X_0}^{\mathbf{A_1}} - \mathbf{X_1}^{\mathbf{A_1}}$$
$$\mathbf{B_0} = \mathbf{Y_0}^{\mathbf{B_0}} - \mathbf{Y_1}^{\mathbf{B_0}}, \qquad \mathbf{B_1} = \mathbf{Y_0}^{\mathbf{B_1}} - \mathbf{Y_1}^{\mathbf{B_1}}$$

with eigenvalues ± 1.

The operators, however, represent different measurements in the non-trivial and trivial operational theory. Consider, for example, the quantum optical realization of the EPR-Bell scenario. In both operational theories, one prepares an ensemble of photon pairs in singlet state and performs certain polarization measurements on the pairs.

In the *non-trivial* theory, one has four *local* measurements: two linear polarization measurements on the left photon, A_0 and A_1, and two linear polarization measurements on the right photon, B_0 and B_1. These measurements are the following:

A_0 : Measure the linear polarization of the left photon along a given transverse axis a_0 (with outcome $+1$ if the photon passes the polarizer and -1 if not)
A_1 : Measure the linear polarization of the left photon along a transverse axis a_1 at $60°$ from the axis a_0
B_0 : Measure the linear polarization of the right photon along a transverse axis b_0 at $60°$ from the axis both a_0 and a_1
B_1 : Measure the linear polarization of the right photon along the transverse axis $b_1 = a_1$

The polarization measurements on the left subsystem can be simultaneously performed with the polarization measurements on the right subsystem realizing the simultaneous measurements $A_0 \wedge B_0$, $A_0 \wedge B_1$, $A_1 \wedge B_0$, and $A_1 \wedge B_1$. The local measurements do not disturb one another, still the ontological model constructed above is simultaneous contextual: performing measurement A_0 or A_1 causally influences the outcomes of B_0 and B_1. Since the events A and Y are spacelike separated, this is a clear violation of local causality.

In the *trivial* operational theory, we replace the local measurements with *global* measurements. We will have four new measurements, each with four outcomes represented by four orthogonal unit vectors in $H_2 \otimes H_2$:

C_{00} : Perform a global polarization measurement on the photon pair with four outcomes corresponding to the basis $\left\{ |X_0^{A_0}\rangle \otimes |Y_0^{B_0}\rangle, |X_0^{A_0}\rangle \otimes |Y_1^{B_0}\rangle, |X_1^{A_0}\rangle \otimes |Y_0^{B_0}\rangle, |X_1^{A_0}\rangle \otimes |Y_1^{B_0}\rangle \right\}$
C_{01} : Perform a global polarization measurement on the photon pair with four outcomes corresponding to the basis $\left\{ |X_0^{A_0}\rangle \otimes |Y_0^{B_1}\rangle, |X_0^{A_0}\rangle \otimes |Y_1^{B_1}\rangle, |X_1^{A_0}\rangle \otimes |Y_0^{B_1}\rangle, |X_1^{A_0}\rangle \otimes |Y_1^{B_1}\rangle \right\}$
C_{10} : Perform a global polarization measurement on the photon pair with four outcomes corresponding to the basis $\left\{ |X_0^{A_1}\rangle \otimes |Y_0^{B_0}\rangle, |X_0^{A_1}\rangle \otimes |Y_1^{B_0}\rangle, |X_1^{A_1}\rangle \otimes |Y_0^{B_0}\rangle, |X_1^{A_1}\rangle \otimes |Y_1^{B_0}\rangle \right\}$

C_{11} : Perform a global polarization measurement on the photon pair with four outcomes corresponding to the basis $\left\{ \left|X_0^{A_1}\right\rangle \otimes \left|Y_0^{B_1}\right\rangle, \left|X_0^{A_1}\right\rangle \otimes \left|Y_1^{B_1}\right\rangle, \left|X_1^{A_1}\right\rangle \otimes \left|Y_0^{B_1}\right\rangle, \left|X_1^{A_1}\right\rangle \otimes \left|Y_1^{B_1}\right\rangle \right\}$

Note that these global polarization measurements are realized by a complicated arrangement of beam splitters, polarizing beam splitters, wave plates, photo detectors and other non-linear optical devices (Mattle et al., 1996; Lütkenhaus et al., 1999; Weihs and Zeilinger 2001). What is important, is that C_{00} is *not* simply performing a linear polarization measurement on the left photon along axis a_0 and performing a linear polarization measurement on the right photon along a given transverse axis b_0. In other words, C_{00} is *not* the same measurement as $A_0 \wedge B_0$; they are only operationally equivalent. Consequently, $C_{00}^{(1)}$ will not be the same as A_0; they will be only operationally equivalent.

