Causal Graphs for EPR Experiments

Paul M. Näger

Department of Philosophy

University of Bremen, Enrique-Schmidt-Str. 7, D-28359 Bremen, Germany

paul.naeger@uni-bremen.de

26 March 2013

We examine possible causal structures of experiments with entangled quantum objects. Previously, these structures have been obscured (i) by assuming a misleading probabilistic analysis of quantum non locality as 'Outcome Dependence or Parameter Dependence' and (ii) by directly associating these correlations with influences. Here we try to overcome these shortcomings: (i) we proceed from a recent stronger Bell argument (Näger, 2012), which provides an appropriate probabilistic description, and (ii) apply the rigorous methods of causal graph theory. Against the standard view that there is only an influence between the measurement outcomes, we show that there must be an influence from one setting (parameter) to its distant outcome: EPR correlations can only come about if one of the outcomes is a common effect of *both* settings. Our discussion makes explicit under which assumptions similar conclusions from information theoretic considerations (Maudlin 2002, ch. 6; Pawlowski et al. 2010) can be interpreted causally.

Contents

1 Introduction

Modern EPR experiments measure the polarisation of entangled photons (Einstein et al., 1935; Bohm, 1951; Aspect et al., 1982; Weihs et al., 1998). In a schematic setup a source is located midway between two measurement devices. Suitably prepared, the source emits a pair of entangled photons (in the quantum state $\boldsymbol{\psi}$), which move in opposite direction towards the measurement devices. While the photons are on their way, the measurement directions of the devices ('settings', 'parameters') are randomly and independently set to one of two possible values ($\boldsymbol{a} = 1, 2$ and $\boldsymbol{b} = 2, 3$). Finally, each apparatus registers whether the photon is polarised in that direction or not ($\boldsymbol{\alpha} = \pm$ and $\boldsymbol{\beta} = \pm$).[1] The statistics of many repeated measurement runs yields that the outcomes are correlated conditional on the settings and the quantum state, $\neg\, \mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\beta} | \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi})$, the famous EPR correlations.[2]

It is a basic assumption in scientific reasoning that correlations have to be explained. On the one hand, these correlations can be derived from the quantum mechanical formalism by Born's rule,[3] which provides a perfect covering law explanation. On the other hand, it is well-known today that covering law explanations can be dissatisfying in different respects (see, e.g., Salmon, 1990). It has become an approved position in the philosophy of science that explaining a phenomenon in a satisfying sense requires to explain it *causally*, i.e. to tell a story how it comes about (Salmon, 1984, 1998). How do EPR correlations come about? An essential part of such causal explanations is to say which properties of the considered events influence which ('causal structure'). Since in our description of EPR experiments we have used the usual convention to represent properties as values of variables, the question for the causal structure is: *which of the variables of an EPR setup influence which?* This is the question we shall try to answer in this paper.

[1] We shall use the convention to denote variables in bold fonts, e.g. $\boldsymbol{\alpha}$, and corresponding values of variables in normal font, e.g. α; *specific* values of variables are indicated by an index, e.g. α_+.

[2] Assuming a maximally entangled state, e.g. $\psi = \frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle)$, the correlations are given by

$$P(\alpha\beta|ab\psi) = \begin{cases} \frac{1}{2}\cos^2\phi_{ab} & \text{if } \alpha = \beta \\ \frac{1}{2}\sin^2\phi_{ab} & \text{if } \alpha \neq \beta. \end{cases} \tag{1}$$

[3] In the present case, Born's rule reads: $P(\alpha\beta|ab\psi) = |\langle\phi_a, \phi_b|\psi\rangle|^2$.

Tackling this question we shall *not* engage in metaphysical debates about the nature of causation or of quantum non-locality. Rather we understand the question in an *epistemic* sense: according to normal standards of causal inference, which influences does one have to assume in order to *explain* the EPR correlations? When we speak of 'causal relations' and 'causal structures' in the following, we shall always mean 'causal' in an epistemic sense (if not otherwise stated). The epistemic causal structures we shall find leave open how the influences are metaphysically realised—which is a different question and cannot be treated here.[4]

In recent decades, the epistemology of causation has made considerable progress due to causal graph theory, which is summarised in the impressive work of Spirtes et al. (2000) and Pearl (2000). Causal graph theory represents causal structures by diagrams which involve variables as nodes and causal relations as arrows. Causal graphs normally do not involve cycles of arrows; they are directed acyclic graphs. We shall present possible causal structures of EPR experiments in graphs which have the empirically accessible variables $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\psi}$ as nodes (plus a possible hidden variable $\boldsymbol{\lambda}$). Causal graphs are a clear way to provide answers to our question. This is one virtue of causal graph theory.

The main distinction when discussing causal structures for a given correlation between variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (here: the measurement outcomes of EPR experiments) is whether it comes about by direct causation (direct cause structure), or by a common cause (common cause structure) or by both (mixed structure). Since in a typical EPR experiment the measurements are space like related (see fig. 1), only common cause structures can be local; but it is a common place that due to the violation of the Bell inequalities local common cause structures cannot explain the correlations (fig. 2). This is still true even if one accounts for the possibility that there might be hidden variables, a hidden common cause $\boldsymbol{\lambda}$ of the outcomes, which complete the quantum state $\boldsymbol{\psi}$ at the source. In some sense there must be a non-local causal connection, either a direct connection between the outcomes (fig. 3) or a causal relation from one setting to its distant outcome (this setting is a *non-local* common cause of both outcomes; fig. 4), or both (fig. 5). Finally, there are two further special cases of common cause structures. According to indirect structures the settings influence the outcomes only via a hidden common cause $\boldsymbol{\lambda}$ of the outcomes (fig. 6). An interesting subclass of indirect structures are retro-structures: here the hidden common cause $\boldsymbol{\lambda}$ lies in the common past of the settings, so that the influence is backwards in time. Finally, superdeterministic structures assume that the hidden state is not only a common cause of the outcomes but also of

[4]However, the two questions are clearly connected: we believe that the metaphysical question can best be answered only when we have found a suitable epistemic causal structure. Epistemic causal structures, we maintain, are our best guide to metaphysical structure (by inference to the best explanation). Speculating about the metaphysical nature of quantum non-locality, as it has been common in the philosophical literature, without having determined exactly which variables influence another according to epistemic standards, does not seem to be a reliable procedure. So the present paper is also meant to create a sound basis for metaphysical explorations.

the settings (fig. 7).[5]

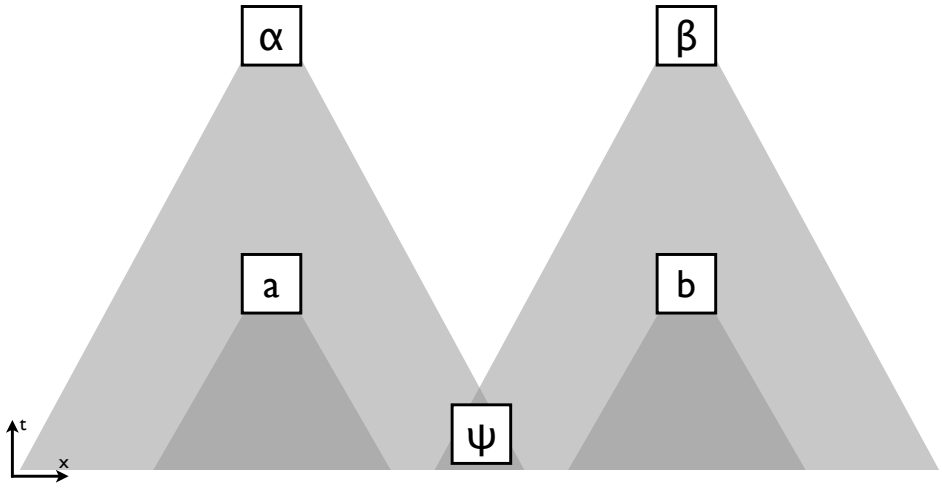

Fig. 1: Schematic space-time diagram of EPR experiments. Each outcome is time-like separated to a local setting ($\boldsymbol{\alpha}$ and $\boldsymbol{a}$, $\boldsymbol{\beta}$ and $\boldsymbol{b}$) and to the quantum state at the source, $\boldsymbol{\psi}$. Any other pair of variables—particularly the outcomes—is space-like separated. Note that we have not marked the location of the hidden common cause $\boldsymbol{\lambda}$—for in principle it can be located anywhere. If it plays the role of hidden variables, however, it must be situated in the common past of the outcomes (as $\boldsymbol{\psi}$).

These six structures are prototypes for causal structures of EPR experiments. Rather than an exhaustive set of possible structures they are the main variants which have been proposed. To each prototype it is easy to find similar structures which fall into the same class, e.g. replace $\boldsymbol{\alpha} \leftarrow \boldsymbol{\beta}$ in figure 23 by $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$. (We shall call such structures which 'mirror' all arrows at the symmetry axis of the experiment 'mirror images'. If an asymmetric structure is allowed or forbidden it is understood that its mirror image, which we shall not explicitly indicate, is allowed or forbidden as well.) One can also have structures which are a mix of two or more prototypes. In our argument below we shall be more precise about these variants. Here, in order to get an overview, we shall constrain our discussion to these prototypes.

Superdeterministic structures have not received much credit, although they constitute the only type of structures which can be completely local (if the hidden variable is in the common past of the settings) *and* violate Bell inequalities. Claiming that the settings which are determined by the experimenter are in effect influenced by a hidden variable they have always provoked the charge of being a conspiratorial theory. We shall not consider these structures as a serious candidate.

[5] Suárez (2004, 2007) and Wood and Spekkens (2012) have provided a similar overview of suggested causal structures for EPR experiments.

Prototypes of possible causal structures for EPR experiments

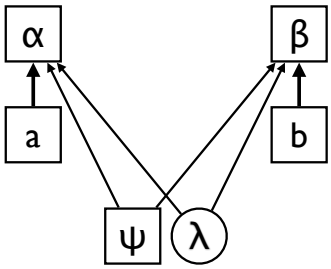

Fig. 2: Local common cause structure

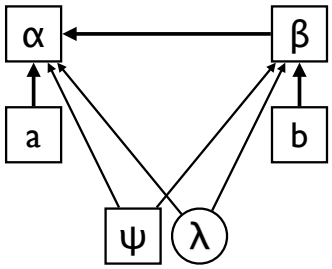

Fig. 3: Direct cause structure

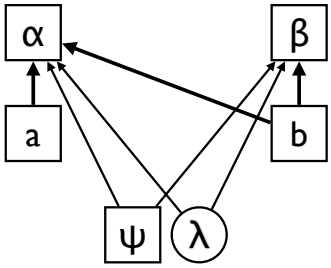

Fig. 4: Non-local common cause structure

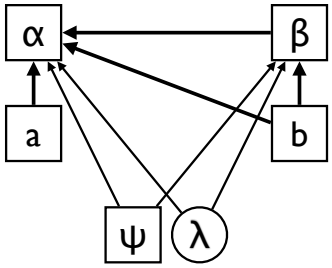

Fig. 5: Mixed structure

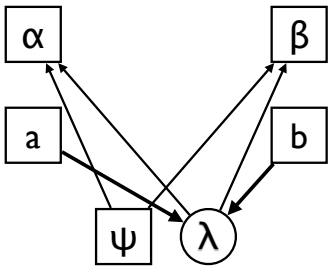

Fig. 6: Indirect structure

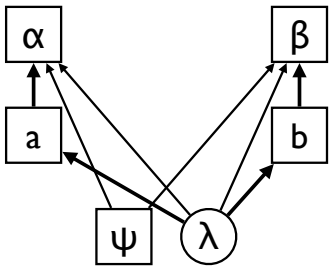

Fig. 7: Superdeterministic structure

(NB: Here and in the following printing some arrows bold is just for reasons of clarity. There is no difference in meaning between bold and normal arrows.)

Among the four remaining structures indirect structures have played only the role of a minority view. In the form of retro-structures they have attracted some attention (Price, 1996). Most philosophers of science interpret the violation of Bell inequalities to show that EPR experiments involve a non-locality of some kind. The non-locality following from Bell's inequality (the failure of the factorisation condition) is on the probabilistic level. Jarrett (1984) famously analysed this quantum non-locality as the disjunction of two probabilistic dependencies, 'Outcome Dependence or Parameter Dependence', suggesting that there is a choice to be made. The predominant criterion for preferring one dependence over the other is the compatibility with relativity: by different arguments (Jarrett, 1984; Shimony, 1984; Arntzenius, 1994) it has been claimed that Parameter Dependence is in conflict with relativity while Outcome Dependence is not. This result, which is still on a probabilistic level, is then interpreted *causally* by associating the dependencies with *influences*: influences between a setting and its distant outcome are inconsistent with relativity, whereas such between the outcomes are not. This rules non-local common cause structures as well as mixed structures out. Rather, according to this standard view, EPR correlations come about by a non-local influence from one outcome to the other (direct cause structure).[6]

In this paper we shall argue that this standard view is deeply flawed. The argument we shall present will have the opposite conclusion: *EPR experiments cannot have a direct cause structure.* For we shall show that direct cause structures imply Bell inequalities, just as local structures do. Thus, for mathematical reasons, direct cause structures cannot explain correlations of the outcomes which are as strong as EPR correlations. Our positive claim will be that *Bell inequalities can only be violated if at least one of the outcomes is a common effect of both settings.* This means that there must be a causal path from at least one of the settings to its distant outcome. This path can be a direct influence (non-local common cause structure), and it might additionally involve an influence between the outcomes (mixed structure), or it is indirect via the hidden variable (indirect structure). In this paper, we shall leave open which of these three possible prototypes likely is the true one. The crucial point here is to show against the standard view that direct structures cannot hold.

The normal way to find causal graphs for a given probability distribution is via causal discovery algorithms (Spirtes et al., 2000; Pearl, 2000). In EPR experiments, however, these algorithms cannot be applied because one of their preconditions, the so called faithfulness assumption, is violated (Glymour, 2006;

[6] We should note that the standard view has it that, *metaphysically*, this influence between the outcomes is grounded in a *non-causal* relation, either a non-causal influence ('passion at-a-distance', Shimony 1984; Redhead 1987) or a holistic 'non-separability' (Howard 1989; Teller 1986; Jarrett 1989; Healey 1991, 1994). There is an ongoing debate whether Lewis' counterfactual analysis of causation judges the outcomes to be causally connected (Butterfield, 1992) or not (Glynn and Kroedel, forthcoming). We stress, however, that in our *epistemic* scheme, which is insensitive to metaphysical realisation, we count any alleged influence as a *causal* relation. For our question here is which variable influences which—not whether a given influence is causal or not. Thus, not any arrow in the causal diagrams we shall consider necessarily is meant to represent a causal relation in a metaphysical sense; the influence it represents might well be due to a different kind of relation.

Wood and Spekkens, 2012). Hence, we have to approach the causal structure in a rather unusual way: we shall give Bell's theorem, which is an argument on the probabilistic level, a causal interpretation.

We shall not use Bell's theorem in its standard form, which says that *any* non-local theory can violate Bell inequalities. For this view has recently been shown to be wrong (Näger, 2012): a stronger version of Bell's theorem makes explicit that certain non-local theories still imply Bell inequalities, among them outcome dependent theories. Jarrett's choice ('Outcome Dependence or Parameter Dependence') is misleading because outcome dependent theories cannot violate Bell inequalities. Proceeding from this false assumption is the main failure of the standard view. We shall overcome this drawback by interpreting the new stronger Bell argument.

