Gerd Graßhoff, Samuel Portmann, Adrian Wüthrich

Causality and Quantum Theory

History and Philosophy of Science

University of Bern

Inhalt:

Graßhoff, G. and M. May (2001). Causal Regularities. In: Current Issues in Causation, edited by W. Spohn, M. Ledwig and M. Esfeld: Mentis Verlag, 85-114.

Graßhoff, G., S. Portmann, and A. Wüthrich (2005). Minimal assumption derivation of a Bell-type inequality. In: Brit. J. Phil. Sci. 56, 663–680.

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

Separat:

Wüthrich, A. (2004). Quantum Correlations and Common Causes - An Analysis of the Assumptions of Bell's Theorem. Bern: Bern Studies in the History and Philosophy of Science.

CAUSAL REGULARITIES

Gerd Graßhoff and Michael May

Abstract: Regularity theories of causality face a standard set of objections related to the identification of epiphenomena, spurious correlation and causal overdetermination. Even John Mackie's refined version of a regularity theory in form of INUS-conditions failed to solve a counter example formulated by Mackie himself - the Manchester factory hooters. This paper proposes an extension of traditional regularity theories of causation which does not only counter the stock objections but also solves the case of the hooters. The crucial definition of causal relevance utilises the idea of minimality: a factor is causally relevant for another factor, if it is part of a minimally necessary condition of minimally sufficient conditions for that factor. This idea of minimality is critical for the identification of true causal relations.

1. Introduction

Causation, Hume said, is a relation between events. In contemporary philosophy, this has been sharpened by a distinction between event types and concrete events. A concrete event is something that happens at a certain time and, at least for physical events, at a certain place. Examples of physical events are a photon hitting the table right now, the car accident yesterday, the football match last Sunday. An *event type* is an abstraction from concrete events; it is an aspect or property of concrete events that can be instantiated repeatedly.

A relation between cause and effect can be stated in different forms. There are singular causal statements, e. g. 'the hot water flowing from this tap causes the melting of the snow under my feet right now', where a concrete event is said to be the cause of a concrete effect. There are also general statements relating event types, e. g. 'Flowing of hot water causes the melting of snow'. It is controversial whether one type of causal statement is reducible to the other.

According to a causal regularity theory, the truth of a singular causal statement '*a* causes *b*' depends on events *a* and *b* instantiating, under some description, a causal regularity. A regularity is a relation between event types. Thus the analysis is shifted from singular causal statements to general ones, the latter being logically prior. A causal regularity is analysed essentially with recourse to a universal material conditional, without invoking necessity, counterfactuals, or forces.

There are two different basic patterns of regularity theories. The weaker one claims that two event types C and E constitute a causal regularity if the cause is *ceteris paribus* sufficient for the effect: under suitable circumstances all events of type C are accompanied by events of type E. The stronger one claims that the cause is not only *ceteris paribus* sufficient, but also necessary for the effect: additionally, under suitable circumstances all E 's are accompanied by C 's. While the first type allows a plurality of alternative causes, the second one does not. Following Hume, these accounts are often supplemented by the conditions of temporal priority and spatial contiguity.

Hardly anyone disputes that a regularity theory of causation has many theoretical attractions. It suits scientific language, as expressed already by Maxwell in the opening sentence of his *Matter and Motion*: "Physical Science is that department of knowledge which relates to the order of nature, or, in other words, to the regular succession of events."[1] This usage of words has survived all scientific revolutions which occurred since Maxwell's days. It satisfies empiricist intuitions since it does not invoke empirically untestable claims of necessary connections. It explains how causal knowledge serves the all-important tasks of prediction, explanation, and control of natural processes. Although causation is a prevalent concept directing a wide range of scientific activities and common life, a comprehensive account of causal relations is still at large.

It is an open task to provide an account which circumvents the many objections with which regularity theories have been confronted. It has been argued that regularity theories cannot account for the directedness or asymmetry of causal relations, that they cannot distinguish between cause and effect in a satisfactory way. If taking arsenic is a sufficient condition for dying, dying together with the absence of whatever else may cause death is sufficient for taking arsenic. But although taking arsenic is a cause of death, dying is no cause of taking arsenic.

Regularity theories typically rely on temporal order to supply the asymmetry: a cause has to take place earlier than the effect. But the concept of time stands as much in need of clarification as the concept of causation itself. It may be tempting to define the direction of time by causal order[2], but this

1 Maxwell (1991(1871)), p. 1.
2 Cf. Reichenbach (1956).

cannot be done without circularity if the direction of causation is defined by temporal order in turn. Additionally, the objections can be finessed so that temporal order still does not supply the wanted asymmetry: a sharp drop in barometer reading is regularly followed by a storm shortly afterwards. But it is no cause of the storm. Instead, both events are caused by changes in atmospheric pressure. They are epiphenomena: events that have a common cause without being causally relevant for each other.

A second challenge concerns the problem of distinguishing regular, though accidental, coincidences from causal regularities. Assume that every sighting of a certain comet was followed by a famine. If causation is nothing more than constant coincidence, how can a regularity theory avoid the conclusion that the – irrelevant – comet is the cause of the famine?

Further counter-examples concern overdetermination with two causes present at the same time, and preempted potential causes, where if the actual cause had not occurred an alternate cause would have brought about the same result (these constitute challenges only to regularity accounts of the second type which take causes to be necessary conditions for the effect).

Many writers who in general are sympathetic to an empiricist approach, among them John Mackie and David Lewis, were convinced by those counter-examples to regularity theories.[3] Lewis writes:

> It remains to be seen whether any regularity analysis can succeed in distinguishing genuine causes from effects, epiphenomena, and preempted potential causes […] I have no proof that regularity analyses are beyond repair […] Suffice it to say that prospects look dark. I think it is time to give up and try something else.[4]

Recently, some philosophers have tried to develop alternative accounts of causation, taking the alleged inviability of a regularity analysis as their main motivation for postulating some form of natural necessity.[5] Meeting the challenges to a regularity theory undercuts the motivation for this approach.

The goal of this paper is the explication of the general causal relation 'C 's cause E 's'. We claim that the main features of causal relations can be successfully analysed within the bounds of a regularity theory of the first type. Our account of a regularity differs from traditional analyses in that being a causal regularity is claimed to be a theory-dependent property: whether a true universal conditional is a causal regularity can not be judged in isolation from a system of such conditionals.

3 Cf. Mackie (1980), p. 86f, Lewis (1993), p. 194.
4 Lewis (1993), p. 194, cf. Lewis (2000) for related problems of a counterfactual causal analysis.
5 Cf. Tooley (1987), p. 67, and, although more concerned with laws of nature, Armstrong (1983), p. 73.

We proceed as follows: first, four principles are introduced which try to capture basic intuitions about causal relations. Then we develop a framework for representing causal relationships that satisfies these principles. Afterwards, the challenges are analysed and a solution is proposed.

2. Causal Principles

With the following four principles we propose a minimal core of criteria, which any theory of deterministic causation – besides its ability to meet the mentioned challenges – should satisfy. Although the first two principles are well known, their exact content is a matter of controversy. For their initial presentation it suffices to state them in general terms.

The first principle captures the idea that for deterministic systems the effect is fixed once its causes have occurred.

Principle of Causal Determinism The same cause is always accompanied by the same effect.

This principle defines the essence of deterministic causation, to which our discussion is confined. If there are *irreducible* probabilistic relations, as it is claimed to be in quantum physics, *and if* they are interpreted to be causal, our analysis does not apply to them. But it should be noted that most laws which are taken to be probabilistic are not irreducible in the sense that there are no further statistically relevant factors not included in the law statement.[6] The application of probabilistic methods should not be seen as an alternative approach to causation. The bulk of scientific investigations, even if it proceeds with statistical methods, is an inquiry of deterministic systems. For those, we claim, a regularity theory of causation is fundamental.

The second principle states that effects do not occur spontaneously:

Principle of Causality If no cause is present, no effect occurs.

This does not constitute a 'Law of Universal Causation', claiming that every event has a cause. Rather, it is claimed that if an event type is classified as a type of effect of some type of cause, then for every occurrence of this type

6 Even authors favouring a probabilistic methodology such as Spirtes, Glymour and Scheines (1993), cannot deal with truly indeterministic quantum mechanical systems. Their assumptions can only be justified strictly for what they term 'pseudo-indeterministic' systems, which are nothing but incompletely specified deterministic systems (op. cit., p. 63f, p. 52).

of effect there is an occurrence of at least one of its causes. This is something not entailed by the Principle of Causal Determinism. It also holds for probabilistic systems: even if there is no sufficient cause for the emission of α-radiation, radiation does not occur without some emitting source.

It is a strong intuition that the connection between cause and effect is more intimate than a mere coincidence: the cause brings about or produces the effect. At the bottom of this intuition lies the notion of relevance. A causal regularity covers only those aspects of a situation which are relevant for the occurrence of the effect. It does not cover accidental aspects.[7] We postulate the following principle:

Principle of Causal Relevance Every type of cause is indispensable in at least one situation.

A stronger version of this principle, that a cause should contribute something to the effect in *any* situation where it occurs, leads Paul Humphreys[8] to the claim that in cases of overdetermination, e. g. if a man's heart is hit by two bullets at the same time, none of the factors, e. g. bullet-shots, is a cause. In our opinion, this consequence should be a reason to reject the principle from which it is derived. If one bullet alone causes death, two bullets do not make it uncaused.

The Principle of Causal Relevance is connected to the idea of the simplicity of nature (see p. 10), and it is mirrored in numerous methodological maxims, such as that one should not assume more causes than necessary, or that one should prefer simpler and more general theories. It is also connected to the idea that causes are somehow operative in bringing about their effects.

The fourth principle seems less conspicuous than the other three, yet it is indispensable for a proper treatment of causation. For an effect, one can normally think up a nearly unlimited multitude of alternative causes and each of these causes would have to contain a nearly unlimited multitude of positive and negative factors. Yet, we often, if not always, state causal regularities which do not specify all factors relevant for a causal process: causal explanations are typically *incomplete*. To retain the sufficiency, a regularity statement has to be prefixed by a *ceteris paribus* clause (or some similar device): under suitable circumstances, the effect always occurs.

The relevant factors that complement the specified factors are implicitly referred to in the *ceteris paribus* clause. We expect each specified cause to maintain its relevance when a more explicit description of the situation is

7 This is especially clear in the context of causal explanation, where the inclusion of irrelevant aspects renders an offered explanation intuitively unacceptable; cf. Salmon (1984), p. 92f.
8 Humphreys (1989), p. 10.

provided. Increase in causal knowledge is, in this sense, monotonic. This idea is captured by the fourth principle:

Principle of Persistent Relevance A causal factor maintains its causal relevance when additional factors are taken into account.

This principle may seem to be rather trivial, since the *ceteris paribus* clause already refers to those additional factors implicitly. Why should it make a difference if we refer to them explicitly? The analysis will show how to exclude certain kinds of 'parasitic' regularities by appealing to this principle.

In sum, for any occurrence of a complex cause there is an occurrence of the effect, and the effect does not occur without one of its causes. A cause does only include factors which are relevant for the occurrence of the effect. If a cause is specified incompletely (as is the case most of the time), the specified factors remain relevant when further factors are taken into account. We will develop a model of causal relationships that satisfies these principles and elucidates the content of key notions used here.

3. A Theory of Causal Regularities

3.1. Complex Event Types

We have already introduced the distinction between concrete events that happen at some time and some place[9] and abstract event types. We follow the common practice in discussions about causation to use the term 'event' in a broad sense, so that it also covers processes and states. An event type E corresponds to a propositional function 'Ex', which we call a *generic event description*. A concrete event a that satisfies 'Ex' is said to instantiate event type E.[10]

A sentence containing just a single predicate, e. g. 'Fa', we call *atomic*. Sentences which are atomic or atomic sentences prefixed by a negation, e. g. 'Fa' and '$\neg Fa$', we call *simple*. Truth functions of simple sentences, e. g. '$Fa \lor Ga$' or '$Fa \land \neg Ga$', we call *complex*. Analogous to event descriptions, there can be

9 We will not deal with non physical events in this paper, although the definitions can easily be adjusted to fit non physical events (if there are any).

10 We use upper case letters 'A', 'B' etc. as predicate letters, lower case letters 'x', 'y' etc. as individual variables, and lower case letters 'a', 'b' as individual constants. '$\forall$' denotes the universal quantifier, '$\exists$' the existential quantifier, '$\rightarrow$' material conditional, '$\land$' conjunction, '$\lor$' disjunction and '$\neg$' negation. Upper case letters are also used for denoting event types. 'Iff' means 'if and only if'.

atomic, simple and complex event types. What counts as atomic is relative to a scheme of classification.

The notion of an event type is intended to explicate the idea of similarity between events. To do this, the notion of a complex event type has to be restricted to those that can be described by a conjunction of simple generic event descriptions: all events satisfying such a description are *homogeneous* with respect to all event types mentioned in the description. Their similarity consists in sharing the properties constituting the complex event type. If, on the other hand, we would allow, e. g. disjunctive or conditional event types, two concrete events could instantiate a disjunctive type consisting of event types A and B without sharing any of those simple properties, i. e. if one event instantiates type A but not B, and the other instantiates type B but not A.

A second reason for not allowing disjunctive types is connected with the Principle of Causal Determinism. A disjunctive 'state' generally does not uniquely determine the subsequent state with respect to the properties mentioned in the disjunction. A system that is in a disjunctive 'state' defined by instantiation of A or B can be, with respect to types A and B in one of the three states defined by instantiation of A and B, instantiation of A and non-B, or instantiation of non-B and A. From any of these states it may evolve into a state which differs drastically from the other states. But it is certainly not compatible with the idea of a deterministic system that two closed systems governed by the same deterministic regularities may be in the same initial state yet evolve into very different final states.

A third reason is that we can always form logically complex generic event descriptions out of the simple ones by using negation, conjunction, disjunction, and conditionals. However, any such description is equivalent to some statement in disjunctive normal form. Therefore, we can restrict our discussion to those descriptions without losing generality. Any disjunct of a sentence in disjunctive normal form is made up of a conjunction of simple event descriptions and may correspond to a complex event type. But the disjunction itself does not correspond to a complex event type. A generic description '$Ax \lor Bx$' means that an event satisfying this description is either of type A, or of type B, or both. By transforming a logically complex description into a disjunctive normal form equivalent and taking disjunctions as alternative event types we can dispense with non-conjunctive complex events.

3.2. Regularity Statements

A regularity holds between a (possibly complex) type of event ϕ and an atomic type of event ψ if and only if for every concrete event instantiating ϕ there is a different concrete event instantiating ψ, so that the second event stands in a suitable spatio-temporal relation to the first.[11] In this case we say that satisfaction of condition ϕ by a concrete event is *sufficient* for satisfaction of ψ by a second concrete event that stands in a suitable spatio-temporal relation to the first. Conversely, satisfaction of condition ψ by a concrete event in a suitable spatio-temporal region is *necessary* for satisfaction of ϕ by a second event in that region. If there is no ψ in that region, there is no ϕ either.

The logical form of a regularity statement relating e. g. two event types A and B is:

$$\forall x[Ax \rightarrow \exists y(By \wedge Rxy)], \tag{1}$$

where variables 'x' and 'y' are defined over the domain of concrete spatio-temporal events with $x \neq y$, A and B are event types, and R is a relation between events x and y demanding those events to stand in a suitable spatio-temporal relation.[12]

For stating that a regularity holds between certain event types we introduce a notational simplification: quantifiers, variables 'x' and 'y', the '$\wedge$' and relation R are omitted and the negation sign is replaced by a bar.

$$A\overline{B} \vee CD \rightarrow E =_{df} \forall x[(Ax \wedge \neg Bx) \vee (Cx \wedge Dx) \rightarrow \exists y(Ey \wedge Rxy)] \tag{2}$$

If there is not only a regularity in one direction but a similar conditional in the other direction as well, e. g.

$$(\forall x[Ax \vee Bx \rightarrow \exists y(Ey \wedge Rxy)]) \wedge (\forall y[Ey \rightarrow \exists x((Ax \vee Bx) \wedge Rxy)] \tag{3}$$

we abbreviate:

$$A \vee B \Rightarrow E. \tag{4}$$

11 Here we use 'ϕ' and 'ψ' as variables standing for event types corresponding to event descriptions ϕ and ψ.

12 For a discussion of this and similar variants compare Kim (1973), p. 229f. Kim prefers to put the relation into the antecedent, but we did not find his reasons convincing. Since our account does not rely on the exact nature of R and is compatible with different choices (e. g. with theories that forbid actions at a distance as well as with theories that allow this) we will not further discuss this relation.

For any such statement there is a unique notation in predicate logic defining its truth conditions. The '$\rightarrow$' we still call a conditional, since it corresponds to a material conditional in its expanded form. The '$\Rightarrow$' we call a *double conditional.*

To model a complex cause by a sufficient condition guarantees satisfaction of the Principle of Causal Determinism: any time the cause type is instantiated the effect type is instantiated in a suitable spatio-temporal region, too.

If an event type is sufficient for several effects (as it is often the case) this is expressed by combining several conditionals. E. g. if events of type A are regularly accompanied by events of type B and type C, this is expressed as

$$(A \rightarrow B) \wedge (A \rightarrow C). \tag{5}$$

It is intuitively unclear whether negative event types such as $\overline{E}$ can stand for effects at all. Furthermore, if they can, then if the causes of E are specified, the presence and absence of $\overline{E}$ can be deduced from this.[13] For this reason we can restrict ourselves to atomic event types in the consequent of a regularity statement.

The four principles allow more than one type of cause for a type of effect, e. g. there are numerous causes of death, or numerous alternative causes for a revolution. Some authors reject the existence of a plurality of causes. But a plurality is admitted not only in our every day concepts, but also in the laws of chemistry, where different reactions yield the same product; physiology, where converging pathways occur especially in catabolism; thermodynamics, where systems in different initial states can end up in the same equilibrium state; or mechanics, where different superpositions of forces generate the same acceleration. There is no reason to suppose that these laws are *therefore* only provisional, or that a future science will abandon plurality.

A plurality of alternative causes is represented by a disjunction of sufficient conditions:

$$\mathrm{AB}\overline{C} \vee \mathrm{FG} \rightarrow \mathrm{E}. \tag{6}$$

If any of these disjuncts is instantiated, or if several are instantiated together, the effect occurs.

If we assume that for any occurrence of an event type E there is at least one of its sufficient causes present, the disjunction of all alternative causes of an effect is a *necessary condition* for the effect.[14] This aspect is not modelled by a

13 See fn. 15.

14 Cf. Mackie (1980), p. 61f.

mere disjunction of sufficient conditions; to do this, we have to make use of the double conditional:

$$AB\bar{C} \vee FG \Rightarrow E. \qquad (7)$$

The disjunction of all individually sufficient conditions is a necessary condition for the effect, and the presence of the effect is necessary and sufficient for the presence of at least one of the causes.[15] This satisfies the Principle of Causality: 'no cause, no effect'.

3.3. Minimality and the Principle of Causal Relevance

How can the Principle of Causal Relevance be satisfied by our approach? Sensitivity to relevance marks an important difference between a causal regularity statement and a simple material conditional. A material conditional is notoriously insensitive to considerations of relevance; e. g. from '$\forall x(Ax \rightarrow Cx)$' one can derive '$\forall x(Ax \wedge Bx \rightarrow Cx)$', whatever '$Bx$' stands for. Supplementing the regularity concept to capture these relevance conditions is crucial.

We claim that these conditions can be incorporated with the help of certain minimality constraints imposed on sufficient conditions. The first minimality constraint is one that is often acknowledged to play an important role in the concept of causation. A sufficient condition can only be a complex cause if it is minimal.[16]

Minimal sufficient condition A conjunctive sufficient condition ϕ of ψ is a *minimal sufficient condition*, iff no proper part of it is sufficient.

'Proper part' means a conjunction ϕ' with fewer conjuncts than ϕ where every conjunct in ϕ' is also a conjunct in ϕ. Example: if FG is a sufficient condition of W, then FGH is not a minimal sufficient condition of W. If it is not minimal it contains a redundancy. If ϕ consists of a single factor we define its proper part to be the empty antecedent, i. e. '$\forall x(Fx \rightarrow Gx)$' reduces to '$\forall x(Gx)$'. '$Fx$' is a minimal sufficient condition only if 'Gx' is not universally satisfied.

