A Stronger Bell Argument for Quantum Non-Locality

Paul M. Näger

Department of Philosophy, University of Münster

Domplatz 6, D-48143 Münster, Germany

paul.naeger@uni-muenster.de

phone: +49 (0)251 83-24339

16th July 2015
Version 4.97

It is widely accepted that the violation of Bell inequalities excludes local theories of the quantum realm. This paper presents a stronger Bell argument which even forbids certain non-local theories. Among these excluded non-local theories are those whose *only* non-local connection is a probabilistic (or functional) dependence between the space-like separated measurement outcomes of EPR/B experiments (a subset of outcome dependent theories). In this way, the new argument shows that the result of the received Bell argument, which requires just any kind of non-locality, is inappropriately weak. Positively, the remaining non-local theories, which can violate Bell inequalities (among them quantum theory), are characterized by the fact that at least one of the measurement outcomes in some sense probabilistically depends both on its local as well as on its distant measurement setting (probabilistic Bell contextuality). Whether an additional dependence between the outcomes holds, is irrelevant for the question whether a certain theory can violate Bell inequalities. This new concept of quantum non-locality is considerably tighter and more informative than the one following from the usual Bell argument. We prove that (given usual background assumptions) the result of the stronger Bell argument presented here is the strongest possible consequence from the violation of Bell inequalities on a qualitative probabilistic level.

Contents

1 Introduction

Bell's argument (1964; 1971; 1975) establishes a mathematical no-go theorem for theories of the micro-world. In its standard form, it derives that theories which are local (and fulfill certain auxiliary assumptions) cannot have correlations of arbitrary strength between events which are space-like separated. An upper bound for the correlations is given by the famous Bell inequalities. Since certain experiments with entangled quantum objects have results which violate these inequalities (EPR/B correlations), it concludes that the quantum realm cannot be described by a local theory. Any correct theory of the quantum realm must involve some kind of non-locality, a 'quantum non-locality'. This result is one of the central features of the quantum realm. It is the starting point for extensive debates concerning the nature of quantum objects and their relation to space and time.

Since Bell's first proof (1964) the theorem has evolved considerably towards stronger forms: there has been a sequence of improvements which derive the inequalities from weaker and weaker assumptions. The main focus has been on getting rid of those premises which are commonly regarded as auxiliary assumptions: Clauser et al. (1969) derived the theorem without assuming perfect correlations; Bell (1971) abandoned the assumption of determinism; Graßhoff et al. (2005) and Portmann and Wüthrich (2007) showed that possible latent common causes do not have to be *common* common causes.[1] What is common to all of these different derivations is that they assume one or another form of locality. Locality seems to be the central assumption in deriving the Bell inequalities—and hence it is the assumption that is assumed to fail when one finds that the inequalities are violated.

In this paper we are going to present another strengthening of Bell's theorem, which relaxes the central assumption: one does not have to assume locality in order to derive

[1] The debate about common common causes vs. separate common causes is to some degree still undecided (cf. Hofer-Szabó 2008).

the Bell inequalities. Certain forms of non-locality, which we shall call 'weakly non-local' suffice: an outcome may depend on the other outcome or on the distant setting—as long as it does not depend on *both* settings, it still implies that the Bell inequalities hold. As a consequence, the violation of the Bell inequalities also excludes those weakly non-local theories. So it does not require *any* kind of non-locality, but a very *specific* one: at least one of the outcomes must depend probabilistically on both settings. While previous strengthenings of Bell's theorem secured that a certain auxiliary assumption is *not* the culprit, our derivation here for the first time strengthens the *conclusion* of the theorem.

Being at its core a mathematical theorem, its direct conclusion is stated in formal terms as well. From the formal result that there are certain non-local probabilistic dependences, typically far reaching conclusions about the existence of certain non-local physical or metaphysical connections are drawn, e.g. a non-separability or non-local causal relations. Since these latter inferences require further assumptions and are far from being trivial (especially they cannot reliably be made en passant), in this paper we shall constrain to establish a strengthening of Bell's core argument on the mathematical level (and leave an appropriate physical and metaphysical interpretation of this result for future work).