Moreover, we shall avoid the standard view's problematic identification of certain probabilistic dependencies with influences. Rather we shall use the clear principles of causal graph theory, which allow to infer statistical properties from causal structures (and, vice versa, to infer causal structures from statistical data). Providing clear translation principles is another virtue of causal graph theory. It is of particular importance because the transition in both directions is susceptible to many fallacies, ranging from naive identifications of statistical dependence with causal dependence (correlation is not causation) up to subtleties about non-faithful representations. A causal interpretation of Bell's theorem by causal graph theory will reveal that also in this case the association of statistical dependencies with influences, such as Outcome Dependence with an influence between the outcomes, is plainly false. This is a second drawback of the standard view. In this sense, our argument shall make explicit the two crucial points at which the argument for the standard view fails.

We should not start before noting that our conclusion has a considerable similarity with certain non-standard views about EPR experiments, which are based on information theoretic considerations. It has been shown that *information* about the distant outcome does not suffice to explain EPR correlations:

> *Bell's inequality can reliably be violated only when the response of one of the particles depends (at least sometimes) on the question asked its partner.* [...] [D]ependence on the distant polarizer setting is crucial. Jarrett's division of theories into those that violate outcome independence and those that violate parameter independence is again seen to be misleading: *any* successful theory must postulate some influence of a distant 'parameter' (i.e. the polarizer angle) on the response of a local photon. Without such dependence the quantum statistics cannot be recovered. (Maudlin, 2002, p. 182f)

> ...it is impossible to model a violation [of the Bell inequalities] without having information in one laboratory about both the setting and the outcome at the distant one. While it is possible that outcome information can be revealed from shared hidden variables, [...] the setting information must be non-locally transferred. (Pawlowski et al., 2010)

Both Maudlin and Pawlowski et al. agree that at least one of the outcomes must depend on the information about the distant setting. Since the information about the local setting is assumed to be available anyway, this amounts to saying that one of the outcomes must depend on the information about *both* settings.

This result in terms of *information* seems very similar to the *causal* claim we shall establish that at least one of the outcomes must be a common effect of both settings. Indeed, Maudlin directly attaches a causal reading to his result ('influence'). (Pawlowski et al. are more cautious, speaking only of information which is 'transferred'.) It is not clear, however, whether Maudlin's leap from 'depending on the information of' to 'is influenced by' is sound. As the transition from statistical to causal facts, that transition might be liable to different fallacies. Under what conditions is our result equivalent to Maudlin's and Pawlowski's? We shall give an answer at the end of this paper. It will turn out that very similar principles which correctly translate probabilistic dependence to causal dependence also translate informational dependence to causal dependence. The result will be that the two approaches—the information theoretic approach by Maudlin and Pawlowski et al. and ours via Bell inequalities—converge: they yield the same causal structures. This will be another strong argument in favour of our claim against the standard view.

The paper is organised in two parts. We first apply causal graph theory to the standard Bell argument. The rough result of this argument is well known (one must have any non-local theory in order to violate Bell inequalities), but it will allow to introduce the causal principles and to demonstrate our concept of causal interpretation at a moderately complicated level. It will also bring home our first point against the standard view that one should not identify singular probabilistic dependencies with influences. In a second part, we shall then apply the same principles to the stronger Bell argument. It will prove our negative main result in opposition to the standard view that direct cause structures cannot violate Bell inequalities. We shall also make precise which positive features causal structures must have in order to be able to account for a violation. Finally, we shall show in what sense our result is in accordance with the information theoretic considerations of Maudlin and Pawlowski et al.

2 A causal interpretation of Bell's theorem

Bell's theorem is a mathematical theorem. In its most general form, it is an argument formulated in probabilistic language. It says that a set of certain probabilistic conditions is in conflict with the empirical results of EPR experiments. Hence, one of the probabilistic conditions must be false. Typically, however, Bell's theorem is said to show certain deep causal or metaphysical facts. How does this work, if the premisses and the conclusion are on a probabilistic level? The answer is that the probabilistic conditions receive a causal or metaphysical *interpretation*. In this section we shall show why the standard way to interpret Bell's theorem, viz. to identify singular probabilistic facts with causal or metaphysical meaning, is wrong. We shall present a reliable way to

endow Bell's theorem with meaning: the clear principles of causal graph theory will provide a sound interpretation. This will also make explicit that Bell's theorem allows not only for the standard reading that local causal structures are ruled out but also for other interpretations.

2.1 What is a causal interpretation?

Let us start by stating Bell's theorem in an explicit form. The probabilistic assumptions underlying the theorem are:

Local Factorisation: $P(\alpha\beta|ab\psi\lambda) = P(\alpha|a\psi\lambda)P(\beta|b\psi\lambda)$

Probabilistic Autonomy: $P(\lambda|ab\psi) = P(\lambda|\psi)$

These two assumptions imply the Bell inequalities (see, e.g., Bell 1975, van Fraassen 1989 or Shimony 1990), which are violated by EPR correlations. Thus, at least one of the probabilistic assumptions that imply the inequality must be false. Here is the argument in an explicit form:

(P1) Autonomy and Local Factorisation imply Bell inequalities: (A) $\wedge$ (LF) $\rightarrow$ (BI)

(P2) Bell inequalities are empirically violated: $\neg$(BI)

(C1) Autonomy or Local Factorisation fails: $\neg$(A) $\vee$ $\neg$(LF) (from P1 & P2, MT)

What does a failure of the probabilistic conditions autonomy and Local Factorisation mean? The usual interpretation is to directly attach a meaning to each probabilistic condition. Local Factorisation, for instance, is said to be a locality condition required by relativity (e.g. Bell, 1975; Jarrett, 1984): it guarantees that there are no influences between space-like separated variables. So its failure is meant to show that there is a non-local influence of some kind. Jarrett's analysis of Local Factorisation seems to paint an even more detailed picture: the failure of Local Factorisation is equivalent to the disjunction of the probabilistic conditions Outcome Dependence and Parameter Dependence. While the former is believed to constitute a (non-causal, in a metaphysical sense) influence between the outcomes the latter is taken to be an influence between a setting and its distant outcome.

These interpretations, however, are highly questionable. For what we have in all these cases is a more or less direct association of singular probabilistic conditions with causal (or metaphysical) claims. Yet, it is well-known that correlation is not causation, and that probabilistic independence does not necessarily mean causal independence. How can we do better?

In the following we shall propose a much more reliable way to assign causal meaning to Bell's theorem. The idea is to add an interpretation premise to the argument:

(P0) The causal assumptions (X) imply autonomy and Local Factorisation:
$(\mathrm{X}) \rightarrow (\mathrm{A}) \wedge (\mathrm{LF})$

Here (X) stands for a set of *causal* assumptions which by the approved principles and methods of causal graph theory imply the probabilistic conditions autonomy and Local Factorisation. Using causal graph theory will put the problematic transition from causal facts to probabilistic facts on a solid basis. Since the overall argument has a modus tollens structure, the new premise will have the effect that not only one of the probabilistic assumptions has to fail, (C1), but also one of the causal assumptions cannot be true:

(C2) (X) fails: $\neg(\mathrm{X})$ (from P0 & C1)

This will give the argument a clear causal conclusion. The task in the following is to find a set of causal assumptions which imply autonomy and Local Factorisation.

2.2 A set of causal assumptions

We start by considering possible causal structures among the five empirical variables which describe an EPR experiment. Without further assumptions one could in principle have a causal relation between any pair of these variables. This does not seem plausible, but why? If we are not willing to accept *any* causal relation, we need principles which restrict causal relations in such EPR experiments.

A first observation might be that in EPR experiments the settings and the quantum state are controlled variables: each is set by an intervention to a certain value. We have furthermore assumed that these interventions occur independently of one another: while the quantum state is prepared identically in every run, each setting is randomly chosen. As a consequence, these variables are statistically independent: $P(ab\psi) = P(a)P(b)P(\psi)$. My suggestion is to assume that variables which fulfil these criteria cannot be effects of other variables in the causal structure. If they were, they would either be overdetermined or they were actually not controlled (contrary to our assumption). Hence, such variables cannot be effects in the given set of variables, they can only be causes—which is the definition of an exogenous variable. These considerations are captured by the following principle:

> **Exogeneity of Controlled Variables (ECV)**: Controlled variables, which are set by independent interventions, are exogenous.

This principle seems so plausible to me that I shall not put it into question throughout this paper. It restricts possible causal structures for EPR experiments considerably, viz. to involve only those relations shown in figure 8. Note that, especially, the principle forbids any causal relation between the controlled variables themselves.

A second constraint comes from relativity theory, the famous locality condition. In a causal formulation it says:

Causal Einstein Locality (CEL): There are no direct causal relations between space-like separated events.

The principle has three possible justifications from relativity theory. One says that according to relativity, causal processes are bound to matter-energy, but matter energy cannot travel faster than light. A second reason is the fact that well-known loop-paradoxes can arise if superluminal signalling were possible. Hence, space-like causal relations which allow for signalling are definitely forbidden. A third reason might be that the temporal asymmetry of causation (causes precede their effects) must be true in all frames of reference. In a relativistic space-time, however, this feature would fail. For if two space-like separated events stand in a causal relation, A causes B, there are always frames of reference in which B is prior to A. So Einstein locality seems to be a well-founded principle as well. It forbids any direct causal relation between the two wings of the experiment. The maximal set of causal relations we can still have, if we assume these two restricting principles, is shown in figure 9: a local common cause structure.

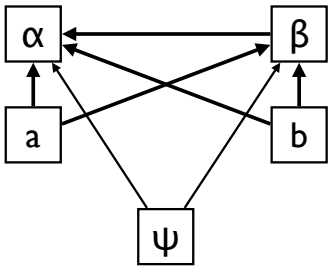

Fig. 8: Possible causal relations if ECV holds. (or $\boldsymbol{\alpha} \leftarrow \boldsymbol{\beta}$ instead of $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$)

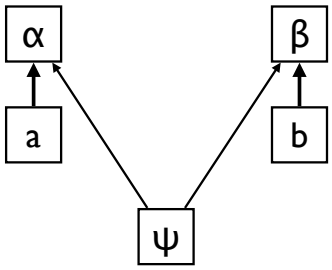

Fig. 9: Possible causal relations if ECV and CEL hold.

We should note, however, that one can circumvent the restrictions that the locality principle imposes, without violating it, if one allowed for backwards causation, i.e. if one allowed that for a pair of time-like separated events A and B the cause A occurs *later* than the effect B. Then we could have an influence, say, from $\boldsymbol{b}$ to $\boldsymbol{\alpha}$, mediated by a variable $\boldsymbol{\mu}$ in their common past, as shown in figure 10. Such backwards causation, however, is highly questionable. Some authors argue that backwards causation is a logical impossibility: the causal asymmetry is conceptually related to the time asymmetry, i.e. causes are always temporarily prior to their effects. Second, even if backwards causation were conceptually possible, it might yield inconsistencies because it enables the coming about of causal loops. Hence, it is plausible to assume:

No Backwards Causation (NBC): Effects cannot precede their time-like separated causes.

With this additional assumption, we now can be sure that no variable on one

wing of the experiment is an effect of a variable on the other. We must have the local common cause structure represented in figure 9.

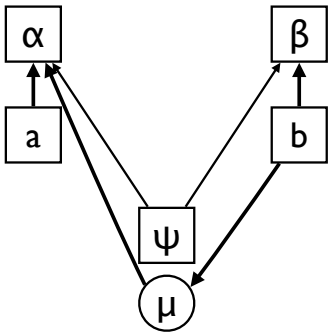

Fig. 10: Causal structure with backwards causation via a latent intervening variable $\boldsymbol{\mu}$.

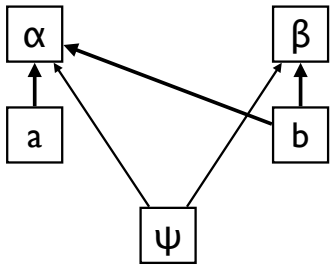

Fig. 11: The structure in fig. 10 omitting the latent intervening variable $\boldsymbol{\mu}$.

So far for principles which restrict possible causal relations. In order to draw mathematical conclusions from causal structures, we still need a principle which translates causal structures to probabilistic facts. The most common principle for this task is the Causal Markov Condition (Spirtes et al. 2000, p. 29f; Pearl 2000, p. 19). It derives probabilistic independence claims from causal structures and is used in nearly all methods of causal inference. It says:

> **Causal Markov Condition (CMC)**: A variable $\boldsymbol{A}$ in a given causal structure is probabilistically independent of its non-effects $\boldsymbol{B}$ conditional on its direct causes $\boldsymbol{C}$:
>
> $$\mathrm{I}(\boldsymbol{A}, \boldsymbol{B}|\boldsymbol{C}) :\leftrightarrow \forall A, B, C : P(A|BC) = P(A|C) \qquad (2)$$

Note that in general $\boldsymbol{B}$ and $\boldsymbol{C}$ are *sets* of variables, $\boldsymbol{B} := \{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n\}$ and $\boldsymbol{C} := \{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_m\}$.

Applying the Causal Markov Condition to the local causal structure would yield the corresponding probabilistic independencies. For this derivation to be correct, however, we have to make sure that we have not forgotten any common cause in our set of variables, because omitting common causes yields independencies which, in fact, do not hold. Given ECV, a hidden common cause $\boldsymbol{\lambda}$ might only be a common cause of the outcomes (fig. 12). Due to the locality condition it would have to be be located in the common past of the outcomes. These restrictions fit very well with the idea of hidden variables, which have been a speculation since the early days of quantum mechanics (Einstein et al., 1935; Schrödinger, 1935). Hidden variables have been thought to provide a more detailed description of the quantum state (here: the photon state at the source), endowing each of the entangled photons with well-defined properties and removing the indeterminacy in their behaviour.

Thus, we allow for one latent common cause $\boldsymbol{\lambda}$, which possibly accounts for hidden variables of the photons at the source. We stress, however, that

by definition $\boldsymbol{\lambda}$ essentially just is a latent common cause of the outcomes: it neither does have to describe hidden variables nor does it have to be located in the common past of the outcomes when we give up on the locality assumption below. We assume that there are no hidden common causes apart from $\boldsymbol{\lambda}$. Having included all true common causes is usually called 'Causal Sufficiency' in the literature on causal graphs (Spirtes et al., 2000, p. 22):

> **Causal Sufficiency$_1$ (CS$_1$)**: The variable set $\boldsymbol{V_1} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}$ contains all common causes of any two (or more) variables in $\boldsymbol{V_1}$.

(Here the index in '$\boldsymbol{V_1}$' indicates the number of hidden common causes in the set, and is referred to in 'Causal Sufficiency$_1$'.)

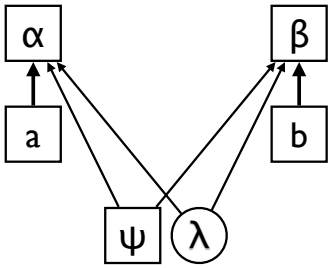

Fig. 12: Local causal structure with hidden variable $\boldsymbol{\lambda}$

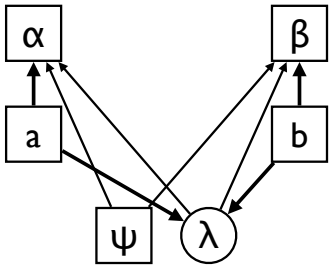

Fig. 13: Non-local causal structure with the hidden variable as common effect of the settings.