15 If we have a necessary and sufficient condition for E, we have also a necessary and sufficient condition for $\bar{E}$: we only have to negate the left hand side of the double conditional.

16 Mackie (1980), p. 62. The idea of minimal conditions can be traced back at least to Broad (1930), p. 131, and Johnson (1963 (1924)), p. 64.

The notion of a minimal sufficient condition already allows one to deal with certain types of irrelevancies:

Case I: If taking a strong narcotic is sufficient for losing consciousness, taking this narcotic while reading a book is sufficient, too. But since taking the narcotic is sufficient in itself, a condition containing both taking a narcotic and reading a book is not minimal. Therefore, the causal regularity is constituted by taking the narcotic alone. Relative to this regularity, reading a book is irrelevant.[17]

The idea of simplicity involved here is not only aesthetic or one required by an 'economy of thought'. Assume we want to bring about a desirable effect, e. g. recovery from an illness, by realising one of its causes. Let treatment with drug X be a minimal sufficient condition; then treatment with drugs X and Y is also a sufficient condition, although not minimal; and let treatment with Y alone not be a sufficient condition. Assume that administering X is unproblematic but administering Y has very undesirable side effects. Under this condition, treatment with X and Y is irresponsible, since X is sufficient in itself. Or assume that X is available while Y is not. If we do not realise that X is sufficient in itself this could prevent us from applying this therapy at all. Knowledge of the *minimal* condition is crucial for a successful therapy. In general, when we plan to bring about some event, we should carefully consider which factors *have* to be present to get the effect and which may be present or not, without influencing the result.

The Principle of Causal Relevance is required a second time. No regularity should contain redundant factors. But regularities which are redundant as a whole are not admitted, too.

Case II: There are several clusters of causes leading to a famine, among them droughts, earthquakes, plant diseases, deliberate crop destruction, sieges. Since any famine can be traced back to such a cluster of unfavourable circumstances, there is no room for cosmological causes such as the presence of a comet – even if the presence of this comet has been regularly followed by a famine.

It is not a causal regularity for reason analogous to the one mentioned before: the regularity as a whole is superfluous, violating the Principle of Causal Relevance.

17 A factor which is a redundant part of a minimal sufficient condition ϕ of ψ need not be totally irrelevant for the effect; e. g. it may be an indirect cause, or itself be caused by ψ. But it does not operate with the other factors in ϕ to bring it about.

This requirement is modelled by making the disjunction of sufficient conditions subject to a minimality constraint, too: it has to be a *minimal necessary disjunction of minimal sufficient conditions.*

Minimal necessary condition A disjunctive necessary condition is called *a minimal necessary condition* iff no proper part of it is necessary.

'Proper part' means a disjunction ϕ' containing fewer disjuncts than ϕ where every disjunct in ϕ' is a disjunct in ϕ. Example: if '$F \vee G$' expresses a necessary condition of an event described by W, '$F \vee G \vee H$' does not express a minimal necessary condition. If the necessary disjunction consists of one disjunct only, e. g. '$\forall x(\neg Fx \rightarrow \neg Gx)$' we define its proper part to be the empty antecedent, e. g. '$\forall x\ (\neg Gx)$'. 'Fx' expresses a necessary condition of 'Gx' only if 'Gx' is not empty; otherwise anything would be necessary.

Minimal theory A minimal necessary condition of E whose disjuncts are minimal sufficient conditions is called a *minimal theory* of E.

To our knowledge, this condition has not been recognised in the literature. Mackie (1963), p. 35, requires a necessary disjunction to consist of 'all the minimal sufficient conditions', i. e. it is a *maximal* set, whereas we require a *minimal* set, which is the opposite. Had Mackie intended a disjunction to be minimal, his discussion of several counterexamples (which we discuss below) would have to be different since the minimality requirement has a great impact on the logic of the case and cannot be ignored. Mackie did not seem to be aware of the problem of redundant disjuncts. Kim (1993), p. 67, considers, but does not endorse the requirement that a full cause has to be in non-redundant disjunctive normal form. This requirement has some analogies with minimal theories but is not identical. Disjunctive normal forms are defined for propositional logic; it is a purely syntactical notion. In contrast, the concept of minimal theories is defined in predicate logic and minimality depends on the extension of the predicates. $AB \vee CD \vee EF$ is a non-redundant disjunctive normal form but the corresponding minimal theory may or may not be redundant, depending on the extensions of the predicates. v. Wright (1952), p. 71, introduces the notion of a *smallest necessary condition*; but the disjuncts of such a condition are not required to be sufficient conditions. They are defined analogous to conjunctive normal form; the same remark applies to Broad (1944).

The notion of a minimal theory is crucial for characterising causal regularities. The following theorems, which are summarised here, are discussed more fully in the appendix.

The most important property is that any minimal sufficient condition contained in a minimal theory T is instantiated at least once when none of the other minimal sufficient conditions is instantiated. This guarantees that every minimal sufficient condition, as well as the conditioned event type, is not empty. Additionally, it guarantees that no disjunction in T is sufficient for any other disjunction in T (except for the trivial case that one disjunction is a part of the other). The minimal sufficient conditions in T are, in this sense, logically and empirically independent.

These properties, together with the properties of minimal sufficient conditions, guarantee that minimal theories do not contain any irrelevancies or redundancies. Since any regularity is instantiated at least once when none of the alternative regularities is instantiated, it does an indispensable job in bringing about the effect in at least one situation. In at least one situation it is the only regularity that can account for the effect, and some cause *must* account for it. Additionally, no regularity contains superfluous elements. If a factor is deleted from a regularity, there is a situation where all other factors are present but the effect does not occur. Therefore, the Principle of Causal Relevance is satisfied. For this, the comet must have been present at least once when none of the other sufficient causes of famine was present. Whether a minimal sufficient condition C of E qualifies as a cause of E does not only depend on C itself but also on the question which other sufficient conditions for E exist. C only qualifies as cause if it is a member of a minimal set.

3.4. Causal Order

The Principle of Causal Determinism and the Principle of Causality operate together to ensure that causally related event types always occur in pairs: any time the effect is present, one of the causes has occurred, and any time one of the causes is present, the effect occurs. Therefore, presence and absence of the effect is completely determined by the presence and absence of its causes.

Note however, that this is not a one-to-one, but a many-to-one correspondence: as we have noted, many different causes can bring about the same effect; the principles allow a plurality of causes, e. g. there are different kinds of force that can set a body in motion. The presence of the cause uniquely determines the effect, but the presence of the effect does not uniquely determine the cause.

This constitutes a conceptual asymmetry between causes and effects: any complex cause is conceptually required to be sufficient, but the effect is not required to be sufficient for any of its complex causes. It is only required that at least one of the alternative complex causes is instantiated. This fact can be used for defining the asymmetry between cause and effect without relying on temporal order: the direction of causation is equated with the direction of determination. Causal order has not to be stipulated explicitly; it emerges from the interdependencies of the principles.

The following example seems to pose difficulties for our account: if taking arsenic is sufficient for dying, dying in the absence of whatever else may cause death is sufficient for taking arsenic. Isn't it always possible to reverse causal order by reversing the double conditional?

For simplicity, let us assume only two causes of death (D): taking arsenic (A) and taking strychnine (S):

$$A \vee S \Rightarrow D. \tag{8}$$

In the above formula, there are two minimal sufficient conditions for D whose disjunction is a necessary condition of D. So, for every occurrence of A the effect D follows, and for every occurrence of S, D follows. The causes determine the effect. But the effect does not determine which of the causes is present. From the mere fact that Joe is dead the minimal theory does not allow to deduce whether he has taken arsenic or strychnine or both.

There are some minimal sufficient conditions, e. g. if Joe is dead but has not taken arsenic, he has taken strychnine.

$$D\bar{A} \rightarrow S. \tag{9}$$

The possibility of forming such minimal sufficient conditions is one of Mackie's reasons to reject the possibility of defining causal order via the direction of determination.[18] The rejection is valid for an account that -- such as Mackie's -- analyses a regularity as a minimal sufficient condition *simpliciter*. Yet, as we have seen, in our analysis a minimal sufficient condition qualifies as a cause only if it is part of a minimal *theory*, which adds additional constraints. These constraints supply the resources for dealing with the critical examples. Although (9) is a minimal sufficient condition, it cannot be transformed into a minimal theory, since

$$D\bar{A} \Rightarrow S \tag{10}$$

18 Mackie (1980), p.160f.

is false for a situation in which a person has taken arsenic and strychnine at the same time. Presence and absence of S is not fully determined by D and A. This can also be expressed by saying that A and S are independent variables while D is dependent.

Mackie has two further objections against the equation of causal order and the direction of determination. The first one applies to indeterministic systems and is not relevant for our context. The second one argues that there is not only a plurality of causes, but also of effects (*op. cit.*, p. 167). Again, this poses no problem for our account. A complex cause may be sufficient for any number of effects; but then it determines any of these effects, while any of them may have several alternative causes. The difference is that the set of effects of a complex cause form a conjunction, whereas the set of alternative causes of an effect form a disjunction.

3.5. Persistent Relevance

Typically, if not always, theories are *incomplete*. There are other factors relevant for those causal interactions. It is obvious that the number of relevant factors is normally huge. Although we often know that a causal theory omits relevant factors, we need an economical construction of causal theories with a limited set of causal factors and few regularities. This commits us to partial theories, even if our knowledge would allow the construction of larger theories with many more factors and regularities.

It is one of the intriguing features of causality that, despite the numerous factors, we can establish causal regularities even if most of the relevant factors are neglected or even unknown. How is this achieved? The standard additional qualification of a causal regularity adds a *ceteris paribus* clause -- that the regularity holds under *suitable conditions*. However, *ceteris paribus* clauses give the impression that causal regularities are vague, ill-defined and open to arbitrary immunisation attempts, once an instance of apparent falsification has been brought up against the claims of a causal theory.

An additional constraint has to be imposed on the relation of an incomplete causal theory to its possible expansions. Consider what happens when we enlarge our view of a causal process by considering additional factors.

Case III: A contraption produces yoghurt under certain conditions. Most of those conditions are controlled according to the set of factors which are part of the causal theory about the fermentation process. These regularities state that *ceteris paribus* if those conditions are given and other blocking factors are prohibited, the fermentation takes place. Now, despite a careful control of all factors given by the theory, occasionally the fermentation process leads to unwanted results. We would then search for other factors relevant to the fermentation process and would include them into the theory. By doing so we can control the fermentation condition by an extended set of factors.

The inclusion of hitherto unconsidered factors leads to an expansion of the set of factors in a theory. It is not to be expected that a firmly established causal relationship has to be dropped in the expanded theory, nor that a cause is suddenly turned into an effect. If this happens, we rather assume that the theory was incorrect all the time. It is a crucial condition for a genuine causally relevant factor that its causal relevance is maintained under any expansion of the theory by any factors, as demanded by the Principle of Persistent Relevance.

The principle shows why one can systematically operate with incomplete causal theories. Even if we do not know the whole story, we know some essential part of it which will not be superseded by further information. If additional causal factors are taken into account, they can not invalidate the relevance of the factors discovered so far. Conversely, if it invalidates the relevance of some factors, our intuition is clear: such regularities are artefacts, or pseudo-regularities.

3.6. A Definition of Causal Regularities

The relations between a theory, *ceteris paribus* clauses and the possibility of a monotonic increase of causal knowledge by expanding a theory are intimately connected to our very concept of causal relevance.

Causal relevance Factor C is causally relevant for factor E iff for any set of factors that contains C and E there is a minimal theory of E that contains C.

If C is causally relevant for E relative to a theory T, then C remains causally relevant for E under every expansion of T. Causal relevancy is a theory-invariant property, being conserved under enlargement of the set of factors taken into consideration. The notion of a causal regularity is defined with the help of the notion of causal relevance:

Causal regularity $\ulcorner \phi \rightarrow \psi \urcorner$ expresses a causal regularity relative to a minimal theory T of ψ iff
(1) ϕ is a minimal sufficient condition in T,
(2) all event types contained in ϕ are causally relevant for ψ.

The second clause in the above definition excludes regularities which are true relative to some restricted set of factors but where the relevance ceases when the set of factors is enlarged. The inclusion of more factors allows one to find situations which differentiate between two alternative causal theories. Note that being a causal regularity is not a theory-invariant property. Something is only a causal regularity if it is a *ceteris paribus* sufficient condition. But if additional relevant factors are taken into account the condition has to be supplemented by them.

4. Meeting the Challenges

4.1. Epiphenomena

Epiphenomena occur whenever two events have a common causal factor without being causally relevant for each other. Mackie (1980) discusses an example of epiphenomena on which he sees a pure form of a regularity analysis fail:

> The sounding of factory hooters in Manchester [at 5 p.m.] may be regularly followed by, but does not cause, London workers leaving their work.[19]

This case is especially interesting because at least two authors who give an account of causation in terms of necessary and sufficient conditions consider this to be a decisive counter-example.[20]

Let us make the logical structure of the example explicit. Any time the Manchester hooters sound (H_M) and it is 5 p.m. (F), Manchester workers leave their work (L_M):

$$H_M \wedge F \rightarrow L_M . \quad (11)$$

19 Mackie (1980), p. 81.
20 The second author is Broad.

Additionally, there are numerous other causes for Manchester workers leaving their work when it is not 5 p.m., or when the hooters do not sound: a fire, a strike etc. For simplicity, we consider only one factor (C_M):

$$C_M \rightarrow L_M . \tag{12}$$

Any time Manchester workers leave, one of the causes specified in (11) or (12) is instantiated. This is represented as a minimal theory:

$$C_M \vee (H_M \wedge F) \Rightarrow L_M . \tag{13}$$

We assume that (13) is a minimal theory for Manchester workers leaving: satisfaction of C_M is minimally sufficient for satisfaction of L_M; satisfaction of $H_M \wedge F$ is minimally sufficient for satisfaction of L_M; satisfaction of their disjunction is necessary for satisfaction of L_M, so that Manchester workers never leave when none of these two sufficient conditions is satisfied.

An exactly analogous description holds for London workers leaving their work (L_L). They leave their work if and only if it is 5 p.m. (F) and London hooters sound (H_L), or some other cause such as fire is present (C_L).

$$C_L \vee (H_L \wedge F) \Rightarrow L_L . \tag{14}$$

From (13) and (14) it can be deduced that any time C_M is not satisfied but Manchester workers leave their work (L_M) and London hooters sound (H_L), London workers leave their work (L_L):

$$\overline{C_M} \wedge L_M \wedge H_L \rightarrow L_L . \tag{15}$$

Any time Manchester workers leave their work, one of its causes must be present; if it is not C_M, it must be $H_M \wedge F$. If in the latter situation additionally the hooters in London sound, London's workers leave their work, since both F and H_L are present. The left side of the conditional (15) is a sufficient condition. And no part of it is sufficient, since $\overline{C_M}$ and L_M are both needed to guarantee that it is 5 p.m.

This leads to the conclusion that Manchester workers leaving their work is a partial cause of London workers leaving their work, if a cause is identified with a minimal sufficient condition.[21] This is the point where Mackie concludes that a pure regularity theory cannot be the whole story about causation.[22]

21 If temporal constraints are added so that the cause has to be earlier than the effect, we can assume that Manchester workers leave an instant before London workers, so that the example still works.

We will show how this puzzle can be solved with the help of the notion of a minimal theory. A minimal sufficient condition of E is only a cause of E if it is part of a minimal theory of E. Of which minimal theory is the left hand side of (15) a part? Can it be simply attached to (14)?

$$(\overline{C_M} \wedge L_M \wedge H_L) \vee (H_L \wedge F) \vee C_L \Rightarrow L_L . \tag{16}$$

A theory is only minimal if there is no shorter disjunction that is a necessary condition for the effect. But we can leave out the first disjunct and still have a necessary disjunction (this shorter disjunction is (14); it is *ex hypothesi* necessary). The spurious minimal sufficient condition is not, according to our definition, a causal regularity : the very fact that it is spurious prevents us from attaching it to the minimal theory.

In general, from a minimal theory we can always derive further true conditionals. But the very fact that they are logical consequences of the minimal theory prevents us from classifying them as *causal regularities*. All instances of a derived conditional are also instances of some causal regularities contained in the minimal theory from which they are derived. To classify a sequence of events as causally related, it is enough that they instantiate *one* causal regularity. Nothing is gained if the causal regularities they instantiate are multiplied.

But, it may be objected, how do we know that the spurious condition is not part of a different minimal theory of L_L other than (14)? How do we know that there is not more than one minimal theory?

Let us check if we can, while sticking to the sequences of events allowed by (13) and (14), omit any of the disjuncts in (14) and add the spurious condition instead. Can we construct an alternative minimal theory this way? That we can not eliminate C_L is obvious. Otherwise the double conditional is falsified by any situation in which London workers leave their work when it is not 5 p.m., or when London hooters do not sound.

But if we leave out $F \wedge H_L$ there are also situations in which L_L is satisfied but the disjunction is false: any situation in which (i) London workers leave their work (L_L) and (ii) there is no fire, strike etc. in London ($\overline{C_L}$), and (iii) the hooters in Manchester do not sound ($\overline{H_M}$).

In such a situation either C_M is satisfied so that Manchester workers leave their work, or Manchester workers continue to work so that L_M is not satisfied. In any case both disjuncts in

22 Mackie (1980), p. 86.

$$C_L \vee (\overline{C_M} \wedge L_M \wedge H_L) \Rightarrow L_L \qquad (17)$$

are false, but London workers leave their work. Therefore, (17) is false. The only minimal theory of L_L is (14), which is the intuitively correct set of causal regularities.

This solution is quite general. The problem arises in the following way: there is a common factor C of two effects E_1 and E_2. In such cases one can always form a minimal sufficient condition for E_2 consisting of the absence of all alternative causes of E_1 that do not contain C, the presence of E_1 and the presence of the factors that are together with C sufficient for E_2. But this condition is not part of a minimal theory. This does not depend on the number of alternative causes, or the complexity of the conjunctions, as long as these factors are not logically dependent.

4.2. Overdetermination

The problem of overdetermination stems from the possibility of a plurality of causes. A neurone transmits a signal if a certain threshold-value is exceeded. But it is possible that this neurone is activated by two different neurones where each activation in itself is sufficient to activate this neurone. A light bulb is wired in parallel. Setting on each of two switches is sufficient for there being light. If both switches are on, the effect is overdetermined. If the situation is symmetric, there is no reason to call one of the switches a cause but not the other. But since light does not occur spontaneously, there has to be a cause. Any theory of causation has to explain how this is possible.

A regularity account faces no problems in providing an explanation. A concrete event *a* causes a second concrete event *b* if there is a causal regularity whose antecedent is instantiated by *a* and whose consequent is instantiated by *b*. If there is a further event instantiating a regularity with the same consequent in the occurrence in question, the effect is overdetermined. In our example each switch is a cause although neither is *the* cause.

4.3. Preempted Potential Causes

The solution of the problem of preempted potential causes is equally straight-forward. Assume that there are two regularity statements '$AB \rightarrow C$' and '$\overline{A}D \rightarrow C$'. Assume further that in a concrete situation the complex

event type ABD is instantiated. In this situation the first regularity is instantiated, but not the second (since A is present). An event of type C follows. If A had not been present, the first regularity would not have been instantiated. Nevertheless, in this case C would still have been instantiated since in this situation the second regularity would have been instantiated. $\overline{A}D$ is a preempted potential cause which would have been instantiated if the first cause had been absent due to the absence of A.

This situation is not as artificial as one might think. In metabolism there are many alternative pathways on which a certain product is formed, but where one pathway is only active if the others are inactive. In the human brain certain activities are controlled by certain regions of the brain; but in case (and only in case) of damage to this region certain other regions take control over these processes. Similar mechanisms are exploited in technology.

4.4. Uninstantiated Regularities

If we knew all events that ever happened, happen or will happen, and wanted to give a minimal description of the regularities holding between them, there would be no use for uninstantiated regularities. Either an uninstantiated regularity is a logical consequence of some statements contained in our description; then it is redundant and should not be included in a minimal description. Or it is an axiom; but since it does not cover any fact, we could exclude it from our description and could still describe all the facts with one axiom less, resulting in a simpler description.
That we sometimes include regularities in our descriptions of the world for which we do not know any positive instances results from our obviously limited knowledge. For all we know, such a regularity may be instantiated tomorrow, and including it in our description may make our description more comprehensive.