This new operational theory has a trivial compatibility structure: C_{01} and C_{11} cannot be performed simultaneously, that is, they cannot be performed on the same pair of photons. Consequently, any ontological model for the theory is (trivially) simultaneously noncontextual. But the model we provided will be measurement contextual: some ontic states will provide different outcomes for the $C_{01}^{(2)}$ and $C_{11}^{(2)}$ contrary to the fact that they are operationally equivalent. Note, however, that measurement contextuality does not lead to the violation of local causality.

In the Introduction, we discerned the Bridgmanian and the standard identification of observables. In the first case, we identified observables with operators, in the second, with measurements. Applying this distinction to the EPR-Bell scenario, one gets the following schema:

$$
\begin{array}{rccc}
 & \textit{Bridgmanian} & & \textit{Standard} \\
\textit{Operator:} & \mathbf{A_0} & & \mathbf{A_0} \\
 & \swarrow \quad \searrow & & \downarrow \\
\textit{Observable:} & \mathcal{O}_1 \qquad \mathcal{O}_2 & & \mathcal{O} \\
 & \downarrow \qquad \downarrow & & \swarrow \quad \searrow \\
\textit{Measurement:} & A_0 \;\sim\; C_{00}^{(2)} & & A_0 \;\sim\; C_{00}^{(2)}
\end{array}
$$

The local and global measurements are represented by the same operator in quantum mechanics. But do they measure the same observable? According to the standard approach: yes; according to the Brigdmannian approach: no.

10 Conclusions

Operational quantum mechanics is a special operational theory in a linear algebraic representation. A distinctive feature of this theory is operational equivalence, the representation of different (and not necessarily simultaneously performable) measurements providing the same outcome statistics in every quantum state by the same self-adjoint operator (or POVM). From the perspective of strict operationalism, the identity of the representation of such measurements does not mean the identity of the measured observables. In this paper, I intended to explore some of the consequences of this Bridgmanian perspective in quantum theory and in general operational theories. We saw, how certain essential properties of the underlying ontological models changed if some measurements were replaced by other operationally equivalent measurements. This change is a straightforward consequence of the sensitivity of the ontological models to the compatibility structure of the theory. To illustrate this change in quantum mechanics, I took the example of the EPR-Bell scenario and compared the ontological models of the non-trivial and the trivial operational theories realizing the EPR-Bell scenario by local and global measurements, respectively. The EPR-Bell situation, however, was not peculiar whatsoever; we could have equally well used the GHZ or the Peres-Mermin case to this goal. The four commuting operators in the horizontal line of the GHZ pentagram

$$\boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z \qquad \boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x \qquad \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_x \qquad \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_z$$

or the three commuting operators in the third column of the Peres-Mermin square

$$\boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z \qquad \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y \qquad \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x$$

can also be represented both by local measurements on individual photons (represented by the graphs in the Figure 1) and also by complicated global GHZ or Bell state measurements on photon pairs or triples (represented by the linegraphs in the Figure 3). These local and global measurements are different and so are the ontological models. All ontological models will be measurement contextual, but those for global measurements will be simultaneously noncontextual and will have a trivial causal structure (Hofer-Szabó, 2021a, b, 2022). All these results point in the same direction which is also the main message of this paper: Operationally equivalent families of measurements represented by the same operators in quantum mechanics can give rise to ontological models with highly different features. Thus, to study these models, it is not enough to simply investigate quantum mechanics at an abstract mathematical level; we also need to take into consideration the measurements represented by the operators. This is the lesson that we can learn from Bridgman.

Appendix

Three outcome-deterministic ontological model for the three operational theories with non-trivial compatibility structure:

Classical theory.