Note that $\boldsymbol{\lambda}$ is neither known nor controlled. Hence, ECV does not secure that the hidden variable is exogenous. It would be perfectly consistent with ECV that they are acted upon by other variables. In the present set of assumptions, however, it is the locality assumption which excludes that, possibly, the settings act non-locally on the hidden variables (figure 13), while No Backwards Causation forbids that there is a directed path from one outcome to the other via the hidden variable in their common past. (Later, when we shall give up on these assumptions, we will come back to this fact, that the hidden variable can be acted upon.)

We are now ready to derive probabilistic consequences from the five causal assumptions we have just introduced. Applying the Causal Markov Condition to the local common cause structure in figure 9 yields the following probabilistic

independencies:

$$\mathrm{I}(\boldsymbol{a}, \{\boldsymbol{\psi}, \boldsymbol{\lambda}, \boldsymbol{b}, \boldsymbol{\beta}\}) \tag{3}$$
$$\mathrm{I}(\boldsymbol{b}, \{\boldsymbol{\psi}, \boldsymbol{\lambda}, \boldsymbol{a}, \boldsymbol{\alpha}\}) \tag{4}$$
$$\mathrm{I}(\psi, \{\boldsymbol{\lambda}, \boldsymbol{a}, \boldsymbol{b}\}) \tag{5}$$
$$\mathrm{I}(\boldsymbol{\lambda}, \{\boldsymbol{\psi}, \boldsymbol{a}, \boldsymbol{b}\}) \tag{6}$$
$$\mathrm{I}(\boldsymbol{\alpha}, \{\boldsymbol{b}, \boldsymbol{\beta}\}|\{\boldsymbol{a}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}) \tag{7}$$
$$\mathrm{I}(\boldsymbol{\beta}, \{\boldsymbol{a}, \boldsymbol{\alpha}\}|\{\boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}) \tag{8}$$

From (6)–(8) we can now very easily derive autonomy and Local Factorisation, the two probabilistic assumptions which are needed for Bell's theorem.[7]

So we have found a set (X) of causal assumptions which imply the probabilistic conditions entailing Bell inequalities. This makes the interpretation premise (P0), that we have added to Bell's argument, explicit. The five causal assumptions are the causal interpretation of Bell's theorem we have been looking for. The complete argument, including the causal interpretation, reads:

(P0) Exogeneity of Controlled Variables, the Causal Markov Condition, Causal Sufficiency$_1$, Causal Einstein Locality and No Backwards Causation imply Probabilistic Autonomy and Local Factorisation:
(ECV) $\wedge$ (CMC) $\wedge$ (CS$_1$) $\wedge$ (CEL) $\wedge$ (NBC) $\rightarrow$ (A) $\wedge$ (LF)

(P1) Probabilistic Autonomy and Local Factorisation imply Bell inequalities:
(A) $\wedge$ (LF) $\rightarrow$ (BI)

(P2) Bell inequalities are empirically violated: $\neg$(BI)

(C1) Probabilistic Autonomy or Local Factorisation fails: $\neg$(A) $\vee$ $\neg$(LF)
(from P1 & P2, MT)

(C2) At least one of the following principles fails: Exogeneity of Controlled Variables or the Causal Markov Condition or Causal Sufficiency$_1$ or Causal Einstein Locality or No Backwards Causation:
$\neg$(ECV) $\vee$ $\neg$(CMC) $\vee$ $\neg$(CS$_1$) $\vee$ $\neg$(CEL) $\vee$ $\neg$(NBC)
(from P0 & C1, MT)

[7] The derivation uses two of the semi-graphoid axioms for conditional independence relations (see Pearl, 2000, p. 11):

$$(8) \wedge \text{`decomposition axiom'} \Rightarrow \mathrm{I}(\boldsymbol{\beta}, \boldsymbol{a}|\{\boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}) \tag{9}$$
$$(6) \wedge \text{`weak union axiom'} \Rightarrow \mathrm{I}(\boldsymbol{\lambda}, \{\boldsymbol{a}, \boldsymbol{b}\}|\boldsymbol{\psi}) \tag{10}$$

(10) just is autonomy, and (7) and (9) imply Local Factorisation:

$$P(\alpha\beta|ab\psi\lambda) \overset{\text{product rule}}{=} P(\alpha|\beta ab\psi\lambda)P(\beta|ab\psi\lambda) \tag{11}$$
$$\overset{(7),(9)}{=} P(\alpha|a\psi\lambda)P(\beta|b\psi\lambda) \tag{12}$$

In this formulation, Bell's theorem involves two levels: a probabilistic and a causal one. The sub-argument from (P1) and (P2) to (C1) is the usual Bell argument on the probabilistic level. Here we have complemented it by a causal interpretation (P0) which yields a causal conclusion (C2).

2.3 Disentangling causal structure from spatio-temporal realisation

The causal result (C2) shows that the standard conclusion from Bell's theorem that there is a non-local influence, ¬(CEL), is not without alternative. In fact, it is just one in five different possible conclusions. The Bell argument might have different conclusions and our arguments explicitly presents the possible options.

There are two problems. First, which of the five causal assumptions is the one that fails? As the Bell argument in itself does not tell, one would have to invoke further arguments (or criteria). For instance, if one opts for the non-locality horn, one has to have good reasons why the other assumptions hold. However, we shall not discuss at this point which of the assumptions fails. The reason is that this will be easier to treat given the conclusion of the stronger Bell argument that we shall present in the following section. For in the stronger form some of the assumptions, that have been necessary here, turn out to be dispensable (so there will be fewer options).

The second problem is that even if knew that it is locality which fails this would leave us with several options for our question concerning the causal structure: either there could be a non-local influence between the outcomes, or between a setting and its distant outcome, or between a setting and the hidden variable. Thus, a failure of this assumption is not really precise in terms of which variable influences which. The same is true for failures of No Backwards Causation: it would leave us with possible influences between the same pairs of variables—just that they are not direct and non-local but indirect and zig zag in time.

So a failure of these two assumptions allows influences between the same pairs of variables but just differs in how these are spatio-temporal realised. The spatio-temporal features, however, are neither crucial for the derivation nor for our present question (we have said that we are interested in the causal structure not in its metaphysical realisation). In our derivation Causal Einstein Locality and No Backwards Causation have served as physical (or metaphysical) justifications for excluding influences between space-like separated variables. But the argument does *not essentially* rely on this spatio-temporal justification: the argument is formal and the *formal* function of the spatio-temporal principles is just to exclude directed causal paths from one wing of the experiment to the other. Crucuial for the derivation (and for our question) is just that some variables are not causes of some others, so that certain probabilistic independencies (Local Factorisation and Probabilistic Autonomy) hold. So it is clear that we can as well derive the Bell inequalities from principles which directly state which variables cannot influence another:

Causal Outcome Independence (COI): None of the outcomes

is a direct effect of the respectively other outcome relative to $\boldsymbol{V}_1$.

Causal Parameter Independence (CPI): None of the outcomes is a direct effect of its distant setting relative to $\boldsymbol{V}_1$.

Causal Autonomy (CA): None of the settings directly influences the hidden variable $\boldsymbol{\lambda}$ relative to $\boldsymbol{V}_1$.

($\boldsymbol{V}_1$ is our usual set of variables, $\boldsymbol{V_1} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}$.) In the context of EPR experiments, these three principles are extensionally equivalent to Causal Einstein Locality and No Backwards Causation, since they impose the same restrictions to causal structures: given these three principles plus ECV, only local structures are possible. Hence, Bell inequalities follow if we add the CMC and Causal Sufficiency$_1$. This allows to replace Causal Einstein Locality and No Backwards Causation in premise (P0) of the causal Bell argument by the new principles:

(P0′) Exogeneity of Controlled Variables, the Causal Markov Condition, Causal Sufficiency$_1$, Causal Outcome Independence, Causal Parameter Independence and Causal Autonomy imply autonomy and Local Factorisation:
$(\text{ECV}) \wedge (\text{CMC}) \wedge (\text{CS}_1) \wedge (\text{COI}) \wedge (\text{CPI}) \wedge (\text{CA}) \rightarrow (\text{A}) \wedge (\text{LF})$

While the other premisses (P1) and (P2) remain unaltered, the new premise (P0′) entails a new conclusion:

(C2′) At least one of the following principles fails: the Exogeneity of Controlled Variables or the Causal Markov Condition or Causal Sufficiency$_1$ or Causal Outcome Independence or Causal Parameter Independence or Causal Autonomy:
$\neg(\text{ECV}) \vee \neg(\text{CMC}) \vee \neg(\text{CS}_1) \vee \neg(\text{COI}) \vee \neg(\text{CPI}) \vee \neg(\text{CA})$

The difference to the former conclusion is that in (C2) the metaphysical principles Causal Einstein Locality and No Backwards Causation were at stake, while here it is influences between certain variables (irrespective of their spatio-temporal realisation). Thus, we have *separated the spatio-temporal question from the question of the causal structure*: in the new form we can discuss which variable influences which *without* considering how these paths are spatio-temporally realised. It is important not to mix these two questions. In the following we shall only consider the former: we shall make explicit possible causal structures but we shall not care about how causal relations are embedded in space-time. Especially, we explicitly allow for causal relations between space-like separated variables. If we find that there is an influence between space-like separated variables, e.g. between a setting and its distant outcome, we do not discuss whether it is due to a non-local influence or an influence backwards and forwards in time. This is an important but separate question, which we can and do leave open here. Note that neglecting the spatio-temporal question also means omitting intervening variables backwards in time, such as $\boldsymbol{\mu}$ in figure 10.

This structure will now be depicted as shown in figure 11: the causal path $\boldsymbol{b} \to \boldsymbol{\mu} \to \boldsymbol{\beta}$ is replaced by a direct influence $\boldsymbol{b} \to \boldsymbol{\beta}$.

Having separated the question of spatio-temporal realisation from that of the causal structure we nevertheless continue to speak of local and non-local causal relations and structures. A local (non-local) causal relation is just one between time-like (space-like) separated variables, e.g. $\boldsymbol{a} \to \boldsymbol{\alpha}$ ($\boldsymbol{a} \to \boldsymbol{\beta}$). A local (non-local) structure is one which contains only local (some non-local) causal relations. But by qualifying a relation as, for instance, non-local we do not mean to say anything about how the causal relation is in fact embedded in space and time. The non-local relation might be embedded non-locally or by backwards and forwards causation.

2.4 Suggested causal structures

So if ECV, the CMC and Causal Sufficiency$_1$ hold, we either have an influence between the outcomes, or between a setting and its distant parameter, or between s setting and the hidden variable (fig. 14–16). These are information about single causal relations—but how do the complete structures look like in these cases? Bell's argument does not tell. It just requires that there must be one of these relations, but does not describe the rest of the structure. Since violating Bell inequalities requires to have *enough* causal dependencies, a maximal structure consistent with ECV and CS$_1$ (fig. 17) will certainly suffice. But what about weaker structures? Which arrows can we delete such that the structure still allows a violation of Bell inequalities?

Actually, the most interesting cases are *minimal* structures, which may not lose any arrow without losing their ability to violate Bell inequalities. We conjecture that they look like the prototypes in figures 3, 4 and 6. But how do we know? Are there alternatives? Which further argument besides Bell's determines how these structures look like? The answer to the latter question is that structures with too few arrows, like those in fig. 14–16, by the CMC yield independencies which are in conflict with the empirical probability distribution. For instance, all three structures would yield $\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\psi})$, which is empirically wrong.[8] This conflict is direct because it only requires comparing independencies following from the structures with the empirical distribution; no Bell inequality or other sophisticated moves are involved.

Therefore, we need further conditions, which guarantee consistency with the empirical distribution (in the just described sense that the CMC does not yield independencies which are empirically wrong). These conditions are:

(a) There is a directed causal path from the quantum state to each outcome ($\boldsymbol{\psi} \to \ldots \to \boldsymbol{\alpha}$ and $\boldsymbol{\psi} \to \ldots \to \boldsymbol{\beta}$).

[8] It might seem unfamiliar that empirically the dependence $\neg\,\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\psi})$ holds, because in experiments with maximally entangled quantum objects, $\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\psi})$ is the case. However, the slightest deviation from *maximal* entanglement yields the dependence; certain independencies of maximally entangled objects are just due to special degeneracies / symmetries of maximal entanglement. Hence, here we presume a statistics of *non*-maximal entangled objects.

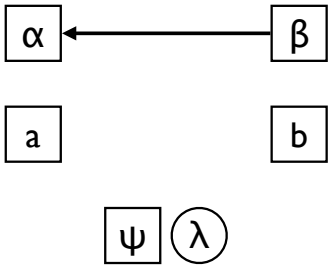

Fig. 14: Violation of Causal Outcome Independence

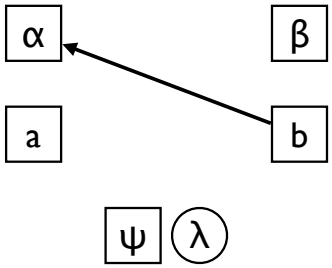

Fig. 15: Violation of Causal Parameter Independence

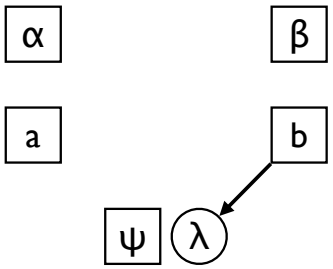

Fig. 16: Violation of Causal Autonomy

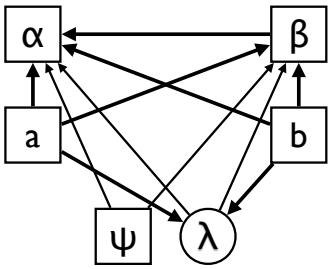

Fig. 17: Maximal structure

(b) There is a directed causal path from each setting to its local outcome ($\boldsymbol{a} \to \ldots \to \boldsymbol{\alpha}$ and $\boldsymbol{b} \to \ldots \to \boldsymbol{\alpha}$).

These principles can be justified by the Causal Markov Condition. However, in this case as well as in others, there is a more appropriate principle which is equivalent to the CMC, but simplifies the reasoning considerably. We shall now introduce this principle, before we come back to the conditions (a) and (b). The principle says (Pearl 2000, p. 16ff; Spirtes et al. 2000, p. 43–46):

> **d-Separation Criterion:** Two variables $\boldsymbol{A}$ and $\boldsymbol{B}$ in a causal graph are probabilistically independent conditional on a set of variables $\boldsymbol{C}$, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are d-separated relative to $\boldsymbol{C}$.

The central concept 'd-separation' is a causal property which indicates how $\boldsymbol{A}$ and $\boldsymbol{B}$ are related in a given causal structure. It is defined as follows:

> **d-separation:** Two variables $\boldsymbol{A}$ and $\boldsymbol{B}$ in a causal graph are d-separated relative to a set of variables $\boldsymbol{C}$, if there is no active causal path between $\boldsymbol{A}$ and $\boldsymbol{B}$ relative to $\boldsymbol{C}$.
>
> **Active path:** A causal path between two variables $\boldsymbol{A}$ and $\boldsymbol{B}$ is active relative to $\boldsymbol{C}$ if all nodes on it are active. A node is active if and only if
>
> – it is a non-collider and not in $\boldsymbol{C}$
>
> *or*
>
> – it is a collider and it (or one of its causes) is in $\boldsymbol{C}$.
>
> **Collider:** A node is a collider on a causal path, if there are two incoming arrows from that path.