We start by introducing an appropriate notation for the underlying experiments with entangled photons and formulate the standard Bell argument in an explicit and clear form (section 2). We then develop a stronger Bell argument in three steps. In section 3 we prove that an important class of non-local theories, viz. purely outcome dependent theories (a subclass of outcome dependent theories according to which the *only* non-local connection is between the outcomes), implies Bell inequalities. Subsequently, we generalize the proof to as many classes as possible, which requires to introduce a comprehensive overview of all logically possible classes (section 4). We call those non-local classes which allow a derivation 'weakly non-local' and elaborate their characteristics. In section 5 we formulate the new stronger Bell argument. A second generalization extends the argument to another partition of the logically possible classes (section 6). Finally, we show that given usual background assumptions our new stronger Bell argument provides the strongest possible consequences from the violation of Bell inequalities on a qualitative probabilistic level (section 7) and discuss some of its immediate consequences (section 8).

2 EPR/B experiments and the standard Bell argument

We consider a usual EPR/B setup with space-like separated polarization measurements of an ensemble of photon pairs in an entangled quantum state $\boldsymbol{\psi} = \psi_0$ (Einstein, Podolsky, and Rosen 1935; Bohm 1951; Clauser and Horne 1974; see fig. 1). Possible hidden variables of the photon pairs are called $\boldsymbol{\lambda}$, so that the *complete* state of the particles at the source is $(\boldsymbol{\psi}, \boldsymbol{\lambda})$. Since in this setup the state $\boldsymbol{\psi}$ is the same in all runs, it will not explicitly be noted in the following (one may think of any probability being conditional on one fixed state $\boldsymbol{\psi} = \psi_0$). We denote Alice's and Bob's measurement setting as $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively, and the corresponding (binary) measurement results as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. On a probabilistic level, the experiment is described by the joint probability distribution

$P(\alpha\beta ab\lambda) := P(\boldsymbol{\alpha} = \alpha, \boldsymbol{\beta} = \beta, \boldsymbol{a} = a, \boldsymbol{b} = b, \boldsymbol{\lambda} = \lambda)$ of these five random variables.[2] We shall consistently use bold symbols $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}, \dots)$ for random variables and normal font symbols $(\alpha, \beta, a, \dots)$ for the corresponding values of these variables. We use indices to refer to *specific* values of variables, e.g. $\alpha_- = -$ or $a_1 = 1$, which provides useful shorthands, e.g. $P(\alpha_-\beta_+a_1b_2\lambda) := P(\boldsymbol{\alpha} = -, \boldsymbol{\beta} = +, \boldsymbol{a} = 1, \boldsymbol{b} = 2, \boldsymbol{\lambda} = \lambda)$. Expressions including probabilities with non-specific values of variables, e.g. $P(\alpha|a) = P(\alpha)$, are meant to hold for all values of these variables (if not otherwise stated).

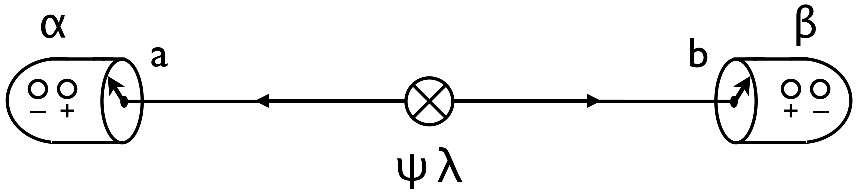

Figure 1: EPR/B setup

Containing the hidden states $\boldsymbol{\lambda}$, which are by definition not measurable, the total distribution is empirically not accessible ('hidden level'), i.e. purely theoretical. Only the marginal distribution which does not involve $\boldsymbol{\lambda}$, $P(\alpha\beta ab)$, *is* empirically accessible and is determined by the results of actual measurements in EPR/B experiments ('observable level'). A statistical evaluation of a series of many runs with similar preparation procedures yields that the outcomes are strongly correlated given the settings and the quantum state.[3] For instance, in case the quantum state is the Bell state $\psi_0 = (|+\rangle|+\rangle + |-\rangle|-\rangle)/\sqrt{2}$ (and the settings are chosen with equal probability $\frac{1}{2}$) the correlations read:

$$P(\alpha\beta|ab) = P(\alpha|\beta ab)P(\beta) = \begin{cases} \cos^2(a-b)\cdot\frac{1}{2} & \text{if } \alpha = \beta \\ \sin^2(a-b)\cdot\frac{1}{2} & \text{if } \alpha \neq \beta \end{cases} \tag{Corr}$$

These famous EPR/B correlations between space-like separated measurement outcomes have first been measured by Aspect et al. (1982) and have been confirmed under strict locality conditions (Weihs et al. 1998) as well as over large distances (Ursin et al. 2007). All these findings are correctly predicted by quantum mechanics: involving only empirically accessible variables, the quantum mechanical probability distribution essentially agrees with the empirical one.