- Set of ontic states: $\mathcal{L} = \{\Lambda_0, \Lambda_1\}$
- Random variable on $\mathcal{L} : \Lambda = 0, 1$
- Probability distribution:

$$p(\Lambda|P_{\text{CL}}) = \frac{1}{2} \tag{43}$$

- Response functions of the non-trivial theory:

$$p(X|A,\Lambda) = \delta_{X,\Lambda} \tag{44}$$
$$p(Y|B,\Lambda) = \delta_{Y,\Lambda} \tag{45}$$
$$p(X,Y|A,B,\Lambda) = \delta_{X,\Lambda} \cdot \delta_{Y,\Lambda} \tag{46}$$

  where δ is the Kronecker delta function.
- Response functions of the trivial theory:

$$p(Z|C,\Lambda) = \delta_{Z_1,\Lambda} \cdot \delta_{Z_2,\Lambda} \tag{47}$$

The EPR-Bell scenario.

- Set of ontic states: $\mathcal{L} \times \mathcal{L}$ where $\mathcal{L} = \{\Lambda_0, \Lambda_1\}$
- Random variable on $\mathcal{L} \times \mathcal{L} : \Lambda_1 \times \Lambda_2$ with $\Lambda_1, \Lambda_2 = 0, 1$
- Probability distribution:

$$p(\Lambda_1, \Lambda_2|P_{\text{EPR}}) = \begin{cases} \frac{1}{8} & \text{if } \Lambda_1 \oplus \Lambda_2 = 1 \\ \frac{3}{8} & \text{otherwise} \end{cases} \tag{48}$$

- Response functions of the non-trivial theory:

$$p(X|A,\Lambda_1,\Lambda_2) = \delta_{X,\Lambda_1} \tag{49}$$
$$p(Y|B,\Lambda_1,\Lambda_2) = \delta_{Y,\Lambda_2} \tag{50}$$
$$p(X,Y|A,B,\Lambda_1,\Lambda_2) = \delta_{X,\Lambda_1} \cdot (\delta_{Y\oplus(A\cdot B),\Lambda_2} \cdot \delta_{\Lambda_1\oplus\Lambda_2,1} + \delta_{Y,\Lambda_2} \cdot \delta_{\Lambda_1\oplus\Lambda_2,0}) \tag{51}$$

- Response functions of the trivial theory:

$$p(Z|C,\Lambda) = \delta_{Z_1,\Lambda_1} \cdot (\delta_{Z_2\oplus(C_1\cdot C_2),\Lambda_2} \cdot \delta_{\Lambda_1\oplus\Lambda_2,1} + \delta_{Z_2,\Lambda_2} \cdot \delta_{\Lambda_1\oplus\Lambda_2,0}) \tag{52}$$

PR box.

- Set of ontic states: $\mathcal{L} = \{\Lambda_0, \Lambda_1\}$
- Random variable on $\mathcal{L} : \Lambda = 0, 1$
- Probability distribution:

$$p(\Lambda|P_{\text{PR}}) = \frac{1}{2} \tag{53}$$

- Response functions of the non-trivial theory:

$$p(X|A,\Lambda) = \delta_{X,\Lambda} \tag{54}$$
$$p(Y|B,\Lambda) = \delta_{Y,\Lambda} \tag{55}$$

$$p(X,Y|A,B,\Lambda) = \delta_{X,\Lambda} \cdot \delta_{Y\oplus(A\cdot B),\Lambda} \tag{56}$$

- Response functions of the trivial theory:

$$p(Z|C,\Lambda) = \delta_{Z_1,\Lambda} \cdot \delta_{Z_2\oplus(C_1\cdot C_2),\Lambda} \tag{57}$$

Acknowledgements. This work has been supported by the Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation, the Hungarian National Research, Development and Innovation Office (K-134275). I would like to thank for the two anonymous reviewers for their valuable comments.

References

Abramsky S., and A. Brandenburger, (2011). The sheaf-theoretic structure of non-locality and contextuality, New J. Phys, 13, 113036.

Abramsky, S., R. S. Barbosa, K. Kishida, R. Lal, S. Mansfield, (2017). Contextuality, cohomology and paradox, URL = https://arxiv.org/abs/1502.03097.