The d-separation criterion implies the same independencies as the CMC, but the derivation procedure is much more elegant. For one can directly read off from a given graph whether a pair of variables is implied to be independent given *any* set of conditional variables. In contrast, by the CMC one can only read off independencies *conditional on all direct causes* of one of the variables. All other independencies have to be derived from these former independencies by logical relations.[9] For instance, applied to the local structure, the CMC directly only tells that the outcome $\boldsymbol{\alpha}$ is independent of the setting $\boldsymbol{b}$ conditional on its direct causes $\boldsymbol{a}$, $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$, i.e. $\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{b} | \{\boldsymbol{a}, \boldsymbol{\psi}, \boldsymbol{\lambda}\})$. Without further inferences the CMC does not tell, whether, e.g., the unconditional independence $\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{b})$ holds. The d-separation criterion, however, does: the independence is *not* implied because both causal paths between $\boldsymbol{\alpha}$ and $\boldsymbol{b}$, one via $\boldsymbol{\psi}$ and one via $\boldsymbol{\lambda}$, are active due to the fact that $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ are non-colliders which do not appear in the conditional.

Since the d-separation criterion is equivalent to the CMC (both yield the same independencies) but is much easier to apply, we shall now use it to justify the conditions (a) and (b). More appropriately, we shall use the following equivalent principle:[10]

[9] See the semi-graphoid axioms in (Pearl, 2000, p. 11).

[10] Logically, the causal connection principle is just the contraposition of the d-separation criterion. The name was recently introduced by (Schurz and Gebharter, preprint).

> **Causal Connection Principle (CCP):** Two variables $\boldsymbol{A}$ and $\boldsymbol{B}$ are d-connected relative to a set of variables $\boldsymbol{C}$, if they are probabilistically dependent relative to $\boldsymbol{C}$, $\neg\,\mathrm{I}(\boldsymbol{A}, \boldsymbol{B}|\boldsymbol{C})$.
>
> **d-connection:** Two variables $\boldsymbol{A}$ and $\boldsymbol{B}$ in a causal graph are d-connected relative to a set of variables $\boldsymbol{C}$, if they are not d-separated relative to $\boldsymbol{C}$.

Condition (a) can be derived from the empirical fact that in EPR experiments with different non-maximally entangled states $\boldsymbol{\psi}$ the probabilistic dependencies $\neg\,\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\psi})$ and $\neg\,\mathrm{I}(\boldsymbol{\beta}, \boldsymbol{\psi})$ hold. By the causal connection principle, the former dependence requires that there is an active causal path between $\boldsymbol{\psi}$ and $\boldsymbol{\alpha}$ (otherwise the d-separation criterion or the CMC, respectively, would imply the independence $\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\psi})$). Since the dependence is unconditional, the path may not contain colliders. It is either a directed causal path, $\boldsymbol{\psi} \rightarrow \boldsymbol{\alpha}$ or $\boldsymbol{\psi} \leftarrow \boldsymbol{\alpha}$, or a path with a common cause of $\boldsymbol{\psi}$ and $\boldsymbol{\alpha}$. Since the latter two paths are forbidden by ECV, the former must hold. This path from $\boldsymbol{\psi}$ to $\boldsymbol{\alpha}$, could either be direct, i.e. $\boldsymbol{\psi} \rightarrow \boldsymbol{\alpha}$, or indirect via the hidden variable, i.e. $\boldsymbol{\psi} \rightarrow \boldsymbol{\lambda} \rightarrow \boldsymbol{\alpha}$, or indirect via the other outcome, i.e. $\boldsymbol{\psi} \rightarrow \boldsymbol{\beta} \rightarrow \boldsymbol{\alpha}$. Analogous causal paths from $\boldsymbol{\psi}$ to $\boldsymbol{\beta}$ prevent inconsistency with to the other dependence $\neg\,\mathrm{I}(\boldsymbol{\beta}, \boldsymbol{\psi})$.

The justification for condition (b) is similar to that of (a): in experiments with non-maximally entangled quantum objects, the outcomes depend on their local settings, $\neg\,\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{a})$ and $\neg\,\mathrm{I}(\boldsymbol{\beta}, \boldsymbol{b})$. So, by the CCP, there must be an active causal path between each setting and its local outcome. Due to ECV one can only have a directed causal path from the setting to the outcome, i.e. $\boldsymbol{a} \rightarrow \boldsymbol{\alpha}$ and $\boldsymbol{b} \rightarrow \boldsymbol{\beta}$. This path can be a direct influence, or it can be indirect via the respective other outcome, i.e. $\boldsymbol{a} \rightarrow \boldsymbol{\beta} \rightarrow \boldsymbol{\alpha}$ or $\boldsymbol{b} \rightarrow \boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$, or indirect via the latent common cause $\boldsymbol{\lambda}$, i.e. $\boldsymbol{a} \rightarrow \boldsymbol{\lambda} \rightarrow \boldsymbol{\alpha}$ or $\boldsymbol{b} \rightarrow \boldsymbol{\lambda} \rightarrow \boldsymbol{\beta}$.

So far for the justification and explication of the conditions (a) and (b). Finally, we should note another condition which does not derive from empirical facts: the hidden variable $\boldsymbol{\lambda}$ was *introduced* as a hidden common cause of the outcomes. According to causal graph theory, there is no reason to account for latent variables unless they are common causes. Hence, we also require:

(c) The hidden variable $\boldsymbol{\lambda}$ is a common cause of the outcomes.

So the minimal structures can be found by starting with one of the causal relations necessary for violating Bell inequalities (fig. 14–16) and then adding causal paths which fulfil conditions (a)–(c). We have seen that (a) and (b) can be realised in different ways.[11] Accordingly, there are several minimal structures, which we do not list here all. We focus on the most plausible cases which fulfil (a) by direct influences and (b) either by direct or by indirect influences via the hidden variable. This subset of minimal structures is shown in figures 18 to 21. The first three are the prototypes from figures 3, 4 and

[11] Note that not all possibilities are consistent with another, e.g. $\boldsymbol{a} \rightarrow \boldsymbol{\beta} \rightarrow \boldsymbol{\alpha}$ is inconsistent with $\boldsymbol{\psi} \rightarrow \boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$, because it requires to both have $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$ and $\boldsymbol{\alpha} \leftarrow \boldsymbol{\beta}$. Furthermore, not all structures which are composed in this way are minimal.

6 (figure 21 is a new structure which we have not considered so far). So we have provided a clear argument that these prototypes are *minimal structures* fulfilling both the indirect empirical requirements from the violation of Bell inequalities (if ECV, the CMC and CS_1 hold) as well as the direct empirical constraints (a) and (b) we have considered in this section.

Minimal structures which can violate Bell inequalities
(according to the standard Bell argument)

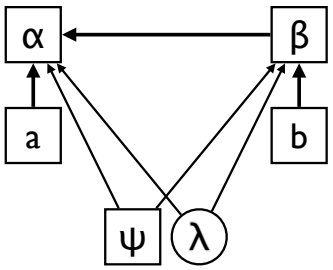

Fig. 18: Minimal structure 1

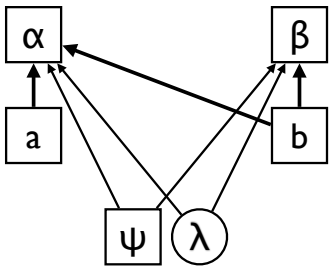

Fig. 19: Minimal structure 2

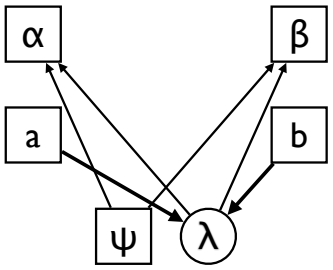

Fig. 20: Minimal structure 3

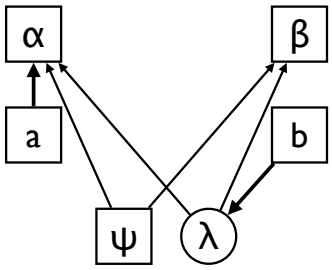

Fig. 21: Minimal structure 4

2.5 Against the standard view 1: a holistic interpretation scheme

We have provided a causal interpretation of the standard Bell argument and have spelled out some of its consequences in terms of causal structures. What is the difference to usual causal interpretations? First, we have consistently used causal graph theory, which is by now not very common in the EPR literature. At every step, the argument is explicit about precisely which principles are at stake. Most of the principles concern features of causal structures which are clearly illustrated by associated graphs. Causal graph theory also points out that besides structural features the translation of structures to probabilistic facts needs special attention: by the CMC (or the d-seperation criterion or the causal connection principle, respectively) it provides a clear connection between causal structures and probabilistic facts.

Second, in the light of our argument the standard interpretation which associates singular probabilistic conditions with a *specific* causal meaning becomes

highly questionable. Consider, for instance, the usual claim that Local Factorisation is a locality condition. It can easily be seen by causal graph theory that this is not true, because Local Factorisation is neither a necessary nor a sufficient condition for Einstein locality. It is not sufficient because there can be non-local causal relations, for instance the relation $a \to \lambda$ in an otherwise local structure (see fig. 22), and the CMC would still imply Local Factorisation. Neither is Local Factorisation necessary for Einstein locality: from a *single* causal claim like Einstein locality nothing follows on a probabilistic level, simply because the translation principle, the CMC, is missing. But even assuming the CMC, Local Factorisation does not follow; the other assumptions are needed as well. Suppose, for instance, Causal Sufficiency$_1$ failed, because, besides $\boldsymbol{\lambda}$, there were another latent common cause of the outcomes. Then, instead of Local Factorisation we had the product form $P(\alpha\beta|ab\psi\lambda) = P(\alpha|\beta ab\psi\lambda)P(\beta|b\psi\lambda)$, i.e. the outcome in the first factor, $\boldsymbol{\alpha}$, would in general still depend on $\boldsymbol{\beta}$ and $\boldsymbol{b}$ due to the forgotten latent common cause.[12] So there really is no correspondence between a *single* causal claim and one of the probabilistic assumptions. Premise (P0) shows that only Einstein locality *and* the CMC *and* No Backwards Causation *and* Causal Sufficiency$_1$ *and* the CMC *jointly imply* that Local Factorisation holds. So the only thing we can say is that, if Local Factorisation fails, one of these causal assumptions cannot be true. But without further assumptions one does not know which.

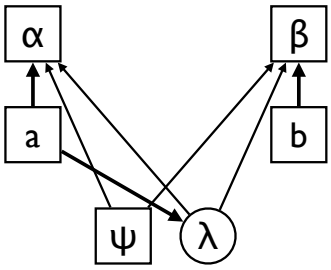

Fig. 22: Non-local structure implying Local Factorisation

This ambiguity remains if one uses Jarrett's analysis (1984) that a failure of Local Factorisation is equivalent to 'Outcome Dependence or Parameter Dependence'. Without further assumptions it is not justified—as it has become usual—to claim that the *probabilistic* condition Outcome Dependence, $\neg\,\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda})$, signifies an *influence* between the outcomes (*Causal* Outcome Dependence) and the other probabilistic condition Parameter Depend-

[12] Since one does not condition on the additional latent common cause, there is an active causal path between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ via $\boldsymbol{\lambda}$ ($\boldsymbol{\lambda}$ is a non-collider and does not appear in the conditional), i.e. the two variables are d-connected. There is also an active path between $\boldsymbol{\alpha}$ and $\boldsymbol{b}$ via $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ ($\boldsymbol{\beta}$ is a collider which appears in the conditional). Hence, the CMC does *not* imply that they are independent. So, except for certain special cases (non-faithful independencies) they are dependent.

ence, $\neg\,\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{b}|\boldsymbol{a}, \boldsymbol{\psi}, \boldsymbol{\lambda}) \vee \neg\,\mathrm{I}(\boldsymbol{\beta}, \boldsymbol{a}|\boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda})$, an influence between at least one setting and its distant outcome (*Causal* Parameter Dependence). Infering Causal Outcome Dependence from probabilistic Outcome Dependence is not valid. It is true, the latter might be due to the former, but it might as well be due to other facts, e.g., another latent common cause which has not been accounted for (i.e. Causal Sufficiency$_1$ fails).

Similarly, probabilistic Parameter Dependence does not imply Causal Parameter Dependence (as in fig. 4). Here, the ambiguity is even more problematic for the standard view. For one of the alternative structures which account for *probabilistic* Parameter Dependence without being *causally* parameter dependent, is the direct cause structure in figure 3: there is no *direct* influence from the setting $\boldsymbol{b}$ to its distant outcome $\boldsymbol{\alpha}$ (i.e. Causal Parameter Independence holds), but due to the indirect path $\boldsymbol{b} \to \boldsymbol{\beta} \to \boldsymbol{\alpha}$ it is nevertheless in general probabilistically parameter dependent.[13] This counterexample is delicate because direct cause structures are commonly regarded as the prototype for structures which are probabilistically outcome dependent but parameter *in*dependent! (According to the standard view, Parameter Independence is important because its violation is said to be incompatible with relativity.) However, the example shows that in general one cannot have Outcome Dependence realised by an influence between the outcomes without Parameter Dependence holding.[14] The standard view's model of quantum non-locality does not seem to be well founded.

The upshot of all this is that the standard way to interpret Bell's theorem, viz. to directly associate probabilistic facts with a specific causal meaning, is flawed. Correlation is not causation. A given probabilistic dependence allows for various different causal interpretations. Especially one should not directly attach a causal meaning to Outcome Dependence or Parameter Dependence. This is a first failure in the argument of the standard view. Rather, we have found a kind of *holistic interpretation scheme*: only a *set* of causal assumptions fixes the meaning of a certain probabilistic condition. So the best one could say about the standard view is that it *implicitly* assumes the other causal assumptions which are required for a sound interpretation. We stress, however, that these assumptions are non-trivial. In this first part of the present paper we have made these assumptions explicit. We have used causal graph theory to introduce an appropriate set, which provides a clear causal interpretation of Bell's theorem.

Our discussion of causally outcome dependent structures which are probabilistically parameter dependent points to still another problem: even if one assumes an appropriate set of causal assumptions, it is problematic to causally interpret a *singular* probabilistic dependence (here: Outcome Dependence). For it might be that the causal assumptions imply other probabilistic facts which are

[13] Parameter independence $\mathrm{I}(\boldsymbol{\alpha}, \boldsymbol{b}|\{\boldsymbol{a}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}) \wedge \mathrm{I}(\boldsymbol{\beta}, \boldsymbol{a}|\{\boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\})$ is not implied by the d-separation criterion because the causal path from $\boldsymbol{b}$ to $\boldsymbol{\alpha}$ via $\boldsymbol{\beta}$ is active.

[14] Jones and Clifton (1993) have made explicit the extra probabilistic conditions under which Outcome Dependence implies Parameter Dependence. They mention that this probabilistic scenario is quite plausibly regarded as a direct cause structure, but they do not make the transition from probabilistic to causal claims as clear as our causal graph theoretic approach does.

not desired or empirically inadequate (here: Parameter Dependence). Causal graph theory tells us that one has to consider *all* dependencies and independencies of given probability distribution. (It is our conditions (a) and (b) which prevent such inconsistencies with the empirical distribution.) Only interpretations of total probability distributions are sound, which is another holistic feature of causal interpretations.

3 Interpreting a stronger Bell argument

Having provided a causal interpretation of the standard Bell argument, we shall now turn to a recent stronger version of Bell's argument (Näger, 2012). The method of interpretation will be very much the same, but the results will differ from those of the standard argument: we shall show that not only local structures but also direct cause structures imply Bell inequalities. This will make the conclusion of the argument considerably stronger.