An uninstantiated regularity may be derived from instantiated regularities and other facts. So we have to admit that they are true universal material conditionals. To deny this would be to deny the regularities or facts from which they are derived. But not every logical consequence of a causal regularity is a causal regularity itself (otherwise, if '$A \rightarrow C$' expresses a causal regularity, '$AB \rightarrow C$' would also, no matter what 'B' stands for). There is no need to count an uninstantiated derived regularity as a causal regularity, since the type of situation to which it applies is obviously already in the scope of a causal regularity. It is only required that for any concrete event that is to be classified as causally related to a second event there is *some* true description of it under which it instantiates a causal regularity. It is not required that it instantiates a regularity under the description we have actually used for referring to it.[23]

In any case there is no room for uninstantiated basic regularities. If we knew all the facts, they would have no place either as axioms, since they do not describe anything, nor as theorems, since per definition they cannot be derived from the axioms.[24]

This is well-captured by our definition of minimal theories since a minimal theory cannot contain uninstantiated regularities. If G is never instantiated, the conditional statement '$G \rightarrow E$' is trivially true. Do we have to admit that 'If a witch is present, evil things happen' states a causal regularity, since there are no witches? No – since there are no witches, the presence of a witch is not in the minimal theory for the happening of evil things, and therefore no causal regularity. For every evil thing that happens there is already a natural cause. The appeal to witchcraft violates the minimality constraint. For the same reason 'If a witch is present, good things happen' does not state a causal regularity.

4.5. Perfect Regularities

4.5.1 *Perfect regularities and persistent relevance*

In philosophy of science, nobody has ever been able to come up with a law that does not have to be prefixed by a *ceteris paribus* clause (Cartwright 1983). Regularity theories have been criticised by authors such as Anscombe[25] for the reason that we do not know any exceptionless generalisation. We take this point very seriously: Our causal generalisations contain a *ceteris paribus* clause that cannot simply be ignored for the purpose of a theoretical analysis. Understanding the *ceteris paribus* clause is crucial for the understanding of causation itself. We hold that the openness of causal generalisations enters into its truth conditions and try to provide an analysis by way of the notion of persistent relevance. Without *ceteris paribus* clause we were never justified in applying (deterministic) causal concepts. This very fact has been one of our

23 Cf. Davidson (1993).

24 Acceptance of uninstantiated basic regularities seems to commit one to platonic realism with respect to natural laws, a position recently revived by Michael Tooley (1987), p. 119. For a detailed critique of such an account, see Van Fraassen (1989), chap. 5. There is no room to discuss functional laws such as $F = ma$ covering causal regularities. Our solution treats those laws as simplifications of conjunctions of causal regularities. Missing values then still constitute no causal regularities though they are derived from the functional law.

25 Anscombe (1993).

main motivations to incorporate the causal persistency condition in our account.

This condition is of crucial importance when discussing perfect correlations. If we look exclusively at the relation between atmospheric pressure, the barometer, and the storm and exclude malfunctions etc. of the barometer by a *ceteris paribus* clause, we may get the result that a storm occurs if and only if the barometer reading has dropped.[26] The drop in reading would be a necessary and sufficient condition for the storm.

Expanding the set of factors for the drop of barometer reading allows to include factors which might be alternatively causally relevant for a change of weather. Knowledge of these alternative causes allows the manipulation of circumstances and devices in such a way that e. g. a drop of barometer reading occurs without a correlated change of weather. The previously perfect correlation would break down. We will finally get a structure similar to the Hooters-example, so that we can apply the minimisation argument developed there. The relation between barometer reading and storm may seem to constitute a causal regularity, but it violates the fourth principle, the Principle of Persistent Relevance.

Most descriptions of perfect accidental correlations falsely suggest a problem by giving an incomplete description of our causal situation. Once such a more complete description is provided the correlations cease to be perfect. So the mistake of these alleged counter-examples is the tacit drop of the *ceteris paribus* clause, which may be omitted for certain simplifying purposes, but which must not be omitted when discussing perfect correlations. Hence this type of example is just a formal artefact.

4.5.2 *Accidental regularities and small extension predicates*

When and only when a certain type of hydrogen bomb was used there was a red label on it. So did the red label cause the thermonuclear fusion process? Besides the first line of argument there is another strategy to deal with such cases.

The second line of argument starts with the observation that typical examples for perfect correlations present cases with a small extension (number of instances) due to the aspects with which the event classes are defined. In the case of the hydrogen bomb one analyses a situation where only relatively

26 If we had omitted factors C_M, C_L, H_M, H_L from the Hooters-example, it would have been simplified to a structure isomorphic to the present example.

few such bombs exist. For this limited set of cases the perfect correlation may hold. But usage of an H-bomb is just a special case of initiating a thermonuclear fusion process; a more general description can be given in terms of collisions of nuclei under suitable conditions of density and temperature. A perfect correlation of fusion processes in general and red labels does of course not exist. Once the more general description of the cause is available, the prima facie cause mentioning the H-bomb becomes redundant, since it is merely a special case. Thereby the red labels become redundant, too. But our notion of a minimal theory prohibits such redundancies. The truth of the singular causal statement 'the detonation of this H-bomb caused a fusion reaction' is guaranteed by the more general regularity which does not specifically mention H bombs, since a detonation of an H-bomb is a case of a collision of nuclei under suitable conditions of density and temperature. The notion of a minimal theory allows us to be selective in counting a regularity as causal.

This strategy gives good prospects to show every plausible prima facie perfect correlation either to be *ceteris paribus* only or to be a special case. In the end this has to be decided on a case by case analysis of plausible examples.

4.5.3 *Residual Cases*

Yet, some readers might still insist, there might be such a strange world (not our world, maybe, just a *possible* world) that the minimalisation requirement is not uniquely solvable. In this case there is not only one minimal theory but several. We could block the counter-examples in a principled manner by requiring that M is a causal theory iff it is a *unique* persistent minimal theory. In this case, the cases such as the one discussed would not allow a causal attribution. If there is no unique persistent minimal theory, our account exhibits this strange situation and provides a criterion which diagnoses undecidable perfect correlations. It also indicates correlations which are either spurious or a genuine causal regularity. An approach which is fully in the spirit of the present account would be to require a causal theory to be a *member of a minimal set* of minimal theories. This further minimisation requirement could then be used for identifying spurious minimal theories. The possibility to identify such a problematic situation is in itself an improvement, and a feature which is missing in most other accounts, which have to count both regularities as causal.

However, instead of complicating our account with the help of further minimisation requirements, at this point we want to change our strategy.

Critics of the regularity account like to take perfect correlations as a strong motivation for their own account with respect to a regularity account. But upon inspection we find that if this counter-examples were to be admitted, analogous examples can be constructed against its main competitors also, be they deterministic or probabilistic, extensional, physically necessitarian, counterfactual, or interventionistic.[27] So this type of example could not be used for arguing in favour of one type of theory against the others.

a) statistical accounts (non-modal) Firstly, statistical accounts that do not rely on modality or counterfactuals. Salmon discusses the problem of perfect correlations under the heading of a 'perfect fork', concluding that "when the probabilities take on the limiting values, it is impossible to tell from the statistical relationships alone whether the fork should be considered interactive or conjunctive" (Salmon 1984, p. 177f). He simply excludes such cases from consideration. As we will argue below, we agree with Salmon that this type of example *should* be excluded from consideration. But once it is excluded, it can of course not be invoked as a counter-example against competing accounts. Cartwright (1984), p. 32, includes an ad hoc condition to exclude perfect correlations (although her account, as she emphasises herself, cannot offer a non-circular definition of causation anyway). This point can be generalised to all statistical accounts that do not rely on modality.

b) statistical accounts (modal) Secondly, statistical accounts that explicitly invoke modality. The difference between a regularity and a modal account is that the latter requires a kind of nomological necessity for causal relationships: A does not only raise the probability of B, but it *has to be* this way. A notion of nomological necessity can of course not remain unanalysed, if one wants to answer Hume's challenge (although authors such as Armstrong and Tooley try to persuade their readers that one can safely take it as primitive). But this is not our point here. However defined, the reader might think that modality is exactly what is needed to deal with such cases. But this is not the case. Paul Humphreys' account is an example for a modal statistical theory: Here we find that Humphreys simply admits that his theory cannot deal with cases where a common cause C is both 'nomologically necessary and sufficient' for both A and B (Humphreys 1987, p. 81f). This gives us a general strategy how to construct counter-examples against a modal account: just add the modal element to the perfect corre-

27 That is, accounts that take logical or statistical concepts as basic. We do not discuss mechanistic accounts such as the conserved quantity view, which are of a very different kind.

lation. Once this is done, those accounts cannot deal with perfect correlations.

c) counterfactuals Thirdly, counterfactual accounts. These come in two varieties. The first type fixes the truth conditions of counterfactuals with respect to natural laws. If it is a law, that *A* happens iff *C* happens and *B* happens iff *C* happens, then a situation in which *A* happens but *B* does not happen would violate the laws of nature. Hence, the counterfactual "if *B* had not happened, *A* had not happened" would be true, indicating a causal relationship where *ex hypothesi* none exists.[28]

David Lewis is aware of this and heroically opts for the second variant: he simply denies the truth of the counterfactual (Lewis 1993, p. 203). For Lewis, the truth conditions of a counterfactual statement are evaluated with respect to a nearest possible world. This, in Lewis' theory, need not be a world in which the same laws as in our world hold. So Lewis' escape is to claim that the nearest possible worlds are not such that the law *L* holds – in which case his analysis would fail – but simply of a different type where *L* does not hold, so that *A* happens but *B* does not happen.

But Lewis answers by presenting just one model where the nearest worlds are such that the law does not hold. He would have to show that for any possible world *W* containing a perfect-correlation-law *L* its nearest worlds are such that this law *L* does not hold. Otherwise we could pick a pair of worlds *W* and *W'*, such that *W'* is among the nearest worlds for *W* and the perfect-correlation law *L* holds both in *W* and *W'*.

So Lewis would have to argue that for every such conceivable case the nearest world is one in which the law does not hold. However, he fails to present an argument in support, and we cannot see any either. We could insist that the perfect-correlation law in question is the fundamental law of this universe, its grand unifying theory. Abandoning this law would be like abandoning the law of gravity for our universe. And a world with a different law of gravity would certainly be farther away from ours than one with just some different facts (if the law is instantiated many times, a lot of facts will be very different). A further drawback is that following this analysis, the notion of 'similarity' becomes rather arbitrary.

d) interventionistic accounts Recently, several varieties of interventionistic accounts have gained ground (e. g. Pearl 1999). Having dealt with statistical and counterfactual accounts, interventionistic accounts are quite easily dealt with. Here, an intervention is a possibly counterfactual disturbance of a causal system. It is crucial to note that any intervention needs a handle to disturb the system; such a handle is nothing but an additional cause: if the experimenter in an ideal experiment intervenes in a system, he can do so only because he triggers a cause which has possibly been left out of the model. In the model, this handle is implicitly set constant by the *ceteris paribus* clause. So a non-modal interventionistic account has to deny a perfect correlation between cause and effect, since in a perfect correlation these handles do not exist. A counterfactual account again founders if we bestow nomological necessity and sufficiency upon the perfect correlation: here the laws of nature forbid the existence of handles. Again the argument that there may exist a world nearest to *W* in which this law does not hold (allowing for a handle). As has been shown in the last section, one would have to argue for the much stronger claim that for every world *W* containing a nomologically perfect correlation law *L*, all its nearest worlds *W'*, *W''*, … are such that *L* does not hold in them. Once we allow to construct arguments on the basis of fictitious, purely formally defined possible worlds, such a position seems to be indefensible.

So far, we have argued that perfect correlations do not offer a valid reason to reject a regularity theory in favour of statistical, counterfactual, or interventionistic accounts (or any combination of these), since if there were a counter-example, analogous examples are easily constructed against all these accounts. But the point we want to emphasise is a different one. Regularity theories, statistical and counterfactual accounts are the main competitors for a theory of causality. The debate is about the ontological status of their respective components and about definitional priority. If all these theories cannot deal with a putative counter-example with merely possible situations, this should make us suspicious: we should have a closer look at the alleged counter-example. It turns out that the counter examples we presented and which are levelled against regularity theories are purely formal examples of the type 'imagine that *A* and *B* are perfectly correlated'. It is not specified what '*A*' or '*B*' stand for. If it is specified, we usually find that the example is just a simplification, as in the barometer case.

Summing up, we find that our regularity account is perfectly apt to deal with the plausible examples of perfect correlations, and that the residual cases are a) ill-devised and b) if they were admitted, could also be constructed in analogous ways against its main competitors, giving – contrary to popular belief – no reason to reject a regularity theory in favour of one of these accounts.

28 Incidentally, Mackie (1980), p. 86, has shown that a counterfactual account also founders on the Hooter's example.

5. Appendix: Logical Properties of Minimal Theories

In this appendix we prove some important properties of minimal necessary disjunctions. This notion is central for our treatment of causal regularities. To our knowledge, no attention has been paid to the following formal properties in the discussion about causation.

ψ and γ are metavariables which take first-predicate-logic formulas as their values, x is a metavariable for individual variables. An expression ψx means that ψ contains at least one free occurrence of variable x. The logical constants have their usual predicate-logical meaning.

Theorem 1 A disjunctive minimal sufficient condition γ for ψ cannot have as a part a (one or n place) disjunction γ_1 which is sufficient for a second (one or n-place) disjunction γ_2 in γ (unless γ_1 is itself a part of γ_2).

Proof. The following is a theorem of predicate logic:

$$\ulcorner \forall x(\psi x \rightarrow \gamma_1 x \vee \gamma_2 x \vee \ldots \vee \gamma_n) \wedge \forall x(\gamma_1 x \rightarrow \gamma_2 x) \vdash \forall x(\psi x \rightarrow \gamma_2 x \vee \ldots \vee \gamma_n) \urcorner$$

If in a necessary condition $\ulcorner \gamma_1 x \vee \gamma_2 x \vee \ldots \vee \gamma_n \urcorner$ for ψ the disjunction γ_1 is sufficient for a second disjunction γ_2, $\ulcorner \gamma_2 x \vee \ldots \vee \gamma_n \urcorner$ taken for itself is a necessary condition for ψ; therefore, $\ulcorner \gamma_1 x \vee \gamma_2 x \vee \ldots \vee \gamma_n \urcorner$ is not minimal.

In other words: if γ_1 is sufficient for γ_2, then γ_1 can be eliminated from the necessary condition and the result is still a necessary condition. In a minimal necessary condition the different disjuncts are neither empirically nor logically contained in each other.

Theorem 2 If the disjunction $\ulcorner \gamma_1 \vee \ldots \vee \gamma_i \vee \ldots \vee \gamma_n \urcorner$ is minimal necessary for ψ, and $\ulcorner \gamma_1 \vee \ldots \vee \neg\gamma_i \vee \ldots \vee \gamma_n \urcorner$ is an arbitrary disjunction resulting from negating exactly one disjunct γ_i, $1 \leq i \leq n$, in the first formula, then there is at least one object satisfying $\ulcorner \neg\gamma_1 \wedge \ldots \wedge \gamma_i \wedge \ldots \wedge \neg\gamma_n \urcorner$.

Proof. Let $\ulcorner \gamma_1 x \vee \ldots \vee \gamma_i x \vee \ldots \vee \gamma_n x \urcorner$ be minimal necessary for ψx. Let $\ulcorner \gamma_1 x \vee \ldots \vee \gamma_{i-1} x \vee \gamma_{i+1} x \vee \ldots \vee \gamma_n x \urcorner$ be the propositional function resulting from elimination of $\gamma_i x$ in the first formula. From theorem 1 it follows that $\gamma_i x$ is not sufficient for $\ulcorner \gamma_1 x \vee \ldots \vee \gamma_{i-1} x \vee \gamma_{i+1} x \vee \ldots \vee \gamma_n x \urcorner$, i. e. $\ulcorner \neg\forall x(\gamma_i x \rightarrow \gamma_1 x \vee \ldots \vee \gamma_{i-1} x \vee \gamma_{i+1} x \vee \ldots \vee \gamma_n x) \urcorner$. But this is logically equivalent to $\ulcorner \exists x(\neg\gamma_1 x \wedge \ldots \wedge \gamma_i x \wedge \ldots \wedge \neg\gamma_n x) \urcorner$.

Since this argument applies to an arbitrary disjunct in a minimal necessary condition, no disjunct is empty.

References

Anscombe, G.: Causality and determination. In: Sosa, E., Tooley, M. (eds.), *Causation*, Oxford 1993, pp. 88-104.

Armstrong, D. M.: *What is a Law of Nature?*, Cambridge University Press 1983.

Broad, C. D.: The principles of demonstrative induction i-ii. In: *Mind* 39, 1930, pp. 302-317, pp. 426-439.

Broad, C. D.: Hr. von Wright on the logic of induction i-iii. In: *Mind* 53, 1944, pp. 1-24, pp. 97-119, pp. 193-214.

Cartwright, N.: *How the Laws of Physics Lie*, Oxford 1983.

Davidson, D.: Causal relations. In: Sosa, E., Tooley, M. (eds.), *Causation*, Oxford 1993, pp. 75-87.

Glymour, C., Cooper, G. F. (eds.): *Computation, Causation, and Discovery*, Cambridge 1999.

Humphreys, P.: *The Chances of Explanation*, Princeton University Press 1989.

Johnson, W. E.: *Logic, Part III*, Dover 1963 (1924).

Kim, J.: Causation, nomic subsumption, and the concept of event. In: *Journal of Philosophy* 70, 1973, pp. 217-236.

Kim, J.: Causes and events: Mackie on causation. In: Sosa, E., Tooley, M. (eds.), *Causation*, Oxford 1993, pp. 60-74.

Lewis, D.: Causation. In: Sosa, E., Tooley, M. (eds.), *Causation*, Oxford 1993, pp. 193-204.

Lewis, D.: Causation as influence. In: *The Journal of Philosophy* 97.5, 2000, pp. 182-197.

Mackie, J. L.: Causes and conditions. In: *American Philosophical Quarterly* 2, 1963, pp. 245-264.

Mackie, J. L.: *The Cement of the Universe. A study of Causation*, Oxford ²1980.

Maxwell, J. C.: *Matter and Motion*, Dover 1991 (1871).

Pearl, J.: Reasoning with cause and effect. In: *16th International Joint Conference on Artificial Intelligence*, 1999, pp. 1437-1449.

Pearl, J.: *Causality*, Cambridge 2000.

Reichenbach, H.: *The Direction of Time*, Berkeley 1956.

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

Spirtes, P., Glymour, C., Scheines, R.: *Causation, Prediction, and Search*, New York, Heidelberg 1993.

Tooley, M.: *Causation*, Oxford 1987.

v. Fraassen, B. C.: *Laws and Symmetry*, Oxford 1989.

v. Wright, G. H.: *A Treatise on Induction and Probability*, London 1952.

Brit. J. Phil. Sci. **56** (2005), 663–680

Minimal Assumption Derivation of a Bell-type Inequality

Gerd Graßhoff, Samuel Portmann and Adrian Wüthrich

ABSTRACT

John Bell showed that a big class of local hidden-variable models stands in conflict with quantum mechanics and experiment. Recently, there were suggestions that empirically adequate hidden-variable models might exist which presuppose a weaker notion of local causality. We will show that a Bell-type inequality can be derived also from these weaker assumptions.

1 Introduction

The violation of Bell's inequality by the outcome of an EPR-type spin experiment (Einstein, Podolsky, and Rosen [1935]; Bohm [1951]) seems to exclude a local theory with hidden variables. The underlying *reductio ad absurdum* proof infers on the grounds of the empirical falsification of the derived inequality that at least one of the required assumptions must be false. The force of the argument requires that the derivation be deductive and that all assumptions be explicit. We aim to extract a minimal set of assumptions needed for a deductive derivation of Bell's inequalities given perfect correlation of outcomes of an EPR-type spin experiment with parallel settings.

One of the assumptions in Bell's original derivation (Bell [1964]) was determinism. Later, he succeeded in deriving a similar inequality without determinism (Bell [1971]), placing in its stead an assumption later dubbed

doi:10.1093/bjps/axi140

Advance Access published on October 21, 2005.