Since according to (Corr), one outcome depends on both the other space-like separ-

[2] While the outcomes are discrete variables and the settings can be considered to be discrete (in typical EPR/B experiments there are two possible settings on each side), the hidden state may be continuous or discrete. In the following I assume $\boldsymbol{\lambda}$ to be discrete, but all considerations can be generalized to the continuous case.

[3] A correlation of the outcomes given the settings and the quantum state means that the joint probability $P(\alpha\beta|ab)$ is in general not equal to the product $P(\alpha|ab)P(\beta|ab) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$.

ated outcome as well as on the distant (and local) setting, the observable part of the probability distribution (or the quantum mechanical distribution, respectively) clearly is non-local. Bell's idea (1964) was to show that EPR/B correlations are so extraordinary that even if one allows for hidden states $\boldsymbol{\lambda}$ one cannot restore locality: given EPR/B correlations the *theoretical* probability distribution (including possible hidden states) must be non-local as well. Hence, *any possible* probability distribution which might correctly describe the experiment must be non-local.

This 'Bell argument for quantum non-locality', as I shall call it, proceeds by showing that the empirically measured EPR/B correlations violate certain inequalities, the famous Bell inequalities. It follows that at least one of the assumptions in the derivation of the inequalities must be false. Indeterministic generalizations (Bell 1971; Clauser and Horne 1974; Bell 1975) of Bell's original deterministic derivation (1964) employ two probabilistic assumptions, 'local factorisation'[4]

$$P(\alpha\beta|ab\lambda) = P(\alpha|a\lambda)P(\beta|b\lambda) \tag{ℓF}$$

and 'autonomy'

$$P(\lambda|ab) = P(\lambda). \tag{A}$$

Another type of derivation (Wigner 1970; van Fraassen 1989; Graßhoff et al. 2005) additionally requires the fact that there are perfect correlations (PCorr) between the outcomes for a certain relative angle of the measurement settings (e.g. for parallel settings given quantum state ψ_0). For both types of derivation we have the dilemma that any empirically correct probability distribution of the quantum realm must either violate autonomy or local factorisation (or both). Since giving up autonomy seems to be ad hoc and implausible ('cosmic conspiracy'), most philosophers conclude that the empirical violation of Bell inequalities implies that local factorisation fails.[5] And since local factorisation states the factorisation of the hidden joint probability distribution into *local* terms, the *failure* of local factorisation indicates a certain kind of *non*-locality, which is specific to the quantum realm—hence 'quantum non-locality'.

For my following critique of this standard Bell argument it is important to have a clear account of its logical structure. Here and in the following I shall presuppose the Wigner-type derivation of Bell inequalities because, as we will see, it is the most powerful one allowing to derive Bell inequalities from the widest range of probability distributions:

(P1) There are EPR/B correlations: (Corr)

(P2) EPR/B correlations violate Bell inequalities: (Corr) $\rightarrow \neg$(BI)

[4] 'Local factorisation' is my term. Bell calls (ℓF) 'local causality', some call it 'Bell-locality', but most often it is simply called 'factorisation' or 'factorisability'. Bell's terminology already suggests a causal interpretation, which I would like to avoid in this paper, and the latter two names are too general since, as I shall show, there are other forms of the hidden joint probability which can be said to factorise; hence '*local* factorisation'.

[5] Though not a majority view, there are suggestions to explain the violation of Bell inequalities by a violation of autonomy (e.g. Sutherland 1983; San Pedro 2012). Our analysis in this paper does not apply to these cases; we shall consistently assume that autonomy holds.

(P3) EPR/B correlations include perfect correlations: (Corr) $\rightarrow$ (PCorr)

(P4) Bell inequalities can be derived from autonomy, perfect correlations and local factorisation: (A) $\wedge$ (PCorr) $\wedge$ (ℓF) $\rightarrow$ (BI)

(P5) Autonomy holds: (A)

(C1) Local factorisation fails: $\neg$(ℓF) (from P1–P5)

The core idea of my critique concerning this standard Bell argument for quantum non-locality is that its result is considerably weaker than it could be. I do not say that the argument is invalid (it is obviously not) nor do I say that one of its premises is not sound, I just say that the argument can be made considerably stronger and that the stronger conclusion will provide a tighter, more informative concept of quantum non-locality: one can be much more precise about what EPR/B correlations imply (if we assume that autonomy holds) than just saying that local factorisation has to fail.