3.1 A recent stronger version of Bell's theorem

Näger (2012) proves that Bell inequalities can be derived from weaker probabilistic assumptions than autonomy and Local Factorisation. Upholding autonomy he shows that the Bell inequalities do not only follow from *local* product forms of the hidden joint probability $P(\alpha\beta|ab\psi\lambda)$ (esp. Local Factorisation), but also from certain *non-local* ones, which he calls 'weakly non-local'. In contrast to 'strongly non-local' product forms they are characterised by the following principle:

> **No Probabilistic Bell Contextuality (NPBC)**: None of the outcomes in the hidden joint probability depends probabilistically on both settings.

So, surprisingly, we can have a dependence on the distant parameters (in the following three equations we underline variables which indicate an interesting dependence),

$$P(\alpha\beta|ab\psi\lambda) = P(\alpha|\underline{b}\psi\lambda)P(\beta|\underline{a}\psi\lambda), \tag{13}$$

or even a dependence between the outcomes,

$$P(\alpha\beta|ab\psi\lambda) = P(\alpha|\underline{\beta}a\psi\lambda)P(\beta|b\psi\lambda), \tag{14}$$

or both,

$$P(\alpha\beta|ab\psi\lambda) = P(\alpha|\underline{\beta}\underline{b}\psi\lambda)P(\beta|\underline{a}\psi\lambda), \tag{15}$$

and the conclusion still follows.[15] Such weakly non-local distributions imply Bell inequalities as do local ones.

Based on this new derivation Näger has formulated a stronger Bell argument on the probabilistic level. It says:

[15] There are further forms, see table 1 in (Näger, 2012). But (14) and (15) are the strongest ones, which yield the most far reaching results.

(P1′) Probabilistic Autonomy and No Probabilistic Bell Contextuality imply Bell inequalities:
(A) ∧ (NPBC) → (BI)

(P2) Bell inequalities are empirically violated: ¬(BI)

(C1′) Probabilistic Autonomy or No Probabilistic Bell Contextuality fails:
¬(A) ∨ ¬(NPBC) (from P2 & P3, MT)

The conclusion of this argument is stronger than that of the standard Bell argument because the failure of No Probabilistic Bell Contextuality excludes not only local product forms but also weakly non-local ones. In the following we shall give this new argument a causal interpretation. Our hope is that the interpretation of the stronger argument will narrow the range of possible causal structures (and also weakens some of the other causal conditions). We shall follow the same interpretation scheme as above. We formulate an interpretation premise which is an implication from a set of causal assumptions (X′) to those probabilistic conditions which by premise (P2′) imply Bell inequalities:

(P0″) The causal assumptions (X′) imply autonomy and No Probabilistic Bell Contextuality:
(X′) → (A) ∧ (NPBC)

(X′) stands for a conjunction of causal assumptions. Since it only has to imply No Probabilistic Bell Contextuality instead of the stronger Local Factorisation, it is to be expected that it can be made weaker than the set of causal assumptions in its predecessor (P1′). If this were true, the new conclusion of the causal interpretation,

(C2″) (X′) must fail: ¬(X′) (from P0″ & C1′)

would be stronger. We now have to discuss, what the weakest form of (X) exactly is.

3.2 A new causal interpretation

To begin with, let us assume that (X′) includes ECV, the CMC and Causal Sufficiency$_1$. What we are looking for are further principles that restrict causal structures such that the assumptions jointly yield Probabilistic Autonomy and No Probabilistic Bell Contextuality.

Let us first consider autonomy: since autonomy is an unconditional independence of $\boldsymbol{\lambda}$ and both settings, it is clear that it is implied by the d-separation criterion (or, equivalently, by the CMC) if there is no active path from $\boldsymbol{\lambda}$ to any of the settings. What do we have to require that there are no such paths? First, we can ignore paths with colliders because they cannot be unconditionally active. Furthermore, ECV, which says that the outcomes are exogenous variables, excludes a common cause of $\boldsymbol{\lambda}$ and one of the settings as well as a directed causal path from $\boldsymbol{\lambda}$ to one of the settings. So the only remaining possibility which might violate autonomy are directed causal paths from the settings to

the hidden variable $\boldsymbol{\lambda}$. Since $\boldsymbol{\lambda}$ is a common cause of the outcomes, the directed paths cannot lead via the outcomes. Neither can they lead via $\boldsymbol{\psi}$, for $\boldsymbol{\psi}$ is exogenous as well (ECV). Only a direct causal connection from the settings to the outcomes can violate autonomy, $\boldsymbol{a} \rightarrow \boldsymbol{\lambda}$ or $\boldsymbol{\lambda} \leftarrow \boldsymbol{b}$. We exclude such influences by a principle that we have already introduced: Causal Autonomy.

We now turn to the product forms of the hidden joint probability: which causal assumptions do have to hold so that the implied product forms obey No Probabilistic Bell Contextuality? In order to find an appropriate principle, we need a link between causal structures and these product forms. Fortunately, given ECV, the CMC and Causal Sufficiency$_1$, there is a clear correspondence between a causal structure and the product form of the hidden joint probability; given a certain causal structure for the variables in $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}$, the hidden joint probability has the following form:

$$P(\alpha\beta|ab\psi\lambda) = P(\alpha|\{\text{direct causes of } \alpha\})P(\beta|\{\text{direct causes of } \beta\}) \qquad (16)$$

The non-local common cause structure in figure 4, for instance, implies the hidden joint probability $P(\alpha\beta|ab\psi\lambda) = P(\alpha|ab\psi\lambda)P(\beta|b\psi\lambda)$.

The rule (16) follows straightforwardly from the assumptions we have made: by the product rule (a theorem of probability theory), the hidden joint probability can in general be written as

$$P(\alpha\beta|ab\psi\lambda) = P(\alpha|\beta ab\psi\lambda)P(\beta|ab\psi\lambda) \qquad (17)$$

The CMC and Causal Sufficiency jointly guarantee that a variable becomes independent of all variables which are not effects given its direct causes. So if we can show that none of the variables in each of the conditionals on the right hand side in (17) is an effect of the outcome in question, then the rule follows. This, however, is easy to show: ECV guarantees that $\boldsymbol{a}$, $\boldsymbol{b}$ and ψ are not effects of the outcomes. Neither is $\boldsymbol{\lambda}$, for we have introduced it as a hidden common cause of the outcomes. Finally, concerning $\boldsymbol{\beta}$, which appears in the conditional of $\boldsymbol{\alpha}$, we have to discern two cases: if $\boldsymbol{\beta}$ causes $\boldsymbol{\alpha}$ or if there is no causal relation between the outcomes at all, $\boldsymbol{\beta}$ is not an effect of $\boldsymbol{\alpha}$ and everything is fine. If, however, $\boldsymbol{\alpha}$ causes $\boldsymbol{\beta}$, one has to choose the equivalent product form $P(\alpha\beta|ab\psi\lambda) = P(\alpha|ab\psi\lambda)P(\beta|\alpha ab\psi\lambda)$: in *this* form every variable in each conditional is not an effect of the respective outcome, so the rule follows. So in any case (16) holds (given ECV, the CMC and Causal Sufficiency$_1$).

According to (16), a certain causal structure determines the form of the hidden joint probability. In turn, by this rule, it is easy to find a generating causal structure for any given form of the hidden joint probability: just assume that every variable in each conditional is a direct cause of the respective outcome. Using the example from above the other way round we can say that the product form $P(\alpha|ab\psi\lambda)P(\beta|b\psi\lambda)$ is generated by the non-local common cause structure in figure 4. With this observation it is easy to formulate the corresponding causal principle to No Probabilistic Bell Contextuality (which is the characteristic of product forms implying Bell inequalities). Suppose we have a product form which obeys No Probabilistic Bell Contextuality, i.e. each of the outcomes depends probabilistically at most on one setting. Then, by the

rule we have just derived, a generating causal structure must be such that each outcome has at most one setting as its direct cause:

> **No Causal Bell Contextuality$_1$ (NCBC$_1$)**: None of the outcomes is a direct common effect of both settings relative to the set of variables $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}$.

If this principle and ECV, the CMC and Causal Sufficiency$_1$ hold, then, by (16), No Probabilistic Bell Contextuality follows. If also the structure in question obeys Causal Autonomy, Probabilistic Autonomy is implied as well. Then, Näger's stronger Bell argument entails Bell inequalities. So the causal assumptions Causal Autonomy and No Causal Bell Contextuality$_1$ are the causal interpretation (X′) of Näger's Bell argument. The new first premise of the stronger Bell argument explicitly reads:

> (P1″) The Exogeneity of Controlled Variables, the Causal Markov Condition, Causal Sufficiency$_1$, Causal Autonomy and No Causal Bell Contextuality$_1$ imply Probabilistic Autonomy and No Probabilistic Bell Contextuality:
> (ECV) ∧ (CMC) ∧ (CS$_1$) ∧ (CA) ∧ (NCBC$_1$) → (A) ∧ (NPBC)

The set of five causal assumptions in the antecedent, which provides the causal interpretation of the stronger Bell argument, differs from that of the standard argument, (P1′), in that we have replaced Causal Outcome Independence and Causal Parameter Independence by No Causal Bell Contextuality$_1$.

3.3 Which structures imply Bell inequalities?

What difference does this replacement make for the causal structures implying Bell inequalities? Up to this point we only have a very abstract understanding of this new causal Bell argument, because we have not shown yet what No Causal Bell Contextuality$_1$ specifically amounts to in terms of causal structures. Which causal structures fall under No Causal Bell Contextuality$_1$, and, hence, imply Bell inequalities?

The most interesting cases are, of course, *maximal* structures implying Bell inequalities. Structures are 'maximal' if one can*not add* any arrow to the structure without violating any of the causal assumptions which restrict them (i.e. ECV, Causal Autonomy and No Causal Bell Contextuality$_1$). We can find such maximal structures if we look for those structures which by rule (16) imply the maximal product forms (14) and (15) falling under No Probabilistic Bell Contextuality. According to the rule an appropriate structure for the first form is the one depicted in figure 23 (or its mirror image structure which has $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$ instead of $\boldsymbol{\alpha} \leftarrow \boldsymbol{\beta}$; the generating structures are in general not unique). Although there is a non-local causal relation between the outcomes, the structure is allowed by No Causal Bell Contextuality$_1$, because none of the outcomes is a *direct* common effect of the settings. It is true, $\boldsymbol{\alpha}$ is a common effect of the settings, but only *indirectly* of $\boldsymbol{b}$ (via $\boldsymbol{\beta}$); this is *not* ruled out by No Causal Bell Contextuality$_1$. The other maximal product form, (15), is generated by the structure in figure 24 (or again its mirror image). It is consistent with No

Causal Bell Contextuality$_1$ because each outcome is only a direct effect of its distant setting but not of its local one ($\boldsymbol{\alpha}$ is only an indirect effect of $\boldsymbol{a}$ via $\boldsymbol{\beta}$).

Maximal causal structures *implying* Bell inequalities

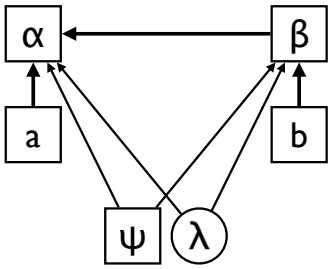

Fig. 23: Direct cause structure 1

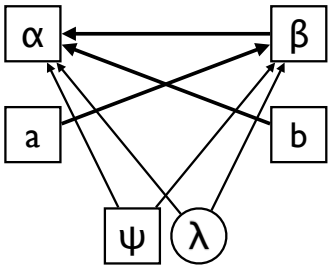

Fig. 24: Direct cause structure 2

Obeying all causal principles of the new Bell argument, each of the two structures in figures 23 and 24 implies Bell inequalities. Hence, we have shown that there are *non-local* causal structures which imply Bell inequalities! This means that some of the causal principles of the standard Bell argument have been too strong: both structures violate Causal Outcome Independence (because of the causal relation from one outcome to the other) and that in figure 24 also violates Causal Parameter Independence (due to the causal relations from each setting to its distant outcome). According to the standard Bell argument these structures wrongly count as being able to violate Bell inequalities. The causal interpretation of the stronger Bell argument, however, shows that this is not the case: these structures do imply Bell inequalities. This makes explicit that the new principle No Causal Bell Contextuality$_1$ is weaker than its predecessors Causal Outcome Independence and Causal Parameter Independence. According to the new argument there are *more* structures which imply Bell inequalities: not only local structures but also certain non-local ones do.

So far we have only considered maximal structures in figures 23 and 24. What other structures are there that imply Bell inequalities according to the new argument? Maximality meant that one may *not add* arrows without producing an inconsistency with the assumptions. However, *removing* any arrows from the maximal ones cannot yield a conflict with the principles (for these only *constrain* possible arrows). By rule (16) it is also clear that structures which derive from the maximal ones in this way have weaker product forms and, hence, also imply Bell inequalities. Let us call the set of all structures which one can gain by removing arrows from the maximal ones (including the latter) 'Bell structures'. For instance, erasing the connection $\boldsymbol{\alpha} \leftarrow \boldsymbol{\beta}$ in figure 23, yields the usual local structure (figure 12). This makes explicit that the local structures are a proper subset of the Bell structures. Since up to now only local structures were thought to imply Bell inequalities, *structures which derive from*

the ones in figures 23 and 24 and leave one of the non-local arrows in place are the news: these non-local Bell structures do imply Bell inequalities as well. So the present set of causal assumptions *extends* the set of structures which imply Bell inequalities by these non-local ones.

3.4 Minimal structures violating Bell inequalities

Since Bell inequalities are violated, (P3), this means that the new Bell argument from (P0″), (P1′) and (P3) possibly *excludes more* causal structures than the standard one. It has the new stronger conclusion:

(C2″) At least one of the following principles fails: Exogeneity of Controlled Variables or the Causal Markov Condition or Causal Sufficiency$_1$ or Causal Autonomy or No Causal Bell Contextuality$_1$:
$\neg(\text{ECV}) \vee \neg(\text{CMC}) \vee \neg(\text{CS}_1) \vee \neg(\text{CA}) \vee \neg(\text{NCBC}_1)$
(from P0″ & C1′)

I have said 'possibly' because in this form of the argument the violation might be explained by the failure of any of the causal assumptions. If, however, ECV, the CMC and Causal Sufficiency$_1$ hold, either Causal Autonomy or No Causal Bell Contextuality$_1$ must fail. Then, the true causal structure cannot be any of the Bell structures, and, as we have just seen, this rules considerably *more* causal structures out than the comparable result of the standard argument that the causal structure cannot be local. In this sense the new causal Bell argument is *stronger* than the standard one.

So if ECV, the CMC and Causal Sufficiency$_1$ hold, Bell structures are excluded. But what does this result mean *positively*? Which structures exactly are *allowed* because they can violate Bell inequalities? By what we have said up to now, it is clear that only those structures can violate Bell inequalities which fulfil one of the following conditions:

(i) At least one of the settings directly influences the hidden common cause $\boldsymbol{\lambda}$ (violation of Causal Autonomy, fig. 25).

or

(ii) At least one of the outcomes is a direct common effect of both settings (violation of No Causal Bell Contextuality$_1$, fig. 26).

These are the requirements from the stronger Bell argument. In section 2.4, we have argued that appropriate structures for EPR experiments have to meet the further conditions (a)–(c). Conditions (a) (there must be a directed path from the quantum state to each outcome) and (b) (there must be a directed path from each setting to its local outcome) guaranteed that the structures do not imply independencies which directly contradict the empirical distribution, while (c) stated that the hidden variable $\boldsymbol{\lambda}$ was defined as a common cause.