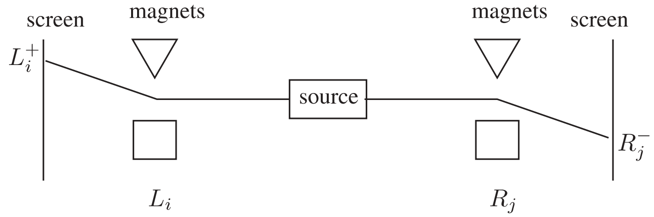

Figure 1. Setup of the EPR–Bohm experiment (cf. Bell [1987], p. 140).

local causality (Bell [1975]). As Bell stressed, the notion of local causality he and others used might be challenged. In Hofer-Szabó, Rédei, and Szabó ([1999]), it was pointed out that Reichenbach's Common Cause Principle (Reichenbach [1956]) indeed suggests a weaker form of local causality. We will prove here, however, that even from this weaker notion Bell's inequality can still be derived.[1]

2 The EPR–Bohm experiment

Consider the so-called EPR–Bohm (EPRB) experiment (Einstein, Podolsky, and Rosen [1935]; Bohm [1951]). Two spin-$\frac{1}{2}$ particles in the *singlet state*

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \tag{1}$$

are separated in such a way that one particle moves to a measurement apparatus in the left wing of the experimental setting and the other particle to a measurement apparatus in the right wing (see Figure 1). The experimenter can choose arbitrarily one of three directions in which the spin is measured with a Stern–Gerlach magnet.

The following terminology follows the reconstruction of Wigner ([1970]), which van Fraassen ([1989]) has subsequently expanded on. The event type[2] that the left (right) measurement apparatus is set to measure the spin in direction $i \in \{1,2,3\}$ is symbolized by L_i (R_i). L_i^a (R_i^a) symbolizes the event type that the measurement outcome in the left (right) wing of a spin measurement in direction i is a. There are two possible measurement outcomes *spin up* ($a = +$) and *spin down* ($a = -$) for each particle in each direction. The letter $j \in \{1, 2, 3\}$ will be used like i to symbolize directions and $b \in \{+,-\}$ like a to symbolize measurement outcomes. Formulae in which the variables i, j, a, and b appear are meant to hold—if not otherwise stated—for all possible

[1] Several of the issues we present in this paper are discussed in more detail in Wüthrich ([2003]).

[2] We will speak of *event types* to distinguish them from the *token events* which instantiate corresponding event types.

values of the variables. $p(X)$ denotes the probability of an event type X, which is empirically measurable as the relative frequency of all runs of an EPRB experiment in which the event type X is instantiated, with respect to all runs. $p(X \wedge Y)$ is the probability of the event type 'X and Y', measurable as the relative frequency of all runs in which both X and Y are instantiated. $p(X|Y) = p(X \wedge Y)/p(Y)$ is the conditional probability of the event type X *given* the event type Y, measurable as the relative frequency of instantiations of X with respect to the subensemble of all runs in which Y is instantiated. For example,

$$p(L_i^a \wedge R_j^b \,|\, L_i \wedge R_j) \tag{2}$$

denotes the probability that the measurement outcome is a on the left and b on the right when measuring in direction i on the left and in direction j on the right. These probabilities are predicted by quantum mechanics as

$$p(L_i^+ \wedge R_j^+ \,|\, L_i \wedge R_j) = \frac{1}{2}\sin^2\frac{\varphi_{ij}}{2}, \tag{3}$$

$$p(L_i^- \wedge R_j^- \,|\, L_i \wedge R_j) = \frac{1}{2}\sin^2\frac{\varphi_{ij}}{2}, \tag{4}$$

$$p(L_i^+ \wedge R_j^- \,|\, L_i \wedge R_j) = \frac{1}{2}\cos^2\frac{\varphi_{ij}}{2}, \tag{5}$$

$$p(L_i^- \wedge R_j^+ \,|\, L_i \wedge R_j) = \frac{1}{2}\cos^2\frac{\varphi_{ij}}{2}, \tag{6}$$

where φ_{ij} denotes the angle between the two measurement directions i and j. Also, the outcomes on each side are predicted separately to be completely random:

$$p\left(L_i^a \,|\, L_i \wedge R_j\right) = \frac{1}{2}, \tag{7}$$

$$p(R_j^b \,|\, L_i \wedge R_j) = \frac{1}{2}. \tag{8}$$

3 Local causality

The derivations of Bell-type inequalities known to us which *do not presuppose determinism* assume instead what John Bell calls *local causality* (Bell [1975];

Clauser and Horne [1974]), that is, the assumption that there is a common cause variable[3] V which takes on values $q \in I = \{q_1, q_2, q_3, \ldots, q_k\}$ such that for event types 'the variable V has the value q' (Vq) we have $\sum_q p(Vq) = 1$ and

$$p(L_i^a \wedge R_j^b \mid Vq \wedge L_i \wedge R_j) = p(L_i^a \mid Vq \wedge L_i)\ p(R_j^b \mid Vq \wedge R_j). \qquad (9)$$

Other frequently used names for this condition are *factorizability* (Butterfield [1989]) and *strong locality* (Jarrett [1984], [1989]). It is usually justified by pointing out that it follows from the conjunction of the following three conditions, which are called *completeness* [Equation (10)] and *locality* [Equations (11) and (12)] (Jarrett [1984], [1989]), *outcome independence* and *parameter independence* (Shimony [1993]), or *causality* and *hidden locality* (van Fraassen [1989]):

$$p(L_i^a \wedge R_j^b \mid Vq \wedge L_i \wedge R_j) = p(L_i^a \mid Vq \wedge L_i \wedge R_j)\ p(R_j^b \mid Vq \wedge L_i \wedge R_j), \qquad (10)$$

$$p(L_i^a \mid L_i \wedge R_j \wedge Vq) = p(L_i^a \mid L_i \wedge Vq), \qquad (11)$$

$$p(R_j^a \mid L_i \wedge R_j \wedge Vq) = p(R_j^a \mid R_j \wedge Vq). \qquad (12)$$

Equation (10) says that event types Vq or the variable V 'screens off' L_i^a and R_j^b from each other (van Fraassen [1989]; Butterfield [1989]). Van Fraassen ([1989]) pointed out that Equation (10) can be motivated through Reichenbach's Principle of Common Cause (PCC) (Reichenbach [1956]). The principle states that whenever two different event types A and B are statistically correlated

$$p(A \wedge B) \neq p(A)p(B) \qquad (13)$$

and neither is A causally relevant for B nor B for A, there exists a common cause variable V with values $q \in \mathrm{I} = \{q_1, q_2, q_3, \ldots, q_k\}$ ($\sum_q p(Vq) = 1$) such that A and B given Vq are uncorrelated:

$$p(A \wedge B|Vq) = p(A|Vq)p(B|Vq). \qquad (14)$$

In its original formulation the principle is stated only for a common cause event type C, which is included in our formulation as the special case where Vq can take only two values: $Vq_1 = C$, $Vq_2 = \neg C$ ('not C'). The principle has been formulated for general common cause variables by Hofer-Szabó and

[3] For the sake of simplicity, we assume that this partition is discrete and finite. As will become clear in the following, the derivation of Bell's inequality can also be done without this restriction.

Rédei ([2004]) and Placek ([2000]). Besides the screening-off condition, Reichenbach ([1956]) and Hofer-Szabó and Rédei ([2004]) stipulate further restrictions on the common cause variable, which are, however, irrelevant for our purposes.

Now, as can be seen from Equations (3)–(6), the event type L_i^a is in general correlated with event type R_j^b. It is

$$p(L_i^a \mid L_i \wedge R_j) = p(R_j^b \mid L_i \wedge R_j) = \frac{1}{2}, \tag{15}$$

and therefore

$$p(L_i^a \wedge R_j^b \mid L_i \wedge R_j) \neq p(L_i^a \mid L_i \wedge R_j)p(R_j^b \mid L_i \wedge R_j) \quad \text{except for } \varphi_{ij} = \frac{\pi}{2} \bmod \pi. \tag{16}$$

Supposing that L_i^a is not causally relevant for R_j^b and vice versa (which is reinforced by the fact that the setup of the experiment can be chosen so that the instantiations of L_i^a and R_j^b in each run of the experiment are space-like separated), PCC requires a common cause variable which fulfills Equation (10). There are several different correlations; for example, L_1^+ is correlated with R_2^+, and L_2^+ is correlated with R_3^+. For each of these correlations PCC enforces the consequence that a common cause variable exists. As stressed in Hofer-Szabó, Rédei, and Szabó ([1999]) nothing in PCC dictates that the common cause variables of the different correlations have to be the same. However, in all the derivations of Bell's inequality known to us this identification is made nevertheless. It is further shown in Hofer-Szabó, Rédei, and Szabó ([1999]) and Hofer-Szabó and Rédei ([2004]) that for any set of correlations it is mathematically possible to construct common cause variables. The authors concluded in Hofer-Szabó, Rédei, and Szabó ([1999]) that the apparent contradiction between this possibility and the claim that the EPRB correlations do not allow for a common cause variable (van Fraassen [1989]; Butterfield [1989]) is resolved by pointing out that in the derivation of Bell's inequality a *common* common cause variable for all measurements is assumed:

> The crucial assumption in the [...] derivation of the [Clauser–Horne] inequality is that [the two-valued common cause variable] is a [two-valued common cause variable] *for all four* correlated pairs, i.e. that [Vq] is a *common* common cause [variable], shared by different correlations. Without this assumption Bell's inequality *cannot* be derived. But there does not seem to be any obvious reason why common causes should also be common common causes, whether of quantum or of any other sort of correlations. (Italics in the original)

Showing the mathematical possibility of constructing common cause variables for any set of correlations and in particular for the correlations found in the EPRB experiment is not sufficient for proving the existence of a physically 'natural' hidden-variable model for that experiment, however. Besides being common cause variables [thus fulfilling Equation (10)], parameter independence should hold, too [Equations (11) and (12)]. Also, they should not be correlated with the measurement choices. As shown by Szabó ([1998]), it is possible to construct a model which fulfils these requirements for each of the common cause variables separately. However, the *conjunctions* and other logical combinations of the event types that the common cause variables have certain values correlate in that model with the measurement operations. Whether a model can be constructed without these correlations was posed as an open question by Szabó. This question is answered negatively by the derivation of Bell's inequality that we present in the remainder of this article.

4 Bell's inequality from separate common causes

4.1 A weak screening-off principle

Consider an EPRB experiment where the same direction i ($i \in \{1,2,3\}$) is chosen in both wings. That is, in each run the event type $L_i \wedge R_i$ is instantiated. With this special setting quantum mechanics predicts [see Equations (3)–(8), with $\varphi_{ij} = 0$] that the measurement outcomes in each wing are random but that the outcomes in one wing are perfectly correlated with the outcomes in the other wing: if and only if the spin of the left particle is up, then the spin of the right particle is down, and vice versa. We refer to this assumption as *perfect correlation*, or PCORR for short.

Assumption 1 (PCORR)

$$p_{ii}\left(R_i^- \mid L_i^+\right) = 1 \text{ and } p_{ii}\left(L_i^+ \mid R_i^-\right) = 1. \tag{17}$$

We use here the definition

$$p_{ij}(\ldots) \doteq p\left(\ldots \mid L_i \wedge R_j\right). \tag{18}$$

Large spatial separation of coinciding events of type L_i^a and R_j^b suggests that the respective instances are indeed distinct events. This excludes an explanation of the correlations by *event identity*, as is the case, for example, with a tossed coin for the perfect correlation of the event types 'heads up' and 'tails down'. Such a perfect correlation is explained in that every instance of 'heads up' is also an instance of 'tails down', and vice versa. Since the separation is even space-like, no L_i^a or R_j^b should be causally relevant for the other. We refer to these two assumptions as *separability*, SEP for short, and *locality 1* (LOC1).

Assumption 2 (SEP) *The coinciding instances of* L_i^a *and* R_j^b *are distinct events.*

Assumption 3 (LOC1) *No* L_i^a *or* R_j^b *is causally relevant for the other.*

Rather, there should be a common cause variable; that is, we assume PCC.

Assumption 4 (PCC) *If two event types A and B are correlated and the correlation cannot be explained by direct causation nor by event identity, then there exists a common cause variable V, with values* $q \in I = \{q_1, q_2, q_3, \ldots, q_k\}$ *such that* $\sum_q p(Vq) = 1$ *and* $p(A \wedge B \mid Vq) = p(A \mid Vq)p(B \mid Vq)$, $\forall q$.

As already mentioned, we omit the other Reichenbachian conditions (Reichenbach [1956]; Hofer-Szabó and Rédei [2004]) since they are not necessary for our derivation.

This principle, together with the assumptions PCORR, SEP, and LOC1, implies that there is for each of the EPRB correlations a (possibly different) common cause variable V_{ij}^{+-} with $q \in I_{ij}^{+-}$. The sub- and superscripts in V_{ij}^{+-} with $q \in I_{ij}^{+-}$ refer to V_{ij}^{+-} being a common cause variable of L_i^+ and R_j^-.

Result 1

$$p_{ii}\left(L_i^+ \wedge R_i^- \mid V_{ii}^{+-} q\right) = p_{ii}\left(L_i^+ \mid V_{ii}^{+-} q\right) \, p_{ii}\left(R_i^- \mid V_{ii}^{+-} q\right). \tag{19}$$

Note that common cause variables can be different for different correlations.

4.2 Perfect correlation and 'determinism'

We now show that from the fact that a *perfect* correlation is screened off by some variable it follows that without loss of generality the common cause variable can be assumed to be two-valued and that the having of one of the two values of the variables is necessary and sufficient for the instantiation of the two perfectly correlated event types, cf. Suppes and Zanotti ([1976]).

Let A and B be perfectly correlated,

$$p(A \mid B) = p(B \mid A) = 1,$$

and screened-off from each other by a common cause variable,

$$p(A \wedge B \mid Vq) = p(A \mid Vq)p(B \mid Vq).$$

We can split the set I of all values V completely into two disjoint subsets, namely into the subset I^+ of those values of V for which $p(A \wedge Vq)$ is not zero and into the subset I^- of those for which it is zero:

$$I^+ = \{q \in I : p(A \wedge Vq) \neq 0\},$$

$$I^- = \{q \in I : p(A \wedge Vq) = 0\},$$

$$I = I^- \cup I^+, \; I^- \cap I^+ = \emptyset.$$

From this definition of I^- it follows already that

$$p(A \mid Vq) = 0, \quad \forall q \in I^-, \tag{20}$$

that is, that Vq with $q \in I^+$ is necessary for A. Moreover, for all $q \in I^+$ we have by screening-off and perfect correlation

$$p(A \mid Vq) = p(A \mid B \wedge Vq) = 1. \tag{21}$$

That the variable V has a value in I^+ is a necessary and sufficient condition for A. The following calculation shows that Vq with $q \in I^+$ is also necessary and sufficient for B.

From perfect correlation it follows that

$$p(B \mid A \wedge Vq) = 1, \quad \forall q \in I^+.$$

That Vq screens off B from A yields

$$p(B \mid A \wedge Vq) = p(B \mid Vq).$$

Together with the previous equation this implies that Vq is sufficient for B for all $q \in I^+$:

$$p(B \mid Vq) = 1, \quad \forall q \in I^+. \tag{22}$$

If $q \in I^-$ we have by definition $p(A \wedge Vq) = 0$, which implies

$$p(A \wedge B \wedge Vq) = 0.$$

By perfect correlation we therefore also have $p(B \wedge Vq) = 0$, which in turn implies that

$$p(B \mid Vq) = 0, \quad \forall q \in I^-, \tag{23}$$

which means that Vq with $q \in I^+$ is also necessary for B.

This calculation shows that in the case of a perfect correlation the set of values of the common cause variable decomposes into two relevant sets. This means that whenever there is an (arbitrarily valued) common cause variable for a perfect correlation, there is also a two-valued common cause variable, namely, the disjunction of all event types Vq for which $q \in I^+$ or $q \in I^-$, respectively.

$$C = \vee_{q \in I^+} Vq,$$

$$\neg C = \vee_{q \in I^-} Vq.$$

We refer to C as a *common cause event type*. In the case of a perfect correlation no generality is achieved by allowing for a more than two-valued common cause variable; if there is a common cause variable for a perfect correlation, there is also a common cause event type. Moreover, the common cause event type is a necessary and sufficient condition for the event types that are screened off by it [Equations (20)–(23)]. Result 1 thus implies that there is a common cause event type C_{ii}^{+-} such that

$$p_{ii}\left(L_i^+ \mid C_{ii}^{+-}\right) = p_{ii}\left(R_i^- \mid C_{ii}^{+-}\right) = 1, \tag{24}$$

$$p_{ii}\left(L_i^+ \mid \neg C_{ii}^{+-}\right) = p_{ii}\left(R_i^- \mid \neg C_{ii}^{+-}\right) = 0. \tag{25}$$

The sub- and superscripts of C_{ii}^{+-} refer to C_{ii}^{+-} being the common cause event type of L_i^+ and R_i^-.

The outcome of a spin measurement is always either $+$ or $-$ and nothing else. We call this assumption *exactly one of exactly two possible outcomes* (EX).

Assumption 5 (EX)

$$p_{ii}\left(L_i^+\right) + p_{ii}\left(L_i^-\right) = 1, \quad p_{ii}\left(L_i^+ \wedge L_i^-\right) = 0, \tag{26}$$

$$p_{ii}\left(R_i^+\right) + p_{ii}\left(R_i^-\right) = 1, \quad p_{ii}\left(R_i^+ \wedge R_i^-\right) = 0. \tag{27}$$

As stressed by Fine ([1982]), among the actual measurements there are always runs in which no outcome is registered, which is normally attributed to the limited efficiency of the detectors and not taken to the statistics. If one assumes, instead, that part of these no-outcome runs are caused by the hidden variable, then it is possible to construct empirically adequate models for the EPRB experiments (Szabó [2000]; Szabó and Fine [2002]). With Assumption 5, we explicitly exclude such models.

With Assumption 5, while C_{ii}^{+-} is necessary and sufficient for L_i^+ and R_i^-, its complement, $\neg C_{ii}^{+-}$, is necessary and sufficient for the opposite outcomes, namely, L_i^- and R_i^+:

$$p_{ii}\left(L_i^- \mid C_{ii}^{+-}\right) = p_{ii}\left(R_i^+ \mid C_{ii}^{+-}\right) = 0, \tag{28}$$

$$p_{ii}\left(L_i^- \mid \neg C_{ii}^{+-}\right) = p_{ii}\left(R_i^+ \mid \neg C_{ii}^{+-}\right) = 1. \tag{29}$$

4.3 A *minimal theory* for spins

In Section 4.2 it was found that C_{ii}^{+-} is sufficient for L_i^+ *given* parallel settings ($L_i \wedge R_i$) [see Equation (24)]. That is, the conjunction $C_{ii}^{+-} \wedge L_i \wedge R_i$ is sufficient for L_i^+. But because of space-like separation of events of type L_i^+ and R_i that are instantiated in the same run, the latter types should not be causally relevant for the former. The measurement choice in one wing should be causally irrelevant for the outcomes (and the choices) in the other wing. Therefore we should discard R_i from the sufficient conjunction. The part $C_{ii}^{+-} \wedge L_i$ *alone* is sufficient for L_i^+. A similar reasoning can be applied to R_j^+, R_j, and $\neg C_{jj}^{+-}$ [cf. Equation (29)]. This is our assumption *locality 2* (LOC2).

Assumption 6 (LOC2) *If $L_i \wedge R_i \wedge X$ is sufficient for L_i^+, then $L_i \wedge X$ alone is sufficient for L_i^+, and similarly for R_j^+*, that is, if $L_j \wedge R_j \wedge Y$ *is sufficient for R_j^+, then $R_j \wedge Y$ alone is sufficient for R_j^+.*

Moreover, the remaining part $C_{ii}^{+-} \wedge L_i$ is *minimally* sufficient, in the sense that none of its parts is sufficient on its own.[4] If, for example, C_{11}^{+-} is instantiated but we do not choose to measure L_1, then L_1^+ will not be instantiated. That is to say, we cannot discard yet another conjunct of $L_i \wedge C_{ii}^{+-}$ as we discarded R_i from $C_{ii}^{+-} \wedge L_i \wedge R_i$.