Bridgman, P. W. (1958). The Logic of Modern Physics (New York: The Macmillan Company).

Busch P., Lahti P., Pellonpää J-P., Ylinen K. (2016). Quantum Measurement. Springer.

Cavalcanti, E. (2018). Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning, Phys. Rev. X, 8, 021018.

Chang, H. (2019). Operationalism, Stanford Encyclopedia of Philosophy. URL = https://plato.stanford.edu/entries/ operationalism.

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

Glymour, C., Scheines, R., and Spirtes, P. (2001). Causation, Prediction, and Search, (Cambridge: The MIT Press).

Greenberger, D. M., Horne, M. A., Shimony, A. and Zeilinger, A. (1990). Bell?s theorem without inequalities, Am. J. Phys. 58, 1131

D'Ariano G. M., Chiribella G., Perinotti P. (2017). Quantum Theory from First Principles: An Informational Approach, Cambridge: Cambridge University Press.

Held, C. (2022). The Kochen-Specker Theorem, Stanford Encyclopedia of Philosophy

Kochen, S., and E. P. Specker (1967). The problem of hidden variables in quantum mechanics, J. Math. Mech., 17, 59–87.

Lin, Y. (2021). Conspiracy in ontological models: λ sufficiency and measurement contextuality, Phys. Rev. A 103, 022211.

Ludwig, G. (1983). Foundations of Quantum Mechanics. Springer.

Lütkenhaus, N., Calsamiglia, J., and Suominen, K-A. (1999). On Bell measurements for teleportation, Phys. Rev. A, 59, 3295.

Mackey G. W. (1957). Quantum Mechanics and Hilbert Space. The American Mathematical Monthly, 64, 45-57.

Mattle, K., Weinfurter, H., Kwiat, P. G., and Zeilinger A. (1996). Dense Coding in Experimental Quantum Communication, Phys. Rev. Lett, 76 (25), 4656-4659.

Mermin, D. (1993). Ontological states and the two theorems of John Bell, Rev. Mod. Phys., 65 (3), 803-815.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference, (Cambridge: Cambridge University Press)

Peres, A. (1990). Incompatible Results of Quantum Measurements,? Phys. Lett. A, 151, 107-108.

Popescu, S., and D. Rohrlich. (1994). Nonlocality as an axiom, Found. Phys., 24, 379-385.

Reichenbach, H. (1927). The Philosophy of Space and Time, New York: Dover Publications, 1958.

Schlick, M. (1930 [1979]). On the Foundations of Knowledge, in Philosophical Papers, vol. 2 (1925–1936), H. L. Mulder and B. F. B. van de Velde-Schlick (eds.), Dordrecht: Reidel, pp. 370–387.

Spekkens, R. W. (2005). Contextuality for preparations transformations and unsharp measurements, Phys. Rev. A 71:052108.

Suarez, M. (2007). Causal Inference in Quantum Mechanics: A Reassessment, in: F. Russo and J. Williamson (eds.), Causality and Probability in the Sciences. College Publications.

Suarez, M. and SanPedro, I. (2009). Causal Markov, Robustness and the Quantum Correlations, in: M. Suarez (ed.), Probabilities, Causes and Propensities in Physics, Synthese Library, Springer, Ch. 8., 173-193.

Suppes, P. (1951). A set of independent axioms for extensive quantities, Portugaliae Mathematica, 10(4): 163–172.

Tal, E. (2020). Measurement in Science, Stanford Encyclopedia of Philosophy. URL = https://plato.stanford.edu/entries/measurement-science

Trout, J.D. (1998). Measuring the intentional world: Realism, naturalism, and quantitative methods in the behavioral sciences, Oxford: Oxford University Press.

Weihs, G., and Zeilinger, A. (2001). Photon statistics at beam splitters: an essential tool in quantum information and teleportation, in: Jan Perina (ed), Coherence and Statistics of Photons and Atoms, (John Wiley and Son, Inc, New York) 262-288.

van Fraassen, B.C. (1980). The Scientific Image, Oxford: Clarendon Press.

von Neumann, J. (1932). Mathematische Grundlagen der Quantenmechanik, Springer.
Mathematische Grundlagen der Quantenmechanik

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 J. Phys. 17, 033002.