Minimal structures complying with these conditions can be found by choosing either (i) or (ii) and then adding causal paths such that also (a)–(c) hold. We have said above that (a) and (b) can be realised in different ways. Again, we do not list all of the possible minimal structures, but we focus on the most

Requirements for violating Bell inequalities
(according to the stronger Bell argument)

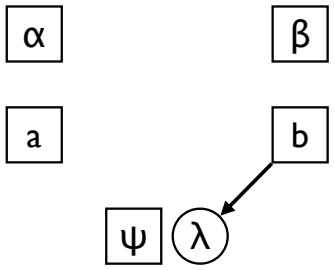

Fig. 25: Violation of Causal Autonomy

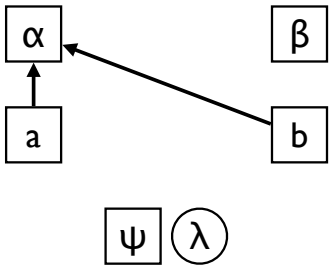

Fig. 26: Violation of No Causal Bell Contextuality$_1$

plausible cases which fulfil (a) by direct influences and (b) either by direct or by indirect influences via the hidden variable. This fraction of minimal structures is shown in figures 27 to 29. Any structure which derives from these by adding one or several arrows also implies Bell inequalities. The maximal structure which one can arrive at in this way is shown in figure 30.

By this construction rule it is clear that there are many variants of causal structures which can violate Bell inequalities. It is not informative to list them all. Rather, it is instructive to come back to the prototypes of causal structures in figures 2 to 7, and ask which of these can violate Bell inequalities according to the new argument. The answer is that the non-local common cause structure (fig. 4 = minimal structure in fig. 29), the mixed structure (fig. 5, non-minimal) and the indirect structure (fig. 6 = minimal structure in fig. 28) are still allowed. (The superdeterministic structure was already excluded by ECV, and the local structure by the standard Bell argument.) Thus, compared to the standard Bell argument, we have excluded one further prototype, viz. the direct cause structure (fig. 3). If ECV, the CMC and Causal Sufficiency$_1$ hold, *direct cause structures cannot violate Bell inequalities because they imply them.* This is the bold and simple message of the stronger causal Bell argument.

4 Discussion

4.1 Against the standard view 2: unmasking a false choice

Excluding direct cause structures the stronger Bell argument refutes the standard view that EPR correlations come about by an influence between the outcomes. It shows that structures whose only non-local influence goes from one outcome to the other still imply Bell inequalities (fig. 3). Such influences are too weak to produce a violation of Bell inequalities.

Against the conclusion of the standard Bell argument, (C2′), the new result

Minimal causal structures which can *violate* Bell inequalities
(according to the stronger Bell argument)

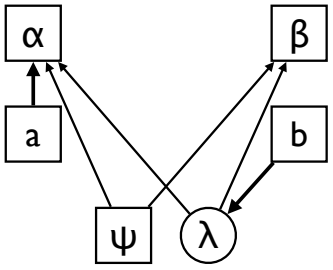

Fig. 27: Minimal structure 1

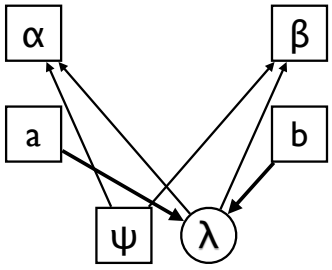

Fig. 28: Minimal structure 2

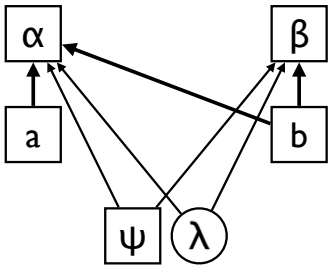

Fig. 29: Minimal structure 3

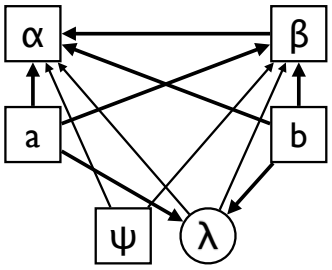

Fig. 30: Maximal structure

(C2″) demonstrates that there is no choice to be made between Causal Outcome Dependence and Causal Parameter Dependence (if the other assumptions—ECV, CMC, CS_1 and CA—hold): there must be a failure of No Causal Bell Contextuality$_1$, and since this means that at least one outcome is a direct common effect of both settings, Causal Parameter Dependence *must* hold. Causal Outcome Dependence, by contrast, does not play any crucial role: it might or might not hold.[16]

This result against the standard view is a causal analogue to Näger's (2012) claim that Jarrett's analysis 'Outcome Dependence or Parameter Dependence' is misleading on the probabilistic level. Näger shows that EPR correlations imply some kind of probabilistic dependence between a setting and its distant parameter (while they might be outcome dependent or not). Here we have derived a similar fact on the causal level: at least one of the settings must influence its distant parameter. The option of the standard causal Bell argument, 'Causal Outcome Dependence or Causal Parameter Dependence', presents us with a false choice. One cannot avoid Causal Parameter Dependence (if ECV, the CMC, CS_1 and CA hold).

Suggesting and supporting options which are not available is a second failure of the standard view. A first failure, we have said above (see section 2.5), was to naively associate Jarrett's probabilistic conditions Outcome Dependence and Parameter Dependence with causal meaning. Here we see that even a sound causal interpretation of the standard argument provides a false choice between Causal Outcome Dependence and Causal Parameter Dependence. The standard Bell argument is just inappropriate to yield correct details about the causal structure of EPR experiments.

4.2 Which assumption should be given up?

In the strongest conclusion of the Bell argument that we have reached, (C2″), there are five variants of how to explain the violation of Bell inequalities: either Exogeneity of Controlled Variables or the Causal Markov Condition or Causal Sufficiency$_1$ or Causal Autonomy or No Causal Bell Contextuality$_1$ fails. But which one is it? Which of the assumptions has to be considered the culprit for the fact that Bell inequalities are violated?

We have already expressed our loyalty to ECV throughout this paper. For failures of ECV would allow that variables which are otherwise controlled can still be influenced. The settings, for instance, are determined by an experimenter (or by a mechanism which has been setup by an experimenter) to take on the values that they do. To claim that nevertheless they are influenced by, say, a hidden variable $\boldsymbol{\lambda}$, would amount to saying that $\boldsymbol{\lambda}$ influences the experimenter or the mechanism in question or what else determines the settings. This is very implausible. Moreover, it would be a mystery how the hidden variables could influence *every* kind of system which happens to determine the setting to the right value. Reducing one mystery, quantum entanglement, to another

[16] Note, however, that Causal Parameter Dependence alone does not suffice. The non-local influence from one setting to its distant outcome must be *combined* with a local influence from the other setting to the same outcome. A non-local influence *per se* would *not* do.

is not a convincing explanation. We do not allow for such superdeterministic models of EPR experiments, according to which, for instance, a hidden variable is a common cause of both the outcomes and the settings (fig. 7).

Concerning Causal Sufficiency$_1$ we cannot be sure that it actually holds. It might be violated because besides $\boldsymbol{\lambda}$ there might be further hidden common causes, which we have not considered in our set of variables $\boldsymbol{V_1} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}\}$. Due to ECV such additional latent variables might only be common causes of any two (or more) variables in $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}\}$. The important fact, however, is that such a failure of Causal Sufficiency$_1$ alone cannot explain EPR correlations. For if only Causal Sufficiency$_1$ fails one can easily reformulate the argument to imply Bell inequalities again: just take into account all forgotten hidden common causes such that Causal Sufficiency with respect to the extended set of variables holds, and the derivation runs analogously as before.

For instance, suppose we have one further hidden common cause $\boldsymbol{\lambda}_2$ of the outcomes in a direct cause structure (fig. 31). The new product form reads $P(\alpha\beta|ab\psi\lambda\lambda_2) = P(\alpha|\beta a\psi\lambda\lambda_2)P(\beta|b\psi\lambda\lambda_2)$ and an appropriate autonomy condition holds, $P(\psi\lambda\lambda_2|ab) = P(\psi\lambda\lambda_2)$.[17] By these two probabilistic assumptions, Bell inequalities follow analogously as before. The causal interpretation of *this* argument assumes Causal Sufficiency$_2$, i.e. sufficiency with respect to the set of variables $\boldsymbol{V_2} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\psi}, \boldsymbol{\lambda}, \boldsymbol{\lambda}_2\}$. (Accordingly, some of the other assumptions have to be adapted slightly to the new set of variables.) So if one starts with one of the Bell structures but forgets a hidden common cause, one can always extend the argument such that Causal Sufficiency holds and Bell inequalities are implied. Since a reformulation works for any number n of ignored latent common causes, we conclude that a failure of Causal Sufficiency$_1$ cannot explain the violation of Bell inequalities.[18]

A similar judgement should be made concerning the Causal Markov Condition. On the one hand, it might fail and indeed it is not unlikely that it fails in the quantum world: the well-known counterexamples (van Fraassen, 1982; Cartwright, 1988) to Reichenbach's principle of the common cause (Reichenbach, 1956) also violate the CMC, because the former is an essential part of the latter. The examples show that in indeterministic worlds there can be common causes which do not screen-off, i.e. which do not make their effects probabilistically independent of another. Such non-screening-off common causes are the only serious threat to the CMC. On the other hand, we claim that a failure of the CMC in this way cannot explain the EPR correlations. This can be seen as follows. Suppose there were a common cause which does not screen-off in any of the Bell structures, say, the quantum state $\boldsymbol{\psi}$ in a local structure without

[17] The idea of Probabilistic Autonomy is that the complete state of the photons at the source, which includes the quantum state and *all* latent common causes, must be unconditionally independent of the settings. Note that the quantum state and the latent common causes do not have to be independent from another: it is consistent with Autonomy that there are influences from $\boldsymbol{\psi}$ to the latent common causes or between the latent common causes, e.g. $\boldsymbol{\lambda}_2 \rightarrow \boldsymbol{\lambda}$.

[18] Our result that even with *several* hidden common causes Bell inequalities follow is supported by a recent series of papers (Graßhoff et al., 2005; Portmann and Wüthrich, 2007; Hofer-Szabó, 2008), who claim that Bell inequalities can be derived from 'separate common causes' (which is a special case of the situation we have considered).

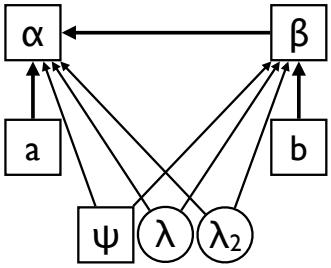

Fig. 31: Direct structure with a second hidden common cause $\boldsymbol{\lambda}_2$ of the outcomes

latent common cause (fig. 32). This would not suffice to explain EPR correlations because a common cause which does not screen-off can only produce the *same kind and strengths of correlations* as two common causes which jointly screen-off.[19] The latter equivalent situation, however, is just the local structure with hidden common cause $\boldsymbol{\lambda}$, which we know to imply Bell inequalities.[20]

[19] Formally, one can always redescribe a non-screening-off common cause C_1 as C_1 plus a latent common cause C_2, which coordinates the correlated effects and makes the pair C_1C_2 a screener-off. In order to justify this claim, consider Cartwright's example (1988) of a molecule in state S which decays into two parts moving in opposite direction. With equal probability the states of the two parts are given by their momenta $p, -p$ or by $p', -p'$, i.e. $P(p, -p|S) = \frac{1}{2} = P(p', -p'|S)$. So the states of the parts after the decay are perfectly anti-correlated, but the state of the molecule S does not screen off the correlation:

$$P(p, -p|S) = \frac{1}{2} \neq P(p|S)P(-p|S) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \tag{18}$$

However, this situation can be redescribed as there being a two-valued latent common cause ($\boldsymbol{\lambda} = \lambda_0, \lambda_1$), which determines whether the parts will move along the axis p (if λ_0 holds) or along the axis p' (if λ_1 is present), i.e. $P(p, -p|S\lambda_0) = 1$ and $P(p', -p'|S\lambda_0) = 1$. Thus, jointly with the latent common cause $\boldsymbol{\lambda}$ the state of the molecule S does screen off:

$$P(p, -p|S\lambda_0) = 1 = P(p|S\lambda_0)P(-p|S\lambda_0) = 1 \cdot 1 = 1 \tag{19}$$
$$P(p, -p|S\lambda_1) = 0 = P(p|S\lambda_0)P(-p|S\lambda_0) = 0 \cdot 0 = 0 \tag{20}$$
$$P(p', -p'|S\lambda_0) = 0 = P(p'|S\lambda_0)P(-p'|S\lambda_0) = 0 \cdot 0 = 1 \tag{21}$$
$$P(p', -p'|S\lambda_1) = 1 = P(p'|S\lambda_1)P(-p'|S\lambda_1) = 1 \cdot 1 = 1 \tag{22}$$

This shows that common causes which do not screen off can be redescribed as screening off jointly with a latent common cause. Note, however, that we do *not* claim that all cases of non-screening-off common causes *in fact* invoke a latent common cause. Rather, we believe that true non-screening-off common causes are possible. Here we just want to make the point that a common cause which does not screen off can explain correlations only to the same degree as a normal common cause plus a latent common cause.

[20] The result does not change if one claimed that even given both common causes $\boldsymbol{\psi}$ *and* $\boldsymbol{\lambda}$ the outcomes would not screen off from another. For this situation can be redescribed as there being another hidden common cause $\boldsymbol{\lambda}_2$ such that $\boldsymbol{\psi}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_2$ screen off—and these scenarios, according to which Causal Sufficiency$_1$ fails, we have just argued, can be

Therefore, non-screening-off common causes in Bell structures cannot *per se* explain the violation of Bell inequalities. A failure of the CMC does not suffice to explain EPR correlations.

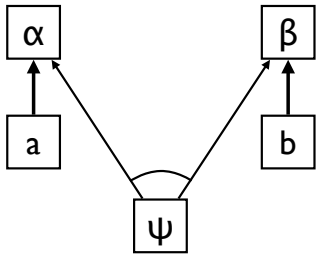

Fig. 32: Local structure with a common cause which does not screen-off

It might be criticised that quantum mechanics is a counterexample to this claim, because it has the described structure (ψ is a common cause which does not screen-off in a local graph) but violates Bell inequalities. This, however, is not correct: besides a common cause which does not screen-off, quantum mechanics involves an influence from one of the settings to its distant outcome. We shall justify this claim in section 4.3 below.

The only remaining assumptions whose failure might reasonably account for EPR correlations are Causal Autonomy and No Causal Bell Contextuality$_1$. We have already made explicit in section 3.4 which structures are required if one of these conditions fails: either there is an influence from one setting to the hidden variable $\boldsymbol{\lambda}$ or one outcome is directly influenced by both settings. This seems to offer a choice between two basic types of violating Bell inequalities. However, we emphasise here that this is only apparently so because there is a *common idea* behind both alternatives:

> **Causal Bell Contextuality$_2$:** One of the outcomes is a common effect of both settings, either directly (violation of No Causal Bell Contextuality$_1$) or (partly) indirectly via the hidden variable $\boldsymbol{\lambda}$ (violation of Causal Autonomy).

One outcome being the common effect of both settings is the main idea of what EPR correlations require in causal terms.[21]

reformulated as to imply Bell inequalities as well.