Let us turn to *necessary* conditions for L_i^+. To begin with, L_i is necessary: if there is no Stern–Gerlach magnet properly set up (L_i) the particle is not deflected either up- or downwards; similarly for L_i^-, R_j^+, and R_j^-. Roughly speaking, *no outcome without measurement* (NOWM).

Assumption 7 (NOWM)

$$p\left(L_i^+ \wedge \neg L_i\right) = 0, \quad p\left(L_i^- \wedge \neg L_i\right) = 0, \tag{30}$$

$$p(R_j^+ \wedge \neg R_j) = 0, \quad p(R_j^- \wedge \neg R_j) = 0. \tag{31}$$

Second, we saw in Section 4.2 that if parallel settings are chosen and $\neg C_{ii}^{+-}$ is instantiated an event of type L_i^+ never occurs. In other words, $\neg C_{ii}^{+-} \wedge L_i \wedge R_i$ implies $\neg L_i^+$:

$$\neg C_{ii}^{+-} \wedge L_i \wedge R_i \rightarrow \neg L_i^+. \tag{32}$$

Again we propose a locality condition based on the idea that the measurement choice in one wing should be causally irrelevant for the outcomes (and the choices) in the other wing:[5] if $\neg C_{ii}^{+-} \wedge L_i \wedge R_i$ is sufficient for $\neg L_i^+$, then $\neg C_{ii}^{+-} \wedge L_i$ *alone* should be sufficient for $\neg L_i^+$. A similar reasoning can be applied to R_j^+, R_j, and C_{jj}^{+-} [cf. Equation (28)].

Assumption 8 (LOC3) *If $L_i \wedge R_i \wedge X$ is sufficient for $\neg L_i^+$, then $L_i \wedge X$ alone is sufficient for $\neg L_i^+$, and similarly for $\neg R_j^+$, that is, if $L_j \wedge R_j \wedge Y$ is sufficient for $\neg R_j^+$, then $R_j \wedge Y$ alone is sufficient for $\neg R_j^+$.*

[4] Minimal sufficient conditions as defined by Graßhoff and May ([2001]) and Baumgartner and Graßhoff ([2004]).

[5] The following version of LOC3 is slightly different from that in an earlier version of the article. We thank Gabor Hofer-Szabó, Miklós Rédei, and Iñaki San Pedro for their comments.

By LOC3 it follows from Equation (32) that

$$\neg C_{ii}^{+-} \wedge L_i \rightarrow \neg L_i^{+}. \tag{33}$$

This is equivalent to

$$L_i^{+} \wedge L_i \rightarrow C_{ii}^{+-}, \tag{34}$$

and also to

$$L_i^{+} \wedge L_i \rightarrow C_{ii}^{+-} \wedge L_i. \tag{35}$$

According to Equation (30), L_i is necessary for L_i^+. That means $L_i^+ \rightarrow L_i$, but also $L_i^+ \rightarrow L_i^+ \wedge L_i$. We have found [Equation (35)] that $L_i^+ \wedge L_i \rightarrow C_{ii}^{+-} \wedge L_i$. Altogether, this entails $L_i^+ \rightarrow L_i \wedge C_{ii}^{+-}$, that is, that $L_i \wedge C_{ii}^{+-}$ is necessary for L_i^+. Moreover, it is a *minimally* necessary condition in the sense of Graßhoff and May ([2001]) since it does not contain any disjuncts. All in all, $C_{ii}^{+-} \wedge L_i$ is a minimally necessary and minimally sufficient condition for L_i^+. In a similar vein we find that $R_j \wedge \neg C_{jj}^{+-}$ is minimally necessary and minimally sufficient for R_j^+. We have thus derived in particular the four *minimal theories* in the sense of Graßhoff and May ([2001]), as illustrated in Figure 2.

In a formal notation the four minimal theories read as the following four equations, where $\leftrightarrow$ is the usual *biconditional*, which means that the left-hand side implies the right-hand side and vice versa.[6] This intermediate result is referred to as *minimal theories* (MTH).

Result 2 (MTH)

$$\left(L_1 \wedge C_{11}^{+-}\right) \leftrightarrow L_1^{+}, \tag{2a}$$

$$\left(L_2 \wedge C_{22}^{+-}\right) \leftrightarrow L_2^{+}, \tag{2b}$$

$$\left(R_2 \wedge \neg C_{22}^{+-}\right) \leftrightarrow R_2^{+}, \tag{2c}$$

$$\left(R_3 \wedge \neg C_{33}^{+-}\right) \leftrightarrow R_3^{+}. \tag{2d}$$

From the logical relations (2a), (2b), (2c) and (2d) the following probabilities can be derived:

$$p\left(L_1^{+} \wedge R_2^{+}\right) = p\left(L_1 \wedge C_{11}^{+-} \wedge R_2 \wedge \neg C_{22}^{+-}\right),$$

$$p\left(L_2^{+} \wedge R_3^{+}\right) = p\left(L_2 \wedge C_{22}^{+-} \wedge R_3 \wedge \neg C_{33}^{+-}\right),$$

$$p\left(L_1^{+} \wedge R_3^{+}\right) = p\left(L_1 \wedge C_{11}^{+-} \wedge R_3 \wedge \neg C_{33}^{+-}\right).$$

6 For details see Graßhoff and May ([2001]) and Baumgartner and Graßhoff ([2004]). Note in particular that a correct formal notation of a minimal theory uses what both call a *double conditional*.

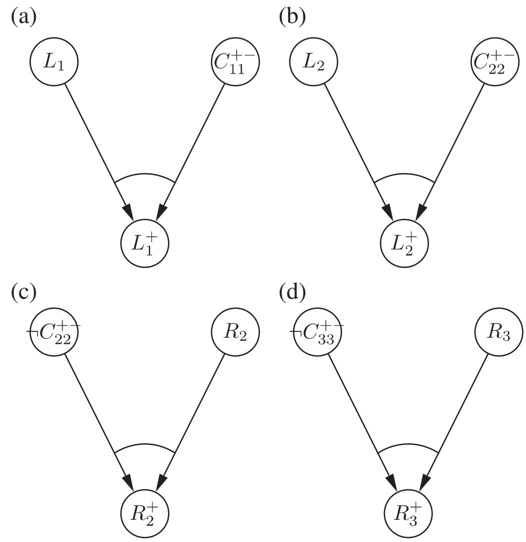

Figure 2. Minimal theories for outcomes of spin measurements according to result 2 (MTH).

By NOWM [Equations (30) and (31)] $p(L_1^+ \wedge R_2^+)$ is the same as $p(L_1^+ \wedge R_2^+ \wedge L_1 \wedge R_2)$ and so on, and the above equations read

$$p(L_1^+ \wedge R_2^+ \wedge L_1 \wedge R_2) = p(L_1 \wedge C_{11}^{+-} \wedge R_2 \wedge \neg C_{22}^{+-}), \tag{36}$$

$$p(L_2^+ \wedge R_3^+ \wedge L_2 \wedge R_3) = p(L_2 \wedge C_{22}^{+-} \wedge R_3 \wedge \neg C_{33}^{+-}), \tag{37}$$

$$p(L_1^+ \wedge R_3^+ \wedge L_1 \wedge R_3) = p(L_1 \wedge C_{11}^{+-} \wedge R_3 \wedge \neg C_{33}^{+-}). \tag{38}$$

4.4 No conspiracy

The events of type C_{ii}^{+-} are not supposed to be influenced by the measuring operations L_i and R_j. One reason for this assumption is that the measurement operations can be chosen arbitrarily before the particles enter the magnetic field of the Stern–Gerlach magnets and that an event of type C_{ii}^{+-} is assumed to happen *before* the particles arrive at the magnets. Therefore a causal influence of the measurement operations on events of type C_{ii}^{+-} would be tantamount to *backward causation*. Also an inverse statement is supposed to hold: the event types C_{ii}^{+-} are assumed not to be causally relevant for the measurement operations. This is meant to rule out some kind of 'cosmic conspiracy' that whenever an event of type C_{ii}^{+-} is instantiated, the experimenter would be 'forced' to use certain measurement operations. This *causal* independence between C_{ii}^{+-} and the measurement operations is assumed to imply the corresponding *statistical* independence. The same is assumed to hold also for *conjunctions* of common cause event types. We refer to this condition as *no conspiracy* (NO-CONS).

Assumption 9 (NO-CONS)

$$p(C_{ii}^{+-} \wedge \neg C_{jj}^{+-} \mid L_i \wedge R_j) = p(C_{ii}^{+-} \wedge \neg C_{jj}^{+-}). \tag{39}$$

By this condition of statistical independence the three probabilities considered above can be transformed. That is, we have, for instance

$$\begin{aligned} p\left(L_1^+ \wedge R_2^+ \mid L_1 \wedge R_2\right) &\doteq \frac{p\left(L_1^+ \wedge R_2^+ \wedge L_1 \wedge R_2\right)}{p(L_1 \wedge R_2)} \\ &\overset{(i)}{=} \frac{p\left(L_1 \wedge C_{11}^{+-} \wedge R_2 \wedge \neg C_{22}^{+-}\right)}{p(L_1 \wedge R_2)} \\ &\doteq p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-} \mid L_1 \wedge R_2\right) \\ &\overset{(ii)}{=} p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-}\right) \\ &\overset{(iii)}{=} p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-} \wedge C_{33}^{+-}\right) \\ &\quad + p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-} \wedge \neg C_{33}^{+-}\right). \end{aligned}$$

The dotted equations are true by definition of conditional probability. In step (i), Equation (36) was used. Step (ii) is valid by 'no conspiracy' [Equation (39)] and (iii) by a theorem of probability calculus, according to which $p(A) = p(A \wedge B) + p(A \wedge \neg B)$ for any A and B. Transforming the other two expressions in a similar way, we arrive at

$$\begin{aligned} p\left(L_1^+ \wedge R_2^+ \mid L_1 \wedge R_2\right) &= p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-} \wedge C_{33}^{+-}\right) \\ &\quad + p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-} \wedge \neg C_{33}^{+-}\right), \end{aligned} \tag{40}$$

$$\begin{aligned} p\left(L_2^+ \wedge R_3^+ \mid L_2 \wedge R_3\right) &= p\left(C_{11}^{+-} \wedge C_{22}^{+-} \wedge \neg C_{33}^{+-}\right) \\ &\quad + p\left(\neg C_{11}^{+-} \wedge C_{22}^{+-} \wedge \neg C_{33}^{+-}\right), \end{aligned} \tag{41}$$

$$\begin{aligned} p\left(L_1^+ \wedge R_3^+ \mid L_1 \wedge R_3\right) &= p\left(C_{11}^{+-} \wedge C_{22}^{+-} \wedge \neg C_{33}^{+-}\right) \\ &\quad + p\left(C_{11}^{+-} \wedge \neg C_{22}^{+-} \wedge \neg C_{33}^{+-}\right). \end{aligned} \tag{42}$$

Since both terms on the right-hand side of the last equation appear in the sum of the right-hand sides of the first two equations, the following version of the Bell inequality (BELL) follows.[7]

[7] It was first derived in this form by Wigner ([1970]).

Result 3 (BELL)

$$p\left(L_1^+ \wedge R_3^+ \mid L_1 \wedge R_3\right) \leq p\left(L_1^+ \wedge R_2^+ \mid L_1 \wedge R_2\right) + p\left(L_2^+ \wedge R_3^+ \mid L_2 \wedge R_3\right). \quad (43)$$

This inequality has been empirically falsified; see, for example, Aspect, Dalibard, and Roger ([1982]).

The inequality was derived from the following assumptions:

- perfect correlation (PCORR)
- separability (SEP)
- locality 1 (LOC1)
- principle of common cause (PCC)
- exactly one of exactly two possible outcomes (EX)
- locality 2 (LOC2)
- no outcome without measurement (NOWM)
- locality 3 (LOC3)
- no conspiracy (NO-CONS)

This is a version of Bell's theorem. It says: if these assumptions are true, the Bell inequality is true. The derivation of the Bell inequality presented here is an improvement on the usual Bell-type arguments, such as Bell ([1975]) and van Fraassen ([1989]), in two respects. First, it does not assume a *common* common cause variable for different correlations. Second, contrary to the usual locality conditions, the ones assumed here do not presuppose a solution to the problems posed by the relation between causal and statistical (in)dependence (see e.g., Spirtes, Glymour, and Scheines [1993]).

5 Discussion

Our claim to have presented a minimal assumption derivation of a Bell-type inequality is relative: our set of assumptions is weaker than any set known to us from which a Bell-type inequality can be derived and that contains the assumption of *perfect* correlation (PCORR). It was one of the achievements of Clauser and Horne ([1974]) to show that a Bell-type inequality can be derived also if the correlations of outcomes of parallel spin measurements are not assumed to be perfect. Our assumption of correlation is stronger than the one used by Clauser and Horne. However, they assume a *common* common cause variable for all correlations, which is a stronger assumption than our assumption of possibly different common cause variables for each correlation (PCC). We have not been able to derive a Bell-type inequality ruling out perfect correlations and allowing different common cause variables. If PCORR is indeed a necessary assumption for our derivation of the Bell inequality, it should be possible to construct a model in which PCORR does not hold (being violated by an arbitrary small deviation, say). Since the actually measured correlations are never perfect—a fact that is usually attributed to experimental imperfections—it is not obvious how such a model could be refuted.

Our notion of local causality might be challenged as follows. Even though nothing in PCC dictates that in general the common cause variables of different correlations have to be the same, there might be strong grounds for why they are the same in the context of the EPRB experiment. Indeed, Bell argued for his choice of local causality along the following lines.[8]

Assume that L_i^a and R_j^b are positively correlated. Then

$$p(L_i^a \mid R_j^b \wedge L_i \wedge R_j) > p(L_i^a \mid L_i \wedge R_j). \tag{44}$$

Since coinciding instances of L_i^a and R_j^b are space-like separated, neither is causally relevant for the other. Rather, the correlation should be explained by exhibiting some common causes in the overlap of the backward light cones of the coinciding instances. An instance of, say, L_i^a raises the probability of an instantiation of one of the common causally relevant factors, and this raises the probability of an instantiation of R_j^b. But *given the total state* of the overlap of the backward light cones of two coinciding instances, the probability of, say, R_j^b is assumed to be the same whether L_i^a is instantiated or not. If the total state of the overlap of the backward light cones is already given, nothing more that could be causally relevant for R_j^b can be inferred from an instance of L_i^a.

Along this line of reasoning the total state V of the overlap of the backward light cones[9] of L_i^a and R_j^b is a common cause variable which screens off the correlation:

$$p(L_i^a \wedge R_j^b \mid L_i \wedge R_j \wedge Vq) = p(L_i^a \mid L_i \wedge R_j \wedge Vq) \times p(R_j^b \mid L_i \wedge R_j \wedge Vq). \tag{45}$$

[8] For a very good and more detailed discussion of this, see Butterfield ([1989]).

[9] One might argue that the total state of the *union* of the backward light cones is a better candidate for a common cause variable (Butterfield [1989]). The following discussion carries over also to this case.

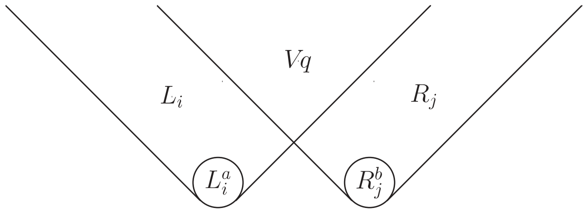

Figure 3. The two backward light cones of two measurement outcomes. The total state of the overlap is taken to define a common common cause variable V which can take on certain values q (cf. Bell [1987], p. 55).

The common past Vq cannot be altered by choosing one or the other direction for the spin measurement—'facta infecta fieri non possunt' (Placek [2000], p. 185). Therefore the total state Vq of the common past is indeed a *common* common cause variable for all correlated outcomes; see Figure 3.

This reasoning can be questioned along the following lines. It is reasonable that not all event types that are instantiated in the overlap of the backward light cones of two coinciding instances of the correlated event types are causally relevant for these latter event types. Therefore conditionalizing on the total state is conditionalizing not only on the relevant factors but also on the irrelevant ones. Moreover, it is conceivable that which event types of the common past are relevant and which are not differs for different measurements. Claiming that the total state of the common past is a common common cause variable, one is thus committed to assume that

> conditionalizing on all other events ... in addition to those affecting [the correlated event types], does not disrupt the stochastic independence induced by conditionalizing on the affecting events (Butterfield [1989]).

In particular, in the light of Simpson's paradox (Simpson [1951]) this assumption has been challenged (Cartwright [1979]). Here, we will not assess arguments in favour of or against the possibility that conditionalizing on irrelevancies yields unexpected statistical dependencies. Our point is that by weakening the assumption in the way we did, our derivation is conclusive whatever may be the answer to this question.

Acknowledgements

We would like to thank Guido Bacciagaluppi, Miklós Rédei, Gabor Hofer-Szabó, Iñaki San Pedro, and Christian Wüthrich for fruitful discussions.

Gerd Graßhoff
History and Philosophy of Science
Exact Sciences

Sidlerstrasse 5
University of Bern
CH-3012 Bern
Switzerland
gerd.grasshoff@philo.unibe.ch

Samuel Portmann and Adrian Wüthrich
Institute of Theoretical Physics
Exact Sciences
Sidlerstrasse 5
University of Bern
CH-3012 Bern
Switzerland
portmann@itp.unibe.ch
awuethr@itp.unibe.ch

REFERENCES

Aspect, A., Dalibard, J. and Roger, G. [1982]: 'Experimental Test of Bell's Inequality Using Time-Varying Analyzers', *Physical Review Letters*, **49**, pp. 1804–7.

Baumgartner, M. and Graßhoff, G. [2004]: *Kausalität und kausales Schliessen*, Bern: Bern Studies in the History and Philosophy of Science.

Bell, J. S. [1964]: 'On the Einstein-Podolsky-Rosen Paradox', *Physics*, **1**, p. 195. Reprinted in Bell ([1987], pp. 14–21).

Bell, J. S. [1971]: 'Introduction to the Hidden-Variable Question', in *Foundations of Quantum Mechanics*, p. 171, *Proceedings of the International School of Physics 'Enrico Fermi'*, New York: Academic. Reprinted in Bell ([1987], pp. 29–39).

Bell, J. S. [1975]: 'The theory of local beables', TH-2053-CERN, presented at the Sixth GIFT Seminar, Jaca, 2–7 June 1975, reproduced in *Epistemological Letters*, March 1976, and reprinted in Bell [1975], pp. 52–62.

Bell, J. S. [1987]: *Speakable and Unspeakable in Quantum Mechanics*, Cambridge: Cambridge University Press.

Bohm, D. [1951]: *Quantum Theory*, New York: Prentice Hall.

Butterfield, J. [1989]: 'A Space-Time Approach to the Bell inequality', in J. T. Cushing and E. McMullin (*eds*), 1989, *Philosophical Consequences of Quantum Theory. Reflections on Bell's Theorem*, Notre Dame: University of Notre Dame Press, pp. 114–44.

Cartwright, N. [1979]: 'Causal Laws and Effective Strategies', *Noûs*, **13**, pp. 419–37.

Clauser, J. F. and Horne, M. A. [1974]: 'Experimental consequences of objective local theories', *Physical Review D*, **10**, pp. 526–35.

Einstein, A., Podolsky, B. and Rosen, N. [1935]: 'Can quantum-mechanical description of physical reality be considered complete?', *Physical Review*, **47**, pp. 777–80.

Fine, A. [1982]: 'Some Local Models for Correlation Experiments', *Synthese*, **50**, p. 279–94.

Graßhoff, G. and May, M. [2001]: ‘Causal Regularities’, in W. Spohn, M. Ledwig and M. Esfeld (*eds*), *Current Issues in Causation*, Paderborn: Mentis, pp. 85–114.

Hofer-Szabó, G., Rédei, M. and Szabó, L. E. [1999]: ‘On Reichenbach’s Common Cause Principle and Reichenbach’s Notion of Common Cause’, *British Journal for the Philosophy of Science*, **50**, pp. 377–99.

Hofer-Szabó, G. and Rédei, M. [2004]: ‘Reichenbachian Common Cause Systems’, *International Journal of Theoretical Physics*, **43**(7/8), pp. 1819–26.