[21] Note that it does *not* suffice that one outcome is a common effect of both settings indirectly *via the other outcome* (i.e. a direct cause structure 3). There are crucial differences between an influence mediated by $\boldsymbol{\lambda}$ and one mediated by one of the outcomes: first, the outcomes are two valued ($\boldsymbol{\lambda}$ can in principle have infinitely many values) and, second, it is an empirical fact that there are perfect (anti-)correlations of the outcomes if the settings agree (disagree by 90°). (Since $\boldsymbol{\lambda}$ is hidden, we do not know whether similar facts hold there.) Due to these two key facts direct cause structures imply Bell inequalities (see Näger, 2012).

One might object that a violation of Causal Autonomy does not by itself imply that one of the outcomes is a common effect of the settings. This is true. But an influence from one setting to the hidden variable, say, $\boldsymbol{b} \to \boldsymbol{\lambda}$, *does* imply the alleged fact if we involve the conditions (b) and (c) from above (see section 2.4). The latter says that $\boldsymbol{\lambda}$ is a common cause of the outcomes, so $\boldsymbol{b}$ influences both outcomes via $\boldsymbol{\lambda}$. The former condition says that there is a directed causal path from each setting to its local outcome, thus $\boldsymbol{a}$ must influence $\boldsymbol{\alpha}$. Hence, $\boldsymbol{\alpha}$ is a common effect of both settings, if Causal Autonomy is violated by $\boldsymbol{b} \to \boldsymbol{\lambda}$ (and mutatis mutandis for violations $\boldsymbol{a} \to \boldsymbol{\lambda}$). This can be seen in the minimal structures which violate Causal Autonomy (fig. 27 and 28).[22]

4.3 The causal structure according to quantum mechanics

Having discussed general principles and structures we shall now turn to the question which specific causal structure quantum mechanics has. Quantum mechanics is widely assumed to have a direct structure. In fact, one of the main motivations for the standard view (Probabilistic and Causal Outcome Dependence) seems to have been the believe that this is the choice that quantum mechanics suggests. By what our argument in this paper, however, it has become clear that quantum mechanics cannot have a direct structure—because otherwise it could not violate Bell inequalities.

Which structure then? Against the standard view, Cartwright has argued that EPR correlations according to quantum mechanics come about by a common cause which does not screen off:

> ... the quantum state consequent on the interaction operates, in conjunction with the separated apparatuses, as a joint cause of the results in each wing, with no direct causal connection between one wing and the other. (Cartwright, 1989, p. 243)

We interpret this claim as suggesting a structure like the one in figure 32—but this cannot be true either. For we have argued above that correlations brought about merely by common causes which do not screen-off, i.e. without any connection between the wings, are too weak as well.

So again: what is the causal structure of quantum mechanics? We agree with Cartwright that quantum mechanics involves a common cause which does not screen off, but we shall show that additionally it involves an influence from one setting to its distant outcome. We shall now give an argument for this claim. The idea is to endow the formalism of (non-relativistic) quantum mechanics with a causal interpretation. Since here we have a full-fledged formalism, we can use a more direct interpretation method than above: we write down the quantum mechanical description of processes in an EPR experiment and directly

[22] Indeed, if $\boldsymbol{b}$ violates Causal Autonomy by $\boldsymbol{b} \to \boldsymbol{\lambda}$ and $\boldsymbol{a}$ would not influence its local or distant outcome, i.e. it would not be a cause at all, Bell inequalities would follow trivially. But if $\boldsymbol{a}$ influences *any* of the other variables (consistent with ECV), one of the outcomes is a common effect of $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e. Bell inequalities are not implied any more and can be violated.

interpret them in a causal way. The criterion of a variable $\boldsymbol{X}$ causing another $\boldsymbol{Y}$ in this procedure is that $\boldsymbol{X}$ is among the variables which determine $\boldsymbol{Y}$ (or determine the probability of $\boldsymbol{Y}$) according to the most detailed description of the formalism. This amounts to reading the quantum mechanical formalism as a kind of structural equation model and associating the usual causal meaning, i.e. to interpret the free variables as causes and the dependent variables as effects.

We describe the experiment in the laboratory frame and assume that one of the measurements is (at least slightly) earlier than the other (here: measurement at B before measurement at A). A detailed quantum mechanical description of one run of an EPR experiment comes in seven steps, which are as follows (one can find the causal sub-structure for each step (1)–(7) in figure 33):

(1) At the source, quantum mechanics ascribes a joint entangled polarisation state to the photons A and B, e.g.

$$\boldsymbol{\psi_{AB}} = \frac{1}{\sqrt{2}}(|+_z\rangle_A|+_z\rangle_B + |-_z\rangle_A|-_z\rangle_B). \tag{23}$$

$|+_z\rangle_A$ means that photon A is polarised in the z-direction, while $|z_-\rangle_A$ means that it is polarised perpendicular to that direction (and analogously for states of photon B). An entangled state $\boldsymbol{\psi_{AB}}$ cannot be written as a product of one state for A and one for B, so the single photons have no defined states, only the compound system has. While the photons move to the measurement devices the polarisation state does not change.

We assume a setup with two-channel measurement devices, i.e. a polarising beam splitter (or a similar analyser) divides the beam into one with polarisation parallel and another perpendicular to the measurement direction. Photons in each beam are registered by a detector, so there are two detectors per device. (Compared to one-channel setups with a polarisation filter and *one* detector the former has the advantage to clearly separate between analysis and measurement process; for filters already absorb the perpendicular beam.)

(2) When the first photon, say B, passes the analyser, its state is split up according to the basis defined by the orientation of the device, $\boldsymbol{b}$. So the state $\boldsymbol{\psi}_{AB}$, which is given in arbitrary basis (in our example the z-Basis), is now expanded relative to the $\boldsymbol{b}$-basis. If $\boldsymbol{\psi}_{AB}$ is rotationally invariant in the plane of the measurement settings (as in our example), the new state $\boldsymbol{\psi}'_{AB}$ is form-invariant:

$$\boldsymbol{\psi'_{AB}} = \frac{1}{\sqrt{2}}(|+_b\rangle_A|+_b\rangle_B + |-_b\rangle_A|-_b\rangle_B). \tag{24}$$

We interpret the new state $\boldsymbol{\psi'_{AB}}$ as an effect of the former state $\boldsymbol{\psi_{AB}}$ and the measurement setting $\boldsymbol{b}$.

(3) The presence of detectors at B triggers that the entangled state $\boldsymbol{\psi'_{AB}}$ collapses indeterministically, in our example either onto $|+_b\rangle_A|+_b\rangle_B$ or onto $|-_b\rangle_A|-_b\rangle_B$ (with probability $\frac{1}{2}$, respectively). This is the central step in the coming about of EPR correlations. It seems natural to interpret it such that $\boldsymbol{\psi'_{AB}}$ and the presence of a detector, $\boldsymbol{D_B}$, cause the new polarisation state. However, as this new state is a product state, each photon now possesses its own separate state, $\boldsymbol{\psi_A}$ or $\boldsymbol{\psi_B}$, respectively. So we should not say that $\boldsymbol{\psi'_{AB}}$

and $\boldsymbol{D_B}$ cause *one joint* state but rather that $\boldsymbol{\psi'_{AB}}$ decays into *two separate* states, and the decay is triggered by $\boldsymbol{D_B}$. The most plausible reading of this process is that $\boldsymbol{D_B}$ brings $\boldsymbol{\psi'_{AB}}$ into a new state $\boldsymbol{\psi''_{AB}}$ which is perfectly similar to $\boldsymbol{\psi'_{AB}}$ but is disposed to collapse instantly. Finally, the indeterministic decay of $\boldsymbol{\psi''_{AB}}$ into the two states $\boldsymbol{\psi_A}$ and $\boldsymbol{\psi_B}$ should be understood as the former being a common cause of the latter two states.

So far for the causal structure of this crucial step. The associated probabilities are:

$$P\Big(|+_b\rangle_A|+_b\rangle_B\Big|\psi''_{AB}\Big) = \frac{1}{2} \tag{25}$$

$$P\Big(\underbrace{|-_b\rangle_A}_{\boldsymbol{\psi_A}}\underbrace{|-_b\rangle_B}_{\boldsymbol{\psi_B}}\Big|\psi''_{AB}\Big) = \frac{1}{2} \tag{26}$$

One can see that the common cause $\boldsymbol{\psi''_{AB}}$ makes the separate states $\boldsymbol{\psi_A}$ and $\boldsymbol{\psi_B}$ perfectly correlated (according to our example). However, and this is another important feature of this central process in EPR experiments, the probabilities imply that the common cause $\boldsymbol{\psi''_{AB}}$ does *not screen-off* the correlation between the new photon states:[23]

$$\underbrace{P\Big(\psi_A\psi_B\Big|\psi''_{AB}\Big)}_{\frac{1}{2}} \neq \underbrace{P\Big(\psi_A\Big|\psi''_{AB}\Big)}_{\frac{1}{2}}\underbrace{P\Big(\psi_B\Big|\psi''_{AB}\Big)}_{\frac{1}{2}} \tag{27}$$

We represent the non-screening-off of $\boldsymbol{\psi''_{AB}}$ by a bow between its outgoing arrows in the causal graph.

This was the crucial part of the quantum mechanical story for entangled states. The rest is unsurprising and can be told quickly:

(4) When photon B is in state $\boldsymbol{\psi}_B = |+_b\rangle_B$ it gives a count at detector β_+, otherwise (when it is in state $\boldsymbol{\psi}_B = |-_b\rangle_B$) at β_-. So it seems obvious that $\boldsymbol{\psi}_B$ influences $\boldsymbol{\beta}$.

(5) When photon A in state $\boldsymbol{\psi_A} = |+_b\rangle_A$ (or $\boldsymbol{\psi_A} = |-_b\rangle_A$) reaches the analyser, the measurement direction $\boldsymbol{a}$ determines the basis relative to which $\boldsymbol{\psi}_A$ is expanded. This gives a new state:

$$\boldsymbol{\psi'_A} = \cos(a-b)|+_a\rangle_A + \sin(a-b)|-_a\rangle_A \tag{28}$$

$$\Big(\text{or} \quad \boldsymbol{\psi'_A} = -\sin(a-b)|+_a\rangle_A + \cos(a-b)|-_a\rangle_A\Big) \tag{29}$$

We interpret this procedure as $\boldsymbol{a}$ and $\boldsymbol{\psi_A}$ causing $\boldsymbol{\psi'_A}$.

(6) The presence of detectors at A, $\boldsymbol{D_A}$, triggers that the state $\boldsymbol{\psi}'_A$ collapses, either onto $\boldsymbol{\psi}'''_A = |+_a\rangle_A$ or onto $\boldsymbol{\psi}'''_A = |-_a\rangle_A$.[24] As in step (3) we understand

[23] This case is perfectly similar to the examples by van Fraassen (1982) and Cartwright (1988).

[24] The probabilities for the different alternatives are:

$$P\Big(\boldsymbol{\psi'''_A} = |\pm_a\rangle_A\Big|\boldsymbol{\psi'_A} = |\pm_b\rangle_A, D_A\Big) = \cos^2(a-b) \tag{30}$$

$$P\Big(\boldsymbol{\psi'''_A} = |\mp_a\rangle_A\Big|\boldsymbol{\psi'_A} = |\pm_b\rangle_A, D_A\Big) = \sin^2(a-b) \tag{31}$$

the triggering as $\boldsymbol{D_A}$ and $\boldsymbol{\psi'_A}$ bringing about a state $\boldsymbol{\psi''_A}$, identical to $\boldsymbol{\psi'_A}$ save that it collapses instantly into one of the possible states $\boldsymbol{\psi'''_A}$. Thus we say that $\boldsymbol{D_A}$ and $\boldsymbol{\psi'_A}$ are causes of $\boldsymbol{\psi''_A}$, and the latter causes $\boldsymbol{\psi'''_A}$.

(7) Finally, when photon A is in state $\boldsymbol{\psi'''_A} = |+_a\rangle_A$ it is registered at detector α_+, otherwise (when it is in state $\boldsymbol{\psi'''_A} = |-_a\rangle_A$) at α_-. As on the other wing, it seems obvious that the measurement outcome at A, $\boldsymbol{\alpha}$, is solely determined by $\boldsymbol{\psi'''_A}$.

The detailed causal graph for these seven steps according to quantum mechanics is shown in figure 33. Besides the five empirically accessible variables that we have used so far, it contains nine more variables. So how can we relate this detailed result to our prototypes? The answer is that we have to eliminate all extra variables. Eliminating a variable $\boldsymbol{x}$ from the set of considered variables means to replace all incoming and outgoing arrows of $\boldsymbol{x}$ by arrows from each direct cause of $\boldsymbol{x}$ to each direct effect of $\boldsymbol{x}$. A simple example for this procedure is demonstrated in figures 34 and 35. If we apply the same method to the detailed causal structure of quantum mechanics (fig. 33) and eliminate all variables which do not belong to the set we have used so far (except the true common cause ψ_{AB}), we arrive at the structure in figure 36.

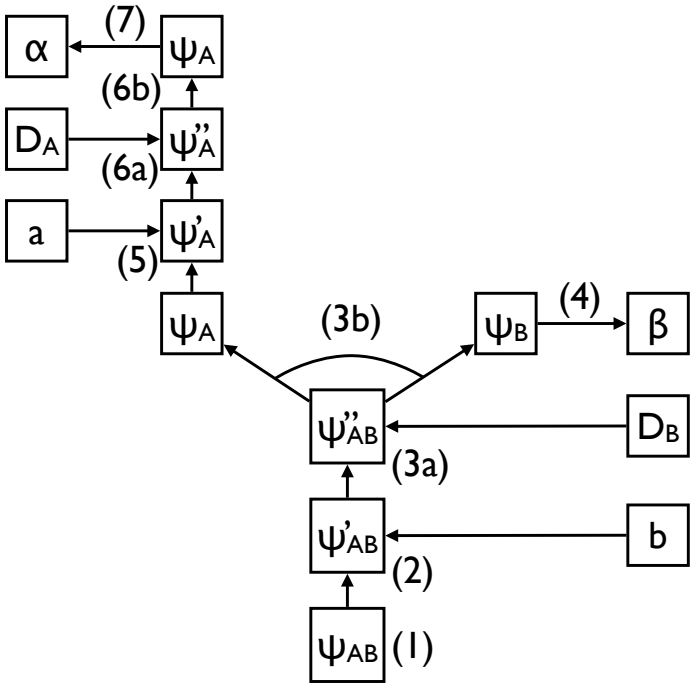

Fig. 33: Detailed causal structure according to quantum mechanics

This structure involves a common cause which does not screen off. However, it is not the quantum state at the source which is the true common cause, but it is the entangled quantum state which has passed the first analyser, and, hence,

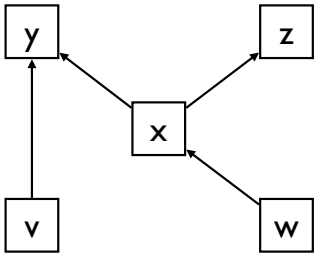

Fig. 34: Example of a causal structure

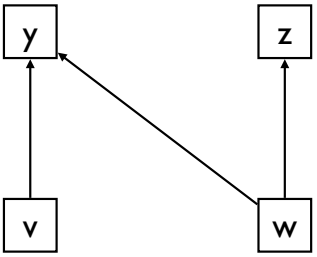

Fig. 35: Eliminating variable $\boldsymbol{x}$ from structure 34

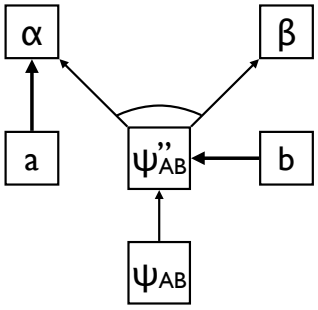

Fig. 36: Reduced causal structure according to quantum mechanics

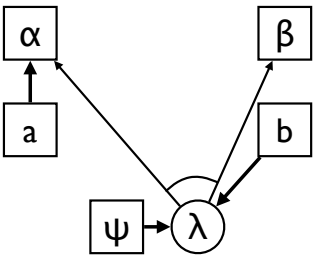

Fig. 37: Structure 36 in the usual variable scheme

has been affected by one of the settings (here: $\boldsymbol{b}$). Therefore, its subsequent decay transfers the influence of this setting to the distant outcome, so there is a directed path from the setting $\boldsymbol{b}$ to the distant outcome $\boldsymbol{\alpha}$. This, however, does not mean that quantum mechanics is a non-local common cause structure (fig. 4). For the influence is *not direct* but mediated via a latent common cause $\boldsymbol{\psi}''_{AB}$. In our considerations so far the role of a latent common cause was played by a variable which we have called $\boldsymbol{\lambda}$. Here it is crucial to recall that $\boldsymbol{\lambda}$ was not necessarily identical to hidden variables but could be any latent common cause located anywhere. In this situation, $\boldsymbol{\lambda}$ is the quantum state $\boldsymbol{\psi}''_{AB}$ which has passed the first analyser, which gives us the structure in figure 37. That structure is similar but not identical to the minimal indirect structure in figure 27. There are two differences: first, according to the former the latent common cause does not screen off, and, second, the directed paths from $\boldsymbol{\psi}$ to each outcome (which must hold according to condition (a)) is not direct but indirect via $\boldsymbol{\lambda}$.

Thus, in our scheme of causal strcutures, quantum mechanics has an *indirect structure*! Neither it is a direct structure, as the standard view says, nor is it a local common cause structure with a non-screening-off common cause, as Cartwright has proposed. Even more surprisingly, the quantum mechanical structure we have found does not conform to any of the prototypes which are commonly discussed! But it is in accordance with the argument we have developed in this paper: it violates Causal Autonomy as well as the Causal Markov Condition.

4.4 The meaning of Bell inequalities

Another lesson of our argument concerns the meaning of Bell inequalities: what do we know about a system if we learn that it fulfils or violates Bell inequalities? Do we learn anything about its causal structure or about its spatio-temporal realisation?

It might sound surprising but the logic of the Bell argument has it that a system *obeying* Bell inequalities does not give you any information about it. Normally Bell inequalities are so closely linked to the notion of locality that one might have got the impression that obeying them is a necessary and sufficient condition thereof. However, our argument only shows that local structures (plus some extra conditions) imply Bell inequalities—but not the other way round. No matter which causal structure a system has, whether the influences are local or non-local, whether it is causally Bell contextual$_2$ or not, in all these cases Bell inequalities can hold. *Fulfilling Bell inequalities does not have any causal or spatio-temporal meaning, especially not that the causal structure must be local.* It is true, we have derived that certain structures, for instance local ones, *necessarily* obey Bell inequalities. But even those which do not, because they are causally Bell contextual$_2$, still *can* obey Bell inequalities. Causal Bell Contextuality$_2$ is only a necessary condition for violating Bell inequalities.

In turn, however, this means that we do learn something about a system if it *violates* Bell inequalities. We then know that the system's causal structure must be Bell contextual$_2$ (if ECV holds). Violating Bell inequalities does have

a causal meaning. We should stress, however, that its meaning does *not* refer to *singular* causal connections. It does not mean that there must be a causal connection of this or that sort. Rather it indicates that there must be the right *combination* of influences, viz. that one of the outcomes is a common effect of the settings. Thus, violating Bell inequalities indicates properties of a causal *structure*, not the existence of singular causal connections like Causal Outcome Dependence or Causal Parameter Dependence.

Besides information about its causal structure (which variable influences which), can we also infer anything about a system's spatio-temporal realisation, if we learn that it violates Bell inequalities? Against a common view that the violation of Bell inequalities implies that there is a non-locality of some kind, here we have shown that a violation of Bell inequalities does *not per se* imply anything about the spatio-temporal realisation of its influences. For we could separate the problem which variable influences which from the spatio-temporal realisation (section 2.3). Our argument, which relied upon the violation of Bell inequalities, told us something about the former question but not about the latter. The violation of Bell inequalities provides information about which variable influences which but not about the fact whether there is a non-local influence or not. In the typical setup it only says that there are influences between space-like separated variables, e.g. one setting and its distant outcome; but these can be realised either directly non-local (fig. 4), or indirectly non-local via a hidden variable at the source (fig. 6), or indirectly backwards and forwards in time (zig-zag causation, fig. 10). Our result leaves this question open. (Especially, we have not touched the question whether and how such structures are compatible with relativity.)

To sum this point up: *intrinsically, the violation of Bell inequalities is not about spatio-temporal features (non-locality) but about certain combinations of influences.* It implies a minimal structural property which variables have to influence which, viz. that at least one of the outcomes is a common effect of both settings (Causal Bell Contextuality$_2$). Only by further arguments a non-locality can be inferred.

4.5 Agreement with information theoretic results

Finally, we come back to the information theoretic results of Maudlin (2002) and Pawlowski et al. (2010) that we mentioned at the start. How does our conclusion, that at least one of the outcomes must be a common effect of both settings, relate to their result, that at least one of the outcomes must depend on the *information* about both settings? (Let us call the latter kind of dependence 'informational dependence'.) The two results sound very similar. Indeed, if 'depending on the information of' were identical to 'being influenced by', the two results would be the same. However, this is not the case. Informational dependence is more like a correlation (probabilistic dependence): a variable $\boldsymbol{X}$ might depend on the information of another variable $\boldsymbol{Y}$ without there being any influence from $\boldsymbol{Y}$ to $\boldsymbol{X}$. Just as for correlations, the influence might be the other way round or there might be a common cause of the two variables. In both cases the value of $\boldsymbol{X}$ can depend on information about the value of $\boldsymbol{Y}$.

Hence, the relation of informational dependence to causal facts seems similar to that of probabilistic dependence to causal facts. There is no one-to-one correspondence, and the translation is subject to multiple possible fallacies. Only with additional assumptions and clear principles the translation can be reliable.

Therefore, Maudlin's unmediated leap from '[*informational*] dependence on the distant polarizer setting is crucial' to 'any successful theory must postulate some *influence* of a distant "parameter" (i.e. the polarizer angle) on the response of a local photon' (my emphasis) is not warranted without further justification. In this paper, we have made explicit the assumptions that one needs in order to do a similar inference from probabilistic facts to causal ones. It is likely that very analogous principles would do the same job for Maudlin's case. In fact, we should use the principle ECV and modify the Causal Markov Condition from relating causal and probabilistic facts to relating causal and informational facts:

> **Causal-informational Markov condition (CIMC)**: A variable $\boldsymbol{A}$ in a given causal structure is *informationally* independent of its non-effects $\boldsymbol{B}$ given *information* about its direct causes $\boldsymbol{C}$.

Then, with these principles, we could justify Maudlin's leap from his result that one of the outcomes must depend informationally on both settings to Causal Bell Contextuality$_2$ as follows. Since the direction of inference is against the direction of the (CIMC) we have to proceed indirectly: suppose Causal Bell Contextuality$_2$ does *not* hold, e.g. because both outcomes are only effects of their local settings (but not of their distant ones). Moreover, by ECV, which says that the settings cannot be effects, there is no other causal connection between the outcomes and their distant parameters: there cannot be a common cause of an outcome and one of the settings, and neither can there be a causal relation from an outcome to the setting. But then, by the CIMC[25] each outcome would be informationally independent of its distant setting—which contradicts Maudlin's result. Hence, one of the assumptions in the short argument we have made must be false. If ECV and the CIMC hold, it can only be the failure of Causal Bell Contextuality$_2$ which—contrary to our premisses—must be wrong.

This is a causal interpretation of Maudlin's result. It has the same causal conclusion, namely that Causal Bell Contextuality$_2$ holds, as our causal interpretation of the stronger Bell argument. It also has the same assumptions: both arguments proceed from EPR correlations (which violate Bell inequalities), ECV and the Causal Markov Condition, either in its original form CMC or in the form of the CIMC. In this sense, our causal Bell argument is in agreement with a causal interpretation of the information theoretic results.[26]

[25] Note that here we need not assume Causal Sufficiency$_1$ because ECV already forbids common causes of an outcome and its distant setting.

[26] We should note that there is also a sense in which the information theoretic approach outdoes our causal argument: the former explicitly determines the *amount* of information about the settings which is needed for reproducing the EPR correlations. Maudlin calculates that on average 1.174 bits of information must be transmitted about the distant parameter. This

Being the stable result of two independent approaches—an information theoretic approach via EPR correlations and a probabilistic approach via the violation of Bell inequalities—Causal Bell Contextuality$_2$ is a good candidate for becoming the new standard view: EPR correlations require that one of the outcomes is a common effect of both outcomes.

Acknowledgements

I would like to thank Alexander Gebharter, Andreas Hüttemann, Kevin Kelly, Meinard Kuhlmann, Thorben Petersen, Gerhard Schurz, Manfred Stöckler, Michael Zehetleitner and audiences at the Heinrich-Heine University Düsseldorf, at the Munich Center for Mathematical Philosophy and at the Workshop 'Physics and Causality' (Munich, 2012) for helpful comments and discussion. This paper is part of a research project funded by the Deutsche Forschungsgemeinschaft (DFG).

provides another explanation, why a direct structure cannot reproduce the correlations: in a direct structure one of the parameters influences the distant outcome *via* the local outcome. Since the outcome is a two-valued variable, it can maximally code *one* bit of information about its local setting—which is not enough to violate Bell inequalities. By the same reasoning it follows that in indirect models where the information of the distant setting is transferred via the hidden variable, the hidden variable cannot be two-valued.

References

Arntzenius, F. (1994). Relativistic hidden variable theories? *Erkenntnis 41*(2), 207–231.

Aspect, A., J. Dalibard, and G. Roger (1982). Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters 49*, 1804–1807.

Bell, J. S. (1975). The theory of local beables. *TH-2053-CERN*.

Bohm, D. (1951). *Quantum Theory*. Englewood Cliffs: Prentice-Hall.

Butterfield, J. N. (1992). David Lewis meets John Bell. *Philosophy of Science 59*, 26–43.

Cartwright, N. (1988). How to tell a common cause. In J. H. Fetzer (Ed.), *Probability and Causality: Essays in Honor of Wesley C. Salmon*, pp. 181–188. Dordrecht: Reidel.

Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.

Einstein, A., B. Podolsky, and N. Rosen (1935). Can quantum mechanical description of physical reality be considered complete? *Physical Review 47*, 777–780.

Glymour, C. (2006). Markov properties and quantum experiments. In W. Demopoulos and I. Pitowsky (Eds.), *Physical Theory and Its Interpretation : Essays in Honor of Jeffrey Bub*. Dordrecht: Springer.

Glynn, L. and T. Kroedel (forthcoming). Relativity, quantum entanglement, counterfactuals, and causation. *British Journal for the Philosophy of Science*.

Graßhoff, G., S. Portmann, and A. Wüthrich (2005). Minimal assumption derivation of a Bell-type inequality. *British Journal for the Philosophy of Science 56*(4), 663–680.

Healey, R. A. (1991). Holism and nonseparability. *Journal of Philosophy 88*, 393–421.

Healey, R. A. (1994). Nonseparable processes and causal explanation. *Studies in History and Philosophy of Science 25*, 337–374.

Hofer-Szabó, G. (2008). Separate- versus common-common-cause type derivations of the Bell inequalities. *Synthese 163*(2), 199–215.

Howard, D. (1989). Holism, separability, and the metaphysical implications of the Bell experiments. In J. T. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, pp. 224–253. Notre Dame: University of Notre Dame Press.

Jarrett, J. P. (1984). On the physical signifance of the locality conditions in the Bell arguments. *Nous 18*, 569–590.

Jarrett, J. P. (1989). Bell's theorem: A guide to the implications. In J. T. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, pp. 60–79. Notre Dame: University of Notre Dame Press.

Jones, M. R. and R. K. Clifton (1993). Against experimental metaphysics. In P. A. French, T. E. Uehling, and H. K. Wettstein (Eds.), *Philosophy of science*, pp. 295–316. Notre Dame, Ind: University of Notre Dame Press.

Maudlin, T. (2002). *Quantum Non-locality and Relativity : Metaphysical Intimations of Modern Physics*. Malden, Mass.; Oxford: Blackwell.

Näger, P. M. (2012). A stronger Bell argument for quantum non-locality. *Preprint: `http://philsci-archive.pitt.edu/9068/`*.

Pawlowski, M., J. Kofler, T. Paterek, M. Seevinck, and C. Brukner (2010). Non-local setting and outcome information for violation of bell's inequality. *New Journal of Physics 12*(8), 083051.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Portmann, S. and A. Wüthrich (2007). Minimal assumption derivation of a weak Clauser-Horne inequality. *Studies in History and Philosophy of Modern Physics 38*, 844–862.

Price, H. (1996). *Time's arrow and Archimedes' point : new directions for the physics of time*. Oxford: Oxford University Press.

Redhead, M. L. G. (1987). *Incompleteness, Nonlocality, and Realism : A Prolegomenon to the Philosophy of Quantum Mechanics*. Oxford New York: Clarendon Press Oxford University Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Salmon, W. C. (1990). *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press.

Salmon, W. C. (1998). *Causality and Explanation*. Oxford: Oxford University Press.

Schrödinger, E. (1935). Die gegenwärtige Situation in der Quantenmechanik. *Die Naturwissenschaften 23*, 807–812; 823–828; 844–849.

Schurz, G. and A. Gebharter (preprint). Causality as a theoretical concept with empirical significance.

Shimony, A. (1984). Controllable and uncontrollable non-locality. In S. Kamefuchi (Ed.), *Foundations of Quantum Mechanics in the Light of New Technology*, pp. 225–230. Tokyo: The Physical Society of Japan.

Shimony, A. (1990). An exposition of Bell's theorem. In A. I. Miller (Ed.), *Sixty-two Years of Uncertainty : Historical, Philosophical, and Physical Inquiries Into the Foundations of Quantum Mechanics*, pp. 33–43. New York ; London: Plenum.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search*. Cambridge, MA: The MIT Press.

Suárez, M. (2004). Causal processes and propensities in quantum mechanics. *Theoria 51*, 271–300.

Suárez, M. (2007). Quantum propensities. *Studies in History and Philosophy of Modern Physics 38*, 418–438.

Teller, P. (1986). Relational holism and quantum mechanics. *Brit. J. Phil. Sci. 37*, 71–81.

van Fraassen, B. C. (1982). Rational belief and the common cause principle. In R. McLaughlin (Ed.), *What? Where? When? Why?*, pp. 193–209. Dordrecht: Reidel.

van Fraassen, B. C. (1989). The charybdis of realism: Epistemological implications of Bell's inequality. In J. T. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, pp. 97–113. Notre Dame: University of Notre Dame Press.

Weihs, G., T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger (1998). Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett. 81*(23), 5039–5043.

Wood, C. J. and R. W. Spekkens (2012). The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning. *Preprint:* `http://arxiv.org/abs/1208.4119`.