Jarrett, J. P. [1984]: ‘On the Physical Significance of the Locality Conditions in the Bell Arguments’, *Noûs*, **18**, pp. 569–89.

Jarrett, J. P. [1989]: ‘Bell’s Theorem: A Guide to the Implications’, in J. T. Cushing and E. McMullin (*eds*), 1989, *Philosophical Consequences of Quantum Theory. Reflections on Bell’s Theorem*, Notre Dame: University of Notre Dame Press, pp. 60–79.

Placek, T. [2000]: *Is Nature deterministic*?, Kraków: Jagiellonian University Press.

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

Shimony, A. [1993]: *Search for a naturalistic world view*, volume 2, Cambridge: Cambridge University Press.

Simpson, E. H. [1951]: ‘The Interpretation of Interaction in Contingency Tables’, *Journal of the Royal Statistical Society, Ser. B.*, **13**, pp. 238–41.

Spirtes, P., Glymour, C. and Scheines, R. [1993]: *Causation, Prediction, and Search*, New York: Springer Verlag.

Suppes, P. and Zanotti, M. [1976]: ‘On the Determinism of Hidden Variable Theories with Strict Correlation and Conditional Statistical Independence of Observables’, in P. Suppes (*ed.*), *Logic and Probability in Quantum Mechanics*, Dordrecht: Reidel, pp. 445–55.

Szabó, L. E. [1998]: ‘On an attempt to resolve the EPR-Bell paradox via Reichenbachian concept of common cause’, electronic preprint; <http://arxiv.org/abs/quant-ph/9806074>.

Szabó, L. [2000]: ‘On Fine’s Resolution of the EPR-Bell Problem’, electronic preprint, <http://arxiv.org/abs/quant-ph/0002030>.

Szabó, L. and Fine, A. [2002]: ‘A Local Hidden Variable Theory for the GHZ Experiment’, PhilSci Archive, <http://philsci-archive.pitt.edu/archive/00000642/>.

van Fraassen, B. C. [1989]: ‘The Charybdis of Realism: Epistemological Implications of Bell’s Inequalities’, in J. T. Cushing and E. McMullin (*eds*), 1989, *Philosophical Consequences of Quantum Theory. Reflections on Bell’s Theorem*, Notre Dame: University of Notre Dame Press, pp. 97–113.

Wigner, E. [1970]: ‘On Hidden Variables and Quantum Mechanical Probabilities’, *American Journal of Physics*, **38**, p. 1005–9.

Wüthrich, A. [2004]: *Quantum Correlations and Common Causes - An Analysis of the Assumptions of Bell’s Theorem*, Bern: Bern Studies in the History and Philosophy of Science.

Studies in History and Philosophy of Modern Physics 38 (2007) 844–862

Studies in History and Philosophy of Modern Physics

www.elsevier.com/locate/shpsb

Minimal assumption derivation of a weak Clauser–Horne inequality

Samuel Portmann[a], Adrian Wüthrich[b,*]

[a]*Institute of Theoretical Physics/History and Philosophy of Science, Sidlerstrasse 5, University of Bern, CH-3012 Bern, Switzerland*

[b]*History and Philosophy of Science, Exact Sciences, Sidlerstrasse 5, University of Bern, CH-3012 Bern, Switzerland*

Received 28 April 2006; received in revised form 9 October 2006; accepted 27 February 2007

Abstract

According to Bell's theorem a large class of hidden-variable models obeying Bell's notion of local causality (LC) conflict with the predictions of quantum mechanics. Recently, a Bell-type theorem has been proven using a weaker notion of LC, yet assuming the existence of perfectly correlated event types. Here we present a similar Bell-type theorem without this latter assumption. The derived inequality differs from the Clauser–Horne inequality by some small correction terms, which render it less constraining.

Keywords: Bell's theorem; Principle of Common Cause; Perfect correlations

1. Introduction

In this article we continue the work of Graßhoff, Portmann, and Wüthrich (2005) and prove a Bell-type theorem from a still weaker set of assumptions. In contrast to Graßhoff et al. (2005), the weakening is reflected in the derived inequality: We get the Clauser–Horne inequality with small correction terms rendering our inequality less constraining.

In order to set the theoretical stage, we would like to set our project in the context of other work aiming to minimalise the strength of the assumptions (see Fig. 1). For more detailed reviews see e.g. Shimony (2005) and Clauser and Shimony (1978).

*Corresponding author.

E-mail addresses: portmann@itp.unibe.ch (S. Portmann), adrian.wuethrich@philo.unibe.ch (A. Wüthrich).

1355-2198/$ - see front matter
doi:10.1016/j.shpsb.2007.02.002

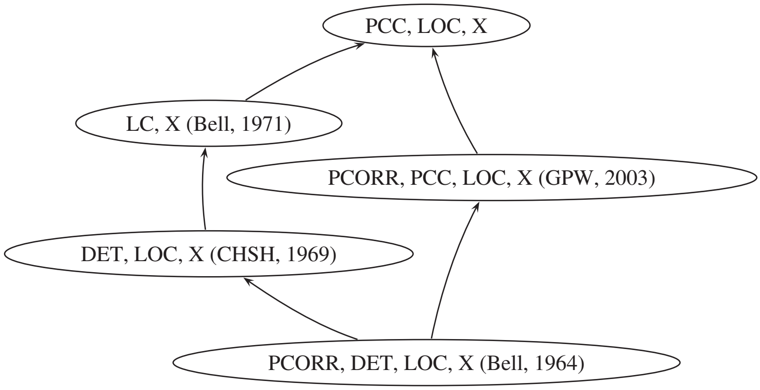

Fig. 1. Comparison of the logical strengths of the different sets of assumptions. Each node stands for a set of assumptions from which a Bell-type inequality was derived. If two nodes are connected by an arrow, the set further upwards is a logical implication of the set further downwards but not vice versa. DET = Determinism, LOC = Locality, PCORR = $\exists$ perfectly correlated event types, LC = local causality, PCC = Principle of Common Cause, X = further assumptions, shared by all derivations.

The experimental context of all these derivations is the EPR-Bohm (EPRB) experiment (see Section 2). Furthermore, they all assume a locality and a causality condition for the observable events in terms of "hidden" variables. In its canonical interpretation, quantum mechanics (QM) violates the locality but not the causality condition. In his seminal derivation, Bell (1964) assumed local determinism (LOC and DET) and, additionally, the existence of perfectly correlated event types (PCORR). Then, Clauser, Horne, Shimony, and Holt (1969) derived the CHSH-inequality—again with LOC and DET, but without PCORR. Moreover, Bell (1971) showed 2 years later, that the same inequality can even be derived if one replaces the assumption of local determinism with a weaker probabilistic notion, which he dubbed "local causality" (LC) (Bell, 1975), and which was later analysed by Suppes and Zanotti (1976), van Fraassen (1982) and Jarrett (1984) as a conjunction of a locality and a causality condition. As the long philosophical discussion demonstrates, it is even difficult to find a necessary condition for probabilistic causation, not to mention sufficient ones. Already Bell (1975) stressed that other definitions of LC are conceivable. Belnap and Szabó (1996) and Hofer-Szabó, Rédei, and Szabó (1999) showed that Reichenbach's Principle of Common Cause does indeed suggest another form of causality (PCC), which, together with LOC, includes Bell's notion only as a special case. They also pointed out that the existing proofs all assume the stronger notion and that it is thus not clear whether a Bell-type theorem can still be proven with PCC. We used PCC as our causality condition in Graßhoff et al. (2005) for a proof of a Bell-type theorem, but the minimality of the logical strength of the assumptions was only relative (see Fig. 1), because we also assumed PCORR. *Given* this assumption, our set of assumptions was minimal. However, there are reasons to think that PCORR might be false (see Section 3.2), which would limit the significance of our result. In this article, we derive a Bell-type inequality without assuming PCORR. Our approach is similarly "straightforward" as the one of Ryff (1997). His intuition "that if a theorem is valid whenever we have perfect correlations, it cannot be totally wrong in the case of almost perfect correlations" can be formulated precisely and proven to be correct in our case. A very similar result for the special case of two-valued common causes was derived independently by Hofer-Szabó (2007).

This article is structured as follows. We describe the EPRB experiment and introduce our notation in Section 2. In the main part, Section 3, we derive a weak Clauser–Horne inequality. In Section 4, we discuss our result and compare it to related work. Specifically, we discuss the significance of the small correction terms in our inequality.

2. The EPRB experiment

Consider the so-called *EPRB experiment* (Bohm, 1951; Einstein, Podolsky, & Rosen, 1935). Two spin-$\frac{1}{2}$ particles in the *singlet state*

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \tag{1}$$

are separated such that one particle moves to the measurement apparatus of Alice on the left and the other particle to the measurement apparatus of Bob on the right (see Fig. 2). The experimenter can arbitrarily choose the direction in which the spin is measured with a Stern–Gerlach magnet.

The event type (henceforth: event, for short) that Alice's (Bob's) measurement apparatus is set to measure the spin in direction a (b) is symbolised by a (b). A_a (B_b) symbolises the measurement outcome of Alice (Bob) for a measurement in direction a (b). For each direction, there are two possible measurement outcomes: *spin up* ($A_a = +_a$, $B_b = +_b$) and *spin down* ($A_a = -_a$, $B_b = -_b$). We will always interpret these events also as elements of a Boolean algebra Ω with a classical probability measure p, constituting a classical probability space (Ω, p). E.g.

$$p(A_a B_b|ab) \tag{2}$$

denotes the probability that Alice' measurement outcome is A_a and Bob's B_b, when measuring in the directions a (Alice) and b (Bob). We will often use the notation

$$p_{a,b}(\cdots) := p(\cdots|ab) \tag{3}$$

with which we can write (2) as

$$p_{a,b}(A_a B_b). \tag{4}$$

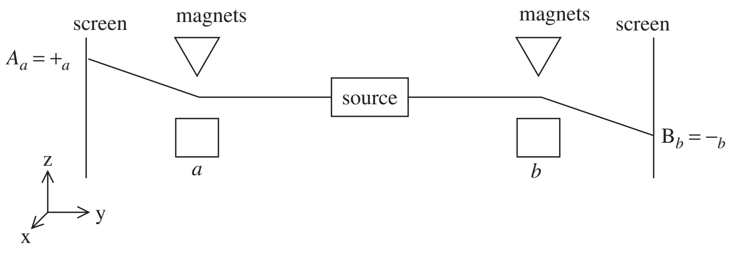

Fig. 2. Setup of the EPR-Bohm experiment.

These probabilities are predicted by QM as

$$p_{a,b}(+_a+_b) = \tfrac{1}{2}\sin^2\frac{\varphi_{a,b}}{2}, \tag{5}$$

$$p_{a,b}(-_a-_b) = \tfrac{1}{2}\sin^2\frac{\varphi_{a,b}}{2}, \tag{6}$$

$$p_{a,b}(+_a-_b) = \tfrac{1}{2}\cos^2\frac{\varphi_{a,b}}{2}, \tag{7}$$

$$p_{a,b}(-_a+_b) = \tfrac{1}{2}\cos^2\frac{\varphi_{a,b}}{2}, \tag{8}$$

where $\varphi_{a,b}$ denotes the angle between the two measurement directions a and b. Also, the outcomes on each side are predicted separately to be completely random:

$$p_{a,b}(A_a) = \tfrac{1}{2}, \tag{9}$$

$$p_{a,b}(B_b) = \tfrac{1}{2}. \tag{10}$$

3. Proof of a Bell-type theorem

We will first introduce our assumptions (Sections 3.1 and 3.2), from which the Bell-type inequality is then shown to follow. In the literature on Bell-type theorems, there is a huge amount of work devoted to the discussion of the assumptions. We do not intend to contribute to this discussion here.

3.1. Locality and causality

The correlation between "heads up" and "tails down" when tossing a coin is explained by the identity of the instances of the respective events: Every instance of "heads up" is also an instance of "tails down", and vice versa. Large spatial separation of coinciding instances of A_a and B_b suggests that this is not the case in the EPRB setup. This is our first assumption.

Assumption 1. The coinciding instances of the events A_a and B_b are distinct.

Given this assumption, we can express

Assumption 2. No A_a or B_b is causally relevant for the other.

This assumption is supported by the fact that EPRB experiments have been performed where in each run of the experiment the instance of A_a is space-like separated from the instance of B_b, see e.g. Aspect, Dalibard, and Roger (1982). If it were violated and if a cause temporally precedes its effects, the direction of causation would depend on the chosen inertial frame.[1]

Assumption 3 (*Principle of Common Cause, PCC*). If two events A and B with distinct coinciding instances are correlated and neither A is causally relevant for B nor vice versa,

[1]Note however that this is per se not a violation of Lorentz invariance and that whether or not this stands in contradiction to the special theory of relativity is an intricate matter. For a discussion, see for example Maudlin (1994) and Weinstein (2006).

then there exists a partition $C = \{C_i\}_{i \in I}$ of Ω, a *common cause*, such that

$$p(AB|C_i) = p(A|C_i)p(B|C_i), \quad \forall i \in I (I \text{ countable}).$$

The common cause can alternatively be thought of as a variable (the "hidden" variable) taking on the elements of the partition as values. Thus, when we say "the value of the common cause", we refer to an element of the partition. In the original formulation, Reichenbach used a partition with two elements, which is here generalised to a partition with countably many elements.[2]

As can be seen from Eqs. (5)–(10), in general, the event A_a is correlated with event B_b:

$$p_{a,b}(A_aB_b) \neq p_{a,b}(A_a)p_{a,b}(B_b), \quad \text{except for } \varphi_{a,b} = \frac{\pi}{2} \bmod \pi. \tag{11}$$

With Assumptions 1 and 2, PCC demands the existence of a common cause $C^{abAB} = \{C_i^{abAB}\}_{i \in I^{abAB}}$ which screens off the correlation:

$$p_{a,b}(A_aB_b|C_i^{abAB}) = p_{a,b}(A_a|C_i^{abAB})p_{a,b}(B_b|C_i^{abAB}), \quad \forall i \in I^{abAB}. \tag{12}$$

As in Graßhoff et al. (2005) there is more than one common cause; in the present case a common cause C^{abAB} for *each* quadruple of measurement directions and outcomes (a, b, A_a, B_b). That is different from other derivations, where a single common cause $\{C_i\}_{i \in I}$ is stipulated for all correlated events:

$$p_{a,b}(A_aB_b|C_i) = p_{a,b}(A_a|C_i)p_{a,b}(B_b|C_i), \quad \forall i \in I. \tag{13}$$

That a *common* common cause (obeying (13)) was assumed in Bell-type theorems was first pointed out and criticised by Belnap and Szabó (1996) (for a deterministic common cause) and by Rédei (1997) (for a probabilistic common cause).

Assumption 4 (*LOC*).

$$p(A_a|abC_i^{abAB}) = p(A_a|aC_i^{abAB}), \tag{14}$$

$$p(B_b|abC_i^{abAB}) = p(B_b|bC_i^{abAB}), \quad \forall i \in I^{abAB}. \tag{15}$$

As in traditional derivations, this assumption is meant to prevent the possibility of superluminal causation.

3.2. *Common causes for the maximal correlations*

In their derivation Graßhoff et al. (2005) exploit that the screening-off condition entails that in the case of perfect correlations common causes determine their effects. The slightest deviation from

$$p_{a=b}(+_a|-_b) = p_{a=b}(+_b|-_a) = 1 \tag{16}$$

leads to a breakdown of that type of derivation. Of course, Eq. (16) is true according to QM and any apparent violation in actual experiments may be attributed to experimental shortcomings, for instance that, in practice, the measurement devices are never set up perfectly parallel.

[2] Reichenbach (1956) and Hofer-Szabó and Rédei (2004) stipulate further conditions, for the two-valued and the general case, respectively. For our derivation, we do not need these assumptions, though. As emphasised in Hofer-Szabó, Rédei, and Szabó (2002), there are not only quantum but quite ordinary correlations that fail to have a two-valued *common* common cause.

Nevertheless, we would like to do without this assumption. Our motivation for this is twofold. First, there are theoretical grounds on which to expect a violation of the quantum mechanical prediction of perfect correlations. Some of the different approaches to quantum gravity, that is, suggest that tiny violations of Lorentz group invariance are to be expected.[3] Seen as an implication of rotation invariance, (16) would not be warranted any more. The second motivation has to do with the prominent claim that Bell-type theorems rule out the existence of empirically adequate local hidden-variable models on empirical grounds alone. However, if besides the assumptions that define the model as a local hidden-variable model, the only constraint was empirical adequacy, PCORR should not be assumed, because small violations of it are consistent with empirical data.[4]

These considerations motivate a weakening of (16) such that we just take the maximal correlations available, without assuming that they are perfect. We do this as follows. For each pair of measurement directions (a, b), we parametrise the conditional probabilities $p_{a,b}(+_a|-_b)$ and $p_{a,b}(+_b|-_a)$ as

$$\begin{aligned} &p_{a,b}(+_a|-_b) = 1 - \varepsilon_{a,b}, \\ &p_{a,b}(+_b|-_a) = 1 - \varepsilon_{b,a} \quad \text{with } \varepsilon_{a,b}, \varepsilon_{b,a} \in [0, 1]. \end{aligned} \tag{17}$$

We will call the set of all measurement directions of Alice (of Bob) D_{A} (D_{B}). For each measurement direction $a \in D_{\mathrm{A}}$ ($b \in D_{\mathrm{B}}$), we pick out the measurement direction $\dot{a} \in D_{\mathrm{B}}$ ($\dot{b} \in D_{\mathrm{A}}$) for which $p_{a,b}(+_a|-_b)$ ($p_{a,b}(+_b|-_a)$) takes on its maximal value, or, equivalently, $\varepsilon_{a,b}$ ($\varepsilon_{b,a}$) takes on its minimal value.[5] If the same minimal value is taken on for more than one direction, we make an arbitrary choice. We denote this minimal value with ε_a (ε_b)[6]:

$$\begin{aligned} &\varepsilon_a \colon \min_b\{\varepsilon_{a,b}\}, \\ &\varepsilon_b \colon \min_a\{\varepsilon_{b,a}\}. \end{aligned} \tag{18}$$

Thus, we have

$$\begin{aligned} &p_{a,\dot{a}}(+_a|-_{\dot{a}}) = 1 - \varepsilon_a, \\ &p_{\dot{b},b}(+_b|-_{\dot{b}}) = 1 - \varepsilon_b. \end{aligned} \tag{19}$$

Because of Assumptions 1–3, we have (in the notation of formula (12)) a common cause $\{C_i^{a\dot{a}+-}\}_{i\in I^{a\dot{a}+-}}$ ($\{C_i^{\dot{b}b-+}\}_{i\in I^{\dot{b}b-+}}$) for the events $+_a$ and $-_{\dot{a}}$ ($+_b$ and $-_{\dot{b}}$). Henceforth, we will use the short hand $\{C_i^a\}_{i\in I^a}$ ($\{C_i^b\}_{i\in I^b}$) for $\{C_i^{a\dot{a}+-}\}_{i\in I^{a\dot{a}+-}}$ ($\{C_i^{\dot{b}b-+}\}_{i\in I^{\dot{b}b-+}}$).With this notation, we get

$$\begin{aligned} &p_{a,\dot{a}}(+_a-_{\dot{a}}|C_i^a) = p_{a,\dot{a}}(+_a|C_i^a)p_{a,\dot{a}}(-_{\dot{a}}|C_i^a), \quad \forall i \in I^a, \\ &p_{\dot{b},b}(-_{\dot{b}}+_b|C_i^b) = p_{\dot{b},b}(-_{\dot{b}}|C_i^b)p_{\dot{b},b}(+_b|C_i^b), \quad \forall i \in I^b. \end{aligned} \tag{20}$$

The postulate that the common causes be not causally relevant for the setting of the measurement apparatuses or vice versa and the assumption that there is not again a common cause for these factors motivate the statistical independence of the measurement

[3] See e.g. Mattingly (2005) for references.

[4] In the context of the Kochen–Specker theorem, a similar loophole was exploited to construct a non-contextual empirically adequate model by Clifton and Kent (2000).

[5] If the number of measurement directions (i.e. the cardinality of D_{A} and D_{B}) is not finite, it is possible, that there is no such minimal value but only an infimum. The proof can be amended also for this case, but we will refrain from doing this here.

[6] Note that we do not assume that the minimal value is taken on for parallel measurement directions.

settings and *any* combination of common causes. In order to derive a Bell-type inequality, however, we only need the following independences:

Assumption 5.

$$p(aC_i^a) = p(a)p(C_i^a), \tag{21}$$

$$p(bC_i^b) = p(b)p(C_i^b), \tag{22}$$

$$p(abC_i^a) = p(ab)p(C_i^a), \tag{23}$$

$$p(abC_i^b) = p(ab)p(C_i^b), \tag{24}$$

$$p(abC_i^a C_j^b) = p(ab)p(C_i^a C_j^b), \quad \forall i,\ j \in I^a,\ I^b. \tag{25}$$

3.3. Constraints for $p_{a,b}(+_a+_b)$, $p(+_a|a)$ and $p(+_b|b)$

To obtain a Bell-type inequality we need an upper and a lower bound for

$$p_{a,b}(+_a+_b), \quad p(+_a|a) \quad \text{and} \quad p(+_b|b). \tag{26}$$

We will need the following proposition.

Proposition 1. *Let two events A and B with $p(A) = p(B) = 0.5$ be almost perfectly correlated ($p(A|B) = 1 - \varepsilon$) and assume a common cause $C = \{C_i\}_{i\in I}$, such that*

$$p(AB|C_i) = p(A|C_i)p(B|C_i), \quad \forall i \in I. \tag{27}$$

Then

$$\sum_{i\in I_1} p(C_i) - \sqrt{\varepsilon} \leqslant p(A) < \sum_{i\in I_1} p(C_i) + 4\sqrt{\varepsilon} - 2\varepsilon, \tag{28}$$

where

$$I_1 \colon \{i \in I \colon p(A|C_i) \geqslant 1 - \sqrt{\varepsilon}\}. \tag{29}$$

We prove Proposition 1 in Appendix A.

With the definition

$$C := \vee_{i\in I_1} C_i, \tag{30}$$

Eq. (28) reads

$$p(C) - \sqrt{\varepsilon} \leqslant p(A) < p(C) + 4\sqrt{\varepsilon} - 2\varepsilon, \tag{31}$$

or, equivalently,

$$p(A) - 4\sqrt{\varepsilon} + 2\varepsilon < p(C) \leqslant p(A) + \sqrt{\varepsilon}. \tag{32}$$

We define

$$\begin{aligned} I_1^a &:= \{i \in I^a \colon p_{a,\dot{a}}(+_a|C_i^a) \overset{(*)}{=} p(+_a|aC_i^a) \geqslant 1 - \sqrt{\varepsilon_a}\}, \\ C^a &:= \vee_{i\in I_1^a} C_i^a, \\ I_1^b &:= \{i \in I^b \colon p_{\dot{b},b}(+_b|bC_i^b) \overset{(*)}{=} p(+_b|bC_i^b) \geqslant 1 - \sqrt{\varepsilon_b}\}, \\ C^b &:= \vee_{i\in I_1^b} C_i^b. \end{aligned} \tag{33}$$

In $(*)$, we use LOC.

With the substitutions

$$
\begin{aligned}
&p(\cdots) \to p_{a,\dot{a}}(\cdots), \quad p(\cdots) \to p_{\dot{b},b}(\cdots),\\
&A \to +_a, \quad A \to +_b,\\
&B \to -_{\dot{a}}, \quad B \to -_{\dot{b}},\\
&C \to C^a, \quad C \to C^b,\\
&\varepsilon \to \varepsilon_a, \quad \varepsilon \to \varepsilon_b,
\end{aligned} \tag{34}
$$

we get

$$
\begin{aligned}
&p_{a,\dot{a}}(+_a) - 4\sqrt{\varepsilon_a} + 2\varepsilon_a < p_{a,\dot{a}}(C^a) \leqslant p_{a,\dot{a}}(+_a) + \sqrt{\varepsilon_a},\\
&p_{\dot{b},b}(+_b) - 4\sqrt{\varepsilon_b} + 2\varepsilon_b < p_{\dot{b},b}(C^b) \leqslant p_{\dot{b},b}(+_b) + \sqrt{\varepsilon_b}.
\end{aligned} \tag{35}
$$

By Assumption 5 we have

$$p_{a,\dot{a}}(C^a) = p(C^a) = p(C^a|a), \tag{36}$$

and $p_{a,\dot{a}}(+_a)$ can be transformed, using Assumptions 4 and 5, as follows:

$$
\begin{aligned}
p_{a,\dot{a}}(+_a) &= \sum_i p(+_a C_i^a|a\dot{a}) = \sum_i p(+_a|C_i^a a\dot{a})p(C_i^a|a\dot{a})\\
&= \sum_i p(+_a|C_i^a a)p(C_i^a|a) = \sum_i p(+_a C_i^a|a) = p(+_a|a).
\end{aligned} \tag{37}
$$

Using the definition

$$\varepsilon := \max_{a,b}\{\varepsilon_a, \varepsilon_b\}, \tag{38}$$

we thus get

$$p(+_a|a) - \Delta^+ < p(C^a) \leqslant p(+_a|a) + \Delta^-, \tag{39}$$

and similarly

$$p(+_b|b) - \Delta^+ < p(C^b) \leqslant p(+_b|b) + \Delta^- \tag{40}$$

with

$$
\begin{aligned}
\Delta^+ &= 4\sqrt{\varepsilon} - 2\varepsilon,\\
\Delta^- &= \sqrt{\varepsilon}.
\end{aligned} \tag{41}
$$

Using again Assumption 5, the following bounds for $p(+_a +_b|ab)$ can be derived (this is shown in Appendix B):

$$p_{a,b}(+_a +_b) - \Delta^+_{a,b} < p(C^a C^b) \leqslant p_{a,b}(+_a +_b) + \Delta^-_{a,b} \tag{42}$$

with

$$
\begin{aligned}
\Delta^+_{a,b} &= \frac{(p(a) + p(b))(5\sqrt{\varepsilon} - 2\varepsilon)}{p(ab)},\\
\Delta^-_{a,b} &= \frac{(p(a) + p(b))\sqrt{\varepsilon}}{p(ab)}.
\end{aligned} \tag{43}
$$

3.4. A weak Clauser–Horne inequality

In the next step, we make use of a constraint, which holds for the probabilities of arbitrary events. For events A and B to be elements of a classical probability space, it is not enough that

$$\begin{aligned} &0 \leqslant p(A) \leqslant 1, \\ &0 \leqslant p(B) \leqslant 1, \\ &0 \leqslant p(AB) \leqslant 1, \end{aligned} \tag{44}$$

and

$$\begin{aligned} &p(AB) \leqslant p(A), \\ &p(AB) \leqslant p(B). \end{aligned} \tag{45}$$

To find further constraints we first note that

$$p(AB) + p(A\bar{B}) + p(\bar{A}B) + p(\bar{A}\bar{B}) = 1, \tag{46}$$

where "$\bar{A}$" means "not A". We also have

$$\begin{aligned} &p(A) = p(AB) + p(A\bar{B}), \quad \text{and} \\ &p(B) = p(AB) + p(\bar{A}B), \end{aligned} \tag{47}$$

and hence

$$p(A) + p(B) - p(AB) = p(AB) + p(A\bar{B}) + p(\bar{A}B), \tag{48}$$

which implies with Eq. (46)

$$0 \leqslant p(A) + p(B) - p(AB) \leqslant 1. \tag{49}$$

For more than two events there are more constraints in the form of such inequalities.[7] For four events A, A', B and B', one constraint reads

$$-1 \leqslant p(AB) + p(AB') + p(A'B') - p(A'B) - p(A) - p(B') \leqslant 0. \tag{50}$$

This is the Clauser–Horne inequality[8] (Clauser & Horne, 1974), which we prove in Appendix C. This inequality is an a priori constraint for arbitrary events. Hence, for the measurement directions $1, 2 \in D_{\mathrm{A}}$ and $3, 4 \in D_{\mathrm{B}}$, we also have

$$-1 \leqslant p(C^1C^3) + p(C^1C^4) + p(C^2C^4) - p(C^2C^3) - p(C^1) - p(C^4) \leqslant 0. \tag{51}$$

Together with inequality (42)

$$p_{a,b}(+_a+_b) - \Delta^+_{a,b} < p(C^aC^b) \leqslant p_{a,b}(+_a+_b) + \Delta^-_{a,b}$$

[7]For a detailed discussion and the beautiful connection to the geometry of convex polytopes, see e.g. Pitowsky (1989).

[8]What Clauser and Horne (1974) have actually derived is inequality (52) without the correction terms (the Δ's). In (52), there are conditional probabilities involved. Nevertheless, we adopt common terminology and refer to both inequalities with the same name, since it will always be clear from the context which is meant.

and inequalities (39) and (40)

$$p(+_a|a) - \Delta^+ < p(C^a) \leqslant p(+_a|a) + \Delta^-,$$
$$p(+_b|b) - \Delta^+ < p(C^b) \leqslant p(+_b|b) + \Delta^-$$

one gets

$$\begin{aligned} &-1 - \Delta^-_{1,3} - \Delta^-_{1,4} - \Delta^-_{2,4} - \Delta^+_{2,3} - 2\Delta^+ \\ &\quad < p_{1,3}(+_1+_3) + p_{1,4}(+_1+_4) + p_{2,4}(+_2+_4) \\ &\qquad - p_{2,3}(+_2+_3) - p(+_1|1) - p(+_4|4) \\ &\quad < \Delta^+_{1,3} + \Delta^+_{1,4} + \Delta^+_{2,4} + \Delta^-_{2,3} + 2\Delta^- \end{aligned} \tag{52}$$

with

$$\begin{aligned} \Delta^-_{a,b} &= \frac{(p(a)+p(b))\sqrt{\varepsilon}}{p(ab)}, \\ \Delta^+_{a,b} &= \frac{(p(a)+p(b))(5\sqrt{\varepsilon}-2\varepsilon)}{p(ab)}, \\ \Delta^- &= \sqrt{\varepsilon}, \\ \Delta^+ &= 4\sqrt{\varepsilon} - 2\varepsilon. \end{aligned} \tag{53}$$

Note that inequality (52) reduces to the Clauser–Horne inequality for $\varepsilon = 0$.

3.5. Contradiction

The predicted values of $p_{1,3}(+_1, +_3)$, $p_{1,4}(+_1, +_4)$, $p_{2,4}(+_2, +_4)$ and $p_{2,3}(+_2, +_3)$ by QM are such that the maximal[9] violation for the lower bound of (52) obtains (among others) for the angles $\varphi_{1,3} = \varphi_{1,4} = \varphi_{2,4} = \pi/4$ and $\varphi_{2,3} = 3\pi/4$:

$$\begin{aligned} &-1 - \Delta^-_{1,3} - \Delta^-_{1,4} - \Delta^-_{2,4} - \Delta^+_{2,3} - \Delta^+ - \Delta^+ \\ &\quad < p_{1,3}(+_1+_3) + p_{1,4}(+_1+_4) + p_{2,4}(+_2, +_4) \\ &\qquad - p_{2,3}(+_2+_3) - p(+_1|1) - p(+_4|4) = -\frac{\sqrt{2}+1}{2}. \end{aligned} \tag{54}$$

The maximal violation for the upper bound obtains (among others) for the angles $\varphi_{1,3} = \varphi_{2,4} = 3\pi/4$, $\varphi_{1,4} = 5\pi/4$ and $\varphi_{2,3} = \pi/4$:

$$\begin{aligned} \frac{\sqrt{2}-1}{2} &= p_{1,3}(+_1+_3) + p_{1,4}(+_1+_4) + p_{2,4}(+_2+_4) \\ &\qquad - p_{2,3}(+_2+_3) - p(+_1|1) - p(+_4|4) \\ &\quad < \Delta^+_{1,3} + \Delta^+_{1,4} + \Delta^+_{2,4} + \Delta^-_{2,3} + \Delta^- + \Delta^-. \end{aligned} \tag{55}$$

With $p(ab) = \frac{1}{4}$ and $p(a) = p(b) = \frac{1}{2}$, one has

$$\Delta^-_{a,b} = 4\sqrt{\varepsilon} \quad \text{and} \quad \Delta^+_{a,b} = 20\sqrt{\varepsilon} - 8\varepsilon. \tag{56}$$

[9] These violations are also maximal in that no other quantum mechanical two-particle state for two spin-$\frac{1}{2}$-particles yields a larger violation (see Cabello, 2002; Tsirelson (Cirel'son), 1980).

With the chosen angles and measurement probabilities one gets

$$\frac{\sqrt{2}-1}{2} < 40\sqrt{\varepsilon} - 12\varepsilon \tag{57}$$

for the lower bound. This inequality is violated for

$$\varepsilon \leqslant \varepsilon^{\mathrm{l}}_{\max}, \quad \varepsilon^{\mathrm{l}}_{\max} = 2.689 \cdot 10^{-5}. \tag{58}$$

The inequality for the upper bound reads

$$\frac{\sqrt{2}-1}{2} < 66\sqrt{\varepsilon} - 24\varepsilon, \tag{59}$$

which is violated for

$$\varepsilon \leqslant \varepsilon^{\mathrm{u}}_{\max}, \quad \varepsilon^{\mathrm{u}}_{\max} = 9.869 \cdot 10^{-6}. \tag{60}$$

Thus the quantum mechanical predictions contradict the predictions of a hidden-variable model obeying our assumptions for

$$\varepsilon \leqslant \varepsilon^{\mathrm{l}}_{\max} = 2.689 \cdot 10^{-5}. \tag{61}$$

4. Discussion

Even though the four sets of assumptions in Bell (1964), Clauser et al. (1969), Bell (1971) and Graßhoff et al. (2005) differ (see Fig. 1), they all imply the same constraints on the correlations as expressed in the Clauser–Horne inequality.[10]

One of the questions left open by Graßhoff et al. (2005) is what constraints are implied *without* assuming the existence of perfectly correlated events. Since with a slightest deviation from perfect correlations the proof by Graßhoff et al. (2005) breaks down, it gives no hint as to whether the same Bell-type constraints follow nor whether a contradiction to QM is entailed at all. In the present paper we have given a partial answer to that question.

The inequality we get at the end of our derivation is stronger than the quantum mechanical predictions (for $\varepsilon \leqslant \varepsilon^{\mathrm{l}}_{\max}$), but weaker than the Clauser–Horne inequality (for $\varepsilon > 0$). Thus, the weakening of the assumptions is also reflected in a resulting weakening of the constraints. Note however that we did not prove that this weakening is really an implication of our assumptions. What we have shown is only that the conditional probabilities *at least* have to obey the constraint (52).[11] To prove the stronger proposition, one could try to construct a separate common cause model obeying all our assumptions that violates the Clauser–Horne inequality without correction terms, but does not violate the weak Clauser–Horne inequality.

$\varepsilon^{\mathrm{l}}_{\max}$ is so small that a deviation of that amount from perfect correlations cannot be ruled out experimentally at present. This means that we cannot rule out the existence of an empirically adequate hidden-variable model obeying all our assumptions. On the other

[10] Although the derived inequalities have a different form in Bell (1964), Bell (1971) and Graßhoff et al. (2005), the respective assumptions are sufficient to derive also the Clauser–Horne inequality.

[11] This proviso is also necessary, because some steps can be optimised in our derivation. For example, one can choose the borders of the partitions in (63) and (78) differently, such that one would get tighter constraints, that is the correction terms to the Clauser–Horne inequality (the Δ's) would become smaller.

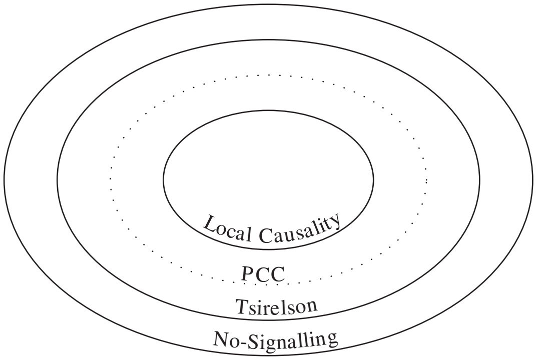

Fig. 3. Comparison with other constraints.

hand, from a theoretical point of view, a deviation from perfect correlation of order $\varepsilon^{l}_{\max}$ is rather large. Modulo some theoretical assumptions, any non-vanishing ε can be interpreted as a violation of rotation invariance (see Section 3.2), which moreover induces a violation of Lorentz invariance.[12] Triggered by theoretical work in various approaches to quantum gravity, which either imply violations of Lorentz invariance or render such a violation natural, there has been a tremendous experimental effort for finding signatures of such violations during the last 10 years or so (for a recent review see e.g. Mattingly, 2005). The constraints coming from negative results of such experiments are rather strong. In view of these findings, one would expect ε to be smaller than $\varepsilon^{l}_{\max}$ and the inequality (52) to be violated.

To conclude we would like to compare our inequality to other prominent constraints.[13] Even though Bell's theorem excludes models which obey LC, the predictions of QM for $p(A_aB_b|ab)$ still obey the no-signalling constraint[14]

$$\sum_{B_b} p(A_aB_b|ab) = \sum_{B_b} p(A_aB_b|ab'),$$
$$\sum_{A_a} p(A_aB_b|ab) = \sum_{A_{a'}} p(A_{a'}B_b|a'b), \tag{62}$$

which states that the probability of the measurement outcome on one side does not depend on the measurement direction on the other side (given the quantum mechanical state of the system). Moreover, there are some correlations obeying no-signalling which are not permitted by QM. The bounds, which are allowed by QM, were first derived by Tsirelson (Cirel'son), 1980. Notoriously, still more constraining are the Bell inequalities. The situation is drawn schematically in Fig. 3. The outmost border represents the least constraining bound coming from the no-signalling condition (62). Next is the Tsirelson-bound, which is again weaker than the bound coming from the Clauser–Horne inequality (LC). The bound coming from inequality (52) lies between the Tsirelson-bound and the

[12] Whether or not a violation under rotation invariance implies also a violation of Lorentz boost invariance is model dependent (see Mattingly, 2005).

[13] For an overview, see e.g. Gisin (2005).

[14] See e.g. Redhead (1987, pp. 113–117) and references therein.

bound coming from the Clauser–Horne inequality, depending on the value of ε. For $\varepsilon \neq 0$ there are quantum mechanical states which do violate the Clauser–Horne inequality but not (52). This reveals the following a priori possibility. As Gisin (1991) showed, the correlations coming from pure entangled states always violate the Clauser–Horne inequality. For $\varepsilon \neq 0$, Gisin's argument is not sufficient to conclude that all entangled states violate inequality (52). Hence, it is an open question, whether or not there exist models obeying all our assumptions, for the correlations of some entangled pure states.

Acknowledgements

We would like to thank Nicolas Gisin, Gerd Graßhoff, Stephanie Kurmann, Peter Minkowski and the audience of the "Workshop Philosophy of Physics", Lausanne, 14th March 2006, for discussions, Gábor Hofer-Szabó for correspondence in connection with this article, and Michael Baumgartner and Tim Räz for discussions and proofreading. We are also grateful to two referees for their comments.

Appendix A. Proof of Proposition 1

We will partition I into the following three subsets:

$$\begin{aligned} &I_1 := \{i \in I : p(A|C_i) \geqslant 1 - \sqrt{\varepsilon}\}, \\ &I_2 := \{i \in I : \sqrt{\varepsilon} < p(A|C_i) < 1 - \sqrt{\varepsilon}\}, \\ &I_3 := \{i \in I : p(A|C_i) \leqslant \sqrt{\varepsilon}\}. \end{aligned} \tag{63}$$

$p(A)$ can be decomposed in three terms:

$$p(A) = \sum_{i \in I_1} p(A|C_i)p(C_i) + \sum_{i \in I_2} p(A|C_i)p(C_i) + \sum_{i \in I_3} p(A|C_i)p(C_i). \tag{64}$$

With the definitions (63) the following inequalities hold:

$$\begin{aligned} &p(A) \geqslant \sum_{i \in I_1} p(A|C_i)p(C_i) \geqslant (1 - \sqrt{\varepsilon}) \sum_{i \in I_1} p(C_i) \geqslant \sum_{i \in I_1} p(C_i) - \sqrt{\varepsilon}, \\ &p(A) \leqslant \sum_{i \in I_1} p(C_i) + \sum_{i \in I_2} p(A|C_i)p(C_i) + \sqrt{\varepsilon}. \end{aligned} \tag{65}$$

Hence, to complete the proof we have to show that

$$\sum_{i \in I_2} p(A|C_i)p(C_i) < 3\sqrt{\varepsilon} - 2\varepsilon. \tag{66}$$

To start with we note that

$$\begin{aligned} \frac{\varepsilon}{2} = \underbrace{p(A)}_{\frac{1}{2}} - \underbrace{p(AB)}_{\frac{1-\varepsilon}{2}} &= \sum_{i \in I} [p(A|C_i) - p(AB|C_i)]p(C_i) \\ &\overset{(*)}{=} \sum_{i \in I} [p(A|C_i) - p(A|C_i)p(B|C_i)]\, p(C_i) \\ &= \sum_{i \in I} p(A|C_i)[1 - p(B|C_i)]\, p(C_i), \end{aligned} \tag{67}$$

where we used screening-off (27) to get equality (∗). Since everything is symmetric in A and B, the same holds if one exchanges A and B for each other. We thus have

$$\sum_{i\in I} p(A|C_i)[1-p(B|C_i)]\,p(C_i)=\frac{\varepsilon}{2}, \tag{68}$$

$$\sum_{i\in I} p(B|C_i)[1-p(A|C_i)]\,p(C_i)=\frac{\varepsilon}{2}. \tag{69}$$

Since all terms in the sums on the left-hand side of Eqs. (68) and (69) are positive, the following inequalities hold for all subsets $I^{\subset}$ of the value space I:

$$0\leqslant\sum_{i\in I^{\subset}} p(A|C_i)[1-p(B|C_i)]\,p(C_i)\leqslant\frac{\varepsilon}{2},\quad \forall I^{\subset}\subset I, \tag{70}$$

$$0\leqslant\sum_{i\in I^{\subset}} p(B|C_i)[1-p(A|C_i)]\,p(C_i)\leqslant\frac{\varepsilon}{2},\quad \forall I^{\subset}\subset I. \tag{71}$$

Subtracting (70) from (71), one gets

$$\left|\sum_{i\in I^{\subset}}[p(A|C_i)-p(B|C_i)]p(C_i)\right|\leqslant\frac{\varepsilon}{2},\quad \forall I^{\subset}\subset I. \tag{72}$$

With the definitions

$$I_2^{A\geqslant B}{:=}\{i\in I_2: p(A|C_i)\geqslant p(B|C_i)\}, \tag{73}$$

$$I_2^{A<B}{:=}\{i\in I_2: p(A|C_i)< p(B|C_i)\} \tag{74}$$

and applying (72) for these sets, one gets

$$\left|\sum_{i\in I_2^{A\geqslant B}}[p(A|C_i)-p(B|C_i)]\,p(C_i)\right|=\sum_{i\in I_2^{A\geqslant B}}|p(A|C_i)-p(B|C_i)|p(C_i)\leqslant\frac{\varepsilon}{2}, \tag{75}$$

$$\left|\sum_{i\in I_2^{A<B}}[p(A|C_i)-p(B|C_i)]\,p(C_i)\right|=\sum_{i\in I_2^{A<B}}|p(A|C_i)-p(B|C_i)|p(C_i)\leqslant\frac{\varepsilon}{2}. \tag{76}$$

Adding these two inequalities, one gets

$$\sum_{i\in I_2}|p(A|C_i)-p(B|C_i)|p(C_i)\leqslant\varepsilon. \tag{77}$$

We partition I_2 in the following two subsets:

$$I_2^{\geqslant\sqrt{\varepsilon}}{:=}\left\{i\in I_2\colon|\,p(A|C_i)-p(B|C_i)|\geqslant\frac{\sqrt{\varepsilon}}{2}\right\},$$

$$I_2^{<\sqrt{\varepsilon}}{:=}\left\{i\in I_2\colon|\,p(A|C_i)-p(B|C_i)|<\frac{\sqrt{\varepsilon}}{2}\right\}. \tag{78}$$

From

$$\sum_{i\in I_2^{\geqslant\sqrt{\varepsilon}}}|p(A|C_i)-p(B|C_i)|p(C_i)\geqslant\frac{\sqrt{\varepsilon}}{2}\sum_{i\in I_2^{\geqslant\sqrt{\varepsilon}}}p(C_i) \tag{79}$$

together with (77), we get

$$\sum_{i \in I_2^{\geqslant \sqrt{\varepsilon}}} p(C_i) \leqslant 2\sqrt{\varepsilon}. \tag{80}$$

Remember that we want to derive an upper bound for

$$\sum_{i \in I_2} p(A|C_i)p(C_i). \tag{81}$$

With (80), we already have

$$\begin{aligned}\sum_{i \in I_2} p(A|C_i)p(C_i) &= \sum_{i \in I_2^{\geqslant \sqrt{\varepsilon}}} p(A|C_i)p(C_i) + \sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)p(C_i) \\ &< (1-\sqrt{\varepsilon})2\sqrt{\varepsilon} + \sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)p(C_i).\end{aligned} \tag{82}$$

We will use again inequality (70), this time for the set $I_2^{<\sqrt{\varepsilon}}$:

$$\sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)[1-p(B|C_i)]p(C_i) \leqslant \frac{\varepsilon}{2}. \tag{83}$$

Because we are looking at the subset $I_2^{<\sqrt{\varepsilon}}$, it is

$$\sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)[1-p(B|C_i)]p(C_i) > \sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)\left[1-p(A|C_i)-\frac{\sqrt{\varepsilon}}{2}\right]p(C_i). \tag{84}$$

With (83), one gets

$$\sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)\left[1-p(A|C_i)-\frac{\sqrt{\varepsilon}}{2}\right]p(C_i) < \frac{\varepsilon}{2}. \tag{85}$$

Now, since $I_2^{<\sqrt{\varepsilon}}$ is a subset of I_2, $p(A|C_i)$ takes on values in the interval $[\sqrt{\varepsilon}, 1-\sqrt{\varepsilon}]$. One can check that each summand is certainly greater than for $p(A|C_i) = 1-\sqrt{\varepsilon}$. We thus have

$$\sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(A|C_i)\left[1-p(A|C_i)-\frac{\sqrt{\varepsilon}}{2}\right]p(C_i) > (1-\sqrt{\varepsilon})\frac{\sqrt{\varepsilon}}{2}\sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(C_i). \tag{86}$$

We get the constraint

$$\sum_{i \in I_2^{<\sqrt{\varepsilon}}} p(C_i) < \frac{\sqrt{\varepsilon}}{(1-\sqrt{\varepsilon})}. \tag{87}$$

With (82) one gets

$$\sum_{i \in I_2} p(A|C_i)p(C_i) < (1-\sqrt{\varepsilon})2\sqrt{\varepsilon} + \sqrt{\varepsilon} = 3\sqrt{\varepsilon} - 2\varepsilon, \tag{88}$$

which is what we wanted to show. We thus arrive at the conclusion of Proposition 1:

$$\sum_{i \in I_1} p(C_i) - \sqrt{\varepsilon} \leqslant p(A) < \sum_{i \in I_1} p(C_i) + 4\sqrt{\varepsilon} - 2\varepsilon. \qquad \square \tag{89}$$

Appendix B. Bounds for $p(+_a +_b|ab)$

With (39) and Assumption 5, we get

$$\begin{aligned} p(+_a|a) &= p(+_a C^a|a) + p(+_a \overline{C^a}|a) < p(C^a|a) + 4\sqrt{\varepsilon} - 2\varepsilon \\ &= p(+_a C^a|a) + p(\overline{+_a} C^a|a) + 4\sqrt{\varepsilon} - 2\varepsilon, \end{aligned} \tag{90}$$

and hence

$$p(+_a a\overline{C^a}) < p(\overline{+_a} a C^a) + p(a)(4\sqrt{\varepsilon} - 2\varepsilon). \tag{91}$$

Furthermore, we have

$$\begin{aligned} p(\overline{+_a} C^a|a) &\overset{(*)}{=} \sum_{i\in I_1^a} p(\overline{+_a}|aC_i^a)p(C_i^a) \\ &= \sum_{i\in I_1^a} (1 - p(+_a|aC_i^a))p(C_i^a) \\ &\leqslant \sqrt{\varepsilon} \sum_{i\in I_1^a} p(C_i^a) \leqslant \sqrt{\varepsilon}, \end{aligned} \tag{92}$$

where we used Assumption 5 to get equality $(*)$. We can write (92) as

$$p(\overline{+_a} a C^a) \leqslant p(a)\sqrt{\varepsilon}, \tag{93}$$

such that we get from (91)

$$p(+_a a\overline{C^a} X) < p(a)(5\sqrt{\varepsilon} - 2\varepsilon), \tag{94}$$

because for any X and any Y, $p(XY) \leqslant p(Y)$. Next, from

$$p(+_a a C^a X) = p(aC^a X) - p(\overline{+_a} a C^a X) \tag{95}$$

together with (93) and because $p(\overline{+_a} a C^a X) \leqslant p(\overline{+_a} a C^a)$ we get

$$p(+_a a C^a X) \geqslant p(aC^a X) - p(a)\sqrt{\varepsilon}. \tag{96}$$

Starting from the inequalities (40), we get inequalities analogue to (94) and (96), that is we get

$$p(+_a a\overline{C^a} X) < p(a)(5\sqrt{\varepsilon} - 2\varepsilon), \tag{97}$$

$$p(+_b b\overline{C^b} X) < p(b)(5\sqrt{\varepsilon} - 2\varepsilon), \tag{98}$$

$$p(+_a a C^a X) \geqslant p(aC^a X) - p(a)\sqrt{\varepsilon}, \tag{99}$$

$$p(+_b b C^b X) \geqslant p(bC^b X) - p(b)\sqrt{\varepsilon}. \tag{100}$$

Now, we can derive an upper bound for $p(+_a +_b ab)$, using (97) and (98):

$$\begin{aligned} p(+_a +_b ab) &= p(+_a +_b abC^a) + p(+_a +_b ab\overline{C^a}) \\ &< p(+_a +_b abC^a) + p(a)(5\sqrt{\varepsilon} - 2\varepsilon) \\ &< p(+_a +_b abC^a C^b) + (p(a) + p(b))(5\sqrt{\varepsilon} - 2\varepsilon) \\ &\leqslant p(abC^a C^b) + (p(a) + p(b))(5\sqrt{\varepsilon} - 2\varepsilon). \end{aligned} \tag{101}$$

Using also Assumption 5, we finally obtain

$$
\begin{aligned}
p(+_a +_b | ab) &\equiv \frac{p(+_a +_b ab)}{p(ab)} \\
&< \frac{p(abC^a C^b) + (p(a) + p(b))(5\sqrt{\varepsilon} - 2\varepsilon)}{p(ab)} \\
&= p(C^a C^b | ab) + (p(a) + p(b)) \frac{5\sqrt{\varepsilon} - 2\varepsilon}{p(ab)} \\
&= p(C^a C^b) + (p(a) + p(b)) \frac{5\sqrt{\varepsilon} - 2\varepsilon}{p(ab)}. \qquad (102)
\end{aligned}
$$

Next, we derive a lower bound.

$$
\begin{aligned}
p(+_a +_b ab) &\geqslant p(+_a +_b abC^a C^b) \\
&\geqslant p(+_b abC^a C^b) - p(a)\sqrt{\varepsilon} \\
&\geqslant p(abC^a C^b) - (p(a) + p(b))\sqrt{\varepsilon}, \qquad (103)
\end{aligned}
$$

$$
\begin{aligned}
p(+_a +_b | ab) &\equiv \frac{p(+_a +_b ab)}{p(ab)} \\
&\geqslant p(C^a C^b) - \frac{(p(a) + p(b))\sqrt{\varepsilon}}{p(ab)}. \qquad (104)
\end{aligned}
$$

(102) and (104) imply

$$
p(+_a +_b | ab) - \Delta^+_{a,b} < p(C^a C^b) \leqslant p(+_a +_b | ab) + \Delta^-_{a,b} \qquad (105)
$$

with

$$
\begin{aligned}
\Delta^-_{a,b} &= \frac{(p(a) + p(b))\sqrt{\varepsilon}}{p(ab)}, \\
\Delta^+_{a,b} &= \frac{(p(a) + p(b))(5\sqrt{\varepsilon} - 2\varepsilon)}{p(ab)}. \qquad (106)
\end{aligned}
$$

Appendix C. Proof of the Clauser–Horne inequality

In this appendix we prove inequality (50). We consider four arbitrary events A, A', B, and B' together with their complements. The sum over all 16 possibilities equals 1:

$$
\sum_{a,a',b,b'} p(a, a', b, b') = 1, \quad \text{where} \quad a \in \{A, \bar{A}\}, \text{ etc.} \qquad (107)
$$

We also have

$$
\begin{aligned}
p(AB) &= p(AA'BB') + p(AA'B\bar{B}') + p(A\bar{A}'BB') + p(A\bar{A}'B\bar{B}'), \\
p(AB') &= p(AA'BB') + p(A\bar{A}'BB') + p(AA'\bar{B}B') + p(A\bar{A}'\bar{B}B'), \\
p(A'B') &= p(AA'BB') + p(AA'\bar{B}B') + p(\bar{A}A'BB') + p(\bar{A}A'\bar{B}B'),
\end{aligned}
$$

$$
\begin{aligned}
p(A'B) &= p(AA'BB') + p(AA'B\bar{B}') + p(\bar{A}A'BB') + p(\bar{A}A'B\bar{B}'),\\
p(A) &= p(AA'BB') + p(AA'B\bar{B}') + p(A\bar{A}'BB') + p(A\bar{A}'B\bar{B}')\\
&\quad + p(AA'\bar{B}B') + p(AA'\bar{B}\bar{B}') + p(A\bar{A}'\bar{B}B') + p(A\bar{A}'\bar{B}\bar{B}'),\\
p(B') &= p(AA'BB') + p(A\bar{A}'BB') + p(AA'\bar{B}B') + p(A\bar{A}'\bar{B}B')\\
&\quad + p(\bar{A}A'BB') + p(\bar{A}\bar{A}'BB') + p(\bar{A}A'\bar{B}B') + p(\bar{A}\bar{A}'\bar{B}B').
\end{aligned} \tag{108}
$$

Thus,

$$
\begin{aligned}
&p(AB) + p(AB') + p(A'B') - p(A'B) - p(A) - p(B')\\
&\quad = -[p(AA'B\bar{B}') + p(AA'\bar{B}\bar{B}') + p(A\bar{A}'\bar{B}B') + p(A\bar{A}'\bar{B}\bar{B}')\\
&\qquad + p(\bar{A}A'BB') + p(\bar{A}A'B\bar{B}') + p(\bar{A}\bar{A}'BB') + p(\bar{A}\bar{A}'\bar{B}B')].
\end{aligned} \tag{109}
$$

Because each term appears only once on the right-hand side of Eq. (109), Eq. (107) implies the Clauser–Horne inequality:

$$
-1 \leqslant p(AB) + p(AB') + p(A'B') - p(A'B) - p(A) - p(B') \leqslant 0. \tag{110}
$$

References

Aspect, A., Dalibard, J., & Roger, G. (1982). Experimental test of Bell's inequality using time-varying analyzers. *Physical Review Letters*, *49*(25), 1804–1807.

Bell, J. S. (1964). On the Einstein–Podolsky–Rosen paradox. *Physics*, *1*, 195 (Reprinted in Bell , 1987, pp. 14–21).

Bell, J.S. (1971). Introduction to the hidden-variable question. In *Foundations of quantum mechanics* (p. 171ff). New York: Academic Press. (Reprinted in Bell, 1987, pp. 29–39).

Bell, J.S. (1975). *The theory of local beables*. TH-2053-CERN (Presented at the sixth GIFT seminar, Jaca, 2–7 June 1975, reproduced in *Epistemological Letters*, March 1976, and reprinted in Bell, 1987, pp. 52–62).

Bell, J. S. (1987). *Speakable and unspeakable in quantum mechanics*. Cambridge: Cambridge University Press.

Belnap, N., & Szabó, L. (1996). Branching space-time analysis of the GHZ theorem. *Foundations of Physics*, *26*, 989–1002.

Bohm, D. (1951). *Quantum theory*. New York: Prentice-Hall.

Cabello, A. (2002). Violating Bell's inequality beyond Cirel'son's bound. *Physical Review Letters*, *88*, 060403.

Clauser, J., & Horne, M. (1974). Experimental consequences of objective local theories. *Physical Review D*, *10*, 526–535.

Clauser, J., Horne, M., Shimony, A., & Holt, R. (1969). Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, *23*, 880–884.

Clauser, J., & Shimony, A. (1978). Bell's theorem: Experimental tests and implications. *Reports on Progress in Physics*, *41*, 1881–1927.

Clifton, R., & Kent, A. (2000). Simulating quantum mechanics by non-contextual hidden variables. In *Proceedings of the Royal Society A* 456, 2101–2114 ⟨http://xxx.lanl.gov/abs/quant-ph/9908031⟩.

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

Gisin, N. (1991). Bell's inequality holds for all non-product states. *Physics Letters A*, *154*, 201–202.

Gisin, N. (2005). *Can relativity be considered complete? From Newtonian nonlocality to quantum nonlocality and beyond* (URL: ⟨http://arxiv.org/abs/quant-ph/0512168⟩).

Graßhoff, G., Portmann, S., & Wüthrich, A. (2005). Minimal assumption derivation of a Bell-type inequality. *British Journal for the Philosophy of Science*, *56*, 663–680 ⟨http://lanl.arxiv.org/abs/quant-ph/0312176⟩.

Hofer-Szabó, G. (2007). *Separate- versus common-common-cause-type derivations of the Bell inequalities*, submitted for publication.

Hofer-Szabó, G., & Rédei, M. (2004). Reichenbachian common cause systems. *International Journal of Theoretical Physics*, *43*, 1819–1826 ⟨http://philsci-archive.pitt.edu/archive/00001246/⟩.

Hofer-Szabó, G., Rédei, M., & Szabó, L. E. (1999). On Reichenbach's common cause principle and Reichenbach's notion of common cause. *British Journal for the Philosophy of Science*, *50*(3), 377–399 ⟨http://xxx.lanl.gov/abs/quant-ph/9805066⟩.

Hofer-Szabó, G., Rédei, M., & Szabó, L. E. (2002). Common-causes are not common common-causes. *Philosophy of Science*, *69*, 623–636.

Jarrett, J. P. (1984). On the physical significance of the locality conditions in the Bell arguments. *Noûs*, *18*, 569–589.

Mattingly, D. (2005). Modern tests of Lorentz invariance. *Living Reviews in Relativity*, 8. URL cited on 9 February 2006: ⟨http://www.livingreviews.org/lrr-2005-5⟩.

Maudlin, T. (1994). *Quantum non-locality and relativity*. Cambridge: Blackwell.

Pitowsky, I. (1989). *Quantum probability—quantum logic (No. 321)*. Berlin: Springer.

Rédei, M. (1997). Reichenbach's common cause principle and quantum field theory. *Foundations of Physics*, *27*, 1309–1321.

Redhead, M. (1987). *Incompleteness, nonlocality and realism*. Oxford: Clarendon Press.

Reichenbach, H. (1956). *The direction of time*. Los Angeles: University of California Press.

Ryff, L. C. (1997). Bell and Greenberger, Horne, and Zeilinger theorems revisited. *American Journal of Physics*, *65*(12), 1197–1199.

Shimony, A. (2005). Bell's theorem. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (summer 2005 edition)* ⟨http://plato.stanford.edu/archives/sum2005/entries/bell-theorem⟩.

Suppes, P., & Zanotti, M. (1976). On the determinism of hidden variable theories with strict correlation and conditional statistical independence of observables. In P. Suppes (Ed.), *Logic and probability in quantum mechanics* (pp. 445–455). Dordrecht: Reidel.

Tsirelson (Cirel'son), B. (1980). Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, *4*, 93–100.

van Fraassen, B. C. (1982). The Charybdis of realism: Epistemological implications of Bell's inequalities. *Synthese*, *52*, 25–38.

Weinstein, S. (2006). Superluminal signaling and relativity. *Synthese*, *148*, 381–399.