Specifically, I shall show that it is premise (P4) which can be made stronger. Stating that autonomy, perfect correlations and local factorisation imply the Bell inequalities, it is clear that one can make (P4) the stronger the weaker one can formulate the antecedent, i.e. the assumptions to derive the inequalities. Former improvements have concentrated on relaxing assumptions *except* the locality condition. In contrast, here I shall try to find *weaker alternatives to local factorisation*, which also imply that Bell inequalities hold. Since local factorization is the weakest possible form of *local* distributions, it is clear that such alternatives have to involve a kind of *non-locality*, i.e. what I am trying to show in the following is that we can derive Bell inequalities from certain non-local probability distributions. This will make the overall argument stronger for it will allow for the conclusion that not only local theories but also those non-local ones that imply the inequalities are ruled out. I shall now first demonstrate this for one central class of non-local probability distributions, before in the subsequent section I consider the general case.

3 Bell inequalities from purely outcome dependent theories

Local factorisation is a specific product form of the *hidden joint probability* of the outcomes, as I shall call $P(\alpha\beta|ab\lambda)$.[6] A prominent non-local product form of this hidden joint probability is the following:

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta a\lambda)P(\beta|b\lambda) \qquad (\mathrm{H}^{\alpha}_{16})$$

(For reasons that will become clear later the product form is tagged (H^{α}_{16}).) It differs from local factorisation in that it involves the distant outcome $\boldsymbol{\beta}$ in the first factor on the right hand side, which makes it a *non-local* product form (at least one of the factors

[6] 'Hidden' because the probability is conditional on the hidden state λ and thus is not empirically accessible.

involves at least one variable that is space-like separated to the respective outcome). Since product forms characterize probability distributions, which represent a whole class of theories, (H_{16}^{α}) represents a class of non-local theories. In the debate about Bell's theorem such theories with a non-local dependence between the outcomes in the product form are usually called *outcome dependent.* They represent physical theories according to which the outcomes are probabilistically or functionally dependent on another. This dependence between the space-like separated outcomes has emerged as the received view of what the violation of Bell inequalities amounts to: adequate theories of the quantum realm are widely believed to be correctly described as outcome dependent theories.

In this section I shall prove that theories having the product form (H_{16}^{α}) are not consistent with the results of EPR/B experiments. In order to avoid misunderstandings, it is important to stress three central facts already at the outset of the argument. First, (H_{16}^{α}) is only one of several possible outcome dependent classes. So proving that (H_{16}^{α}) is impossible does not rule out all outcome dependent theories, but only very specific ones. For instance, the quantum mechanical distribution, which is well-known to be outcome dependent, is not correctly described by (H_{16}^{α}), but rather has the product form

$$P(\alpha\beta|ab) = P(\alpha|\beta ab)P(\beta), \tag{1}$$

i.e. according to quantum mechanics the outcome $\boldsymbol{\alpha}$ additionally depends on the distant setting $\boldsymbol{b}$ (and there is no dependence on a hidden variabel $\boldsymbol{\lambda}$). In order to distinguish (H_{16}^{α}) from such other outcome dependent classes I denote it as *purely* outcome dependent. So when in the following we show that purely outcome dependent theories are not consistent with results of EPR/B experiments, this does not mean that quantum mechanics is not consistent with these results.

Second, while not ruling out well established theories, the result still has far reaching implications, because it informs us about the role that outcome dependence plays in the violation of Bell inequalities. Against a widely spread belief, the new result that purely outcome dependent theories *imply* the Bell inequalities shows that outcome dependence *per se* cannot explain the violation of the inequalities. We shall see in later sections that some kind of dependence on the distant setting is required (but not necessarily, as the quantum mechanical example shows, the kind of dependence that usually is called 'parameter dependence').

Finally, we emphasize that the claims we shall be arguing for here are exclusively probabilistic ones: they are about *probabilistic* dependences (between the variables in the setup) and not about physical or metaphysical relations. It is important to stress the difference, because too often correlations are naively interpreted to indicate physical interactions, causal relations or the like. But correlation is not causation, and establishing the latter by the former involves non-trivial inferences, also invoking further assumptions. So saying that outcome dependence (which is by definition a probabilistic dependence) cannot explain the violation of the Bell inequalities is first of all not to say that a physical connection between the outcomes cannot explain the Bell inequalities. And likewise, saying that a violation requires probabilistic dependence of one outcome on both settings does not *per se* say that a physical connection between a setting and its distant outcome is implied. Since inferring (meta-)physical relations from probabilistic facts requires careful discussion, in this paper we shall not have the resources to treat that question as well. Rather we shall concentrate on deriving the probabilistic consequences of the violation of Bell inequalities, and therefore, if not explicitly stated otherwise, when I speak of 'dependence' in general or 'outcome dependence' in particular in the following I always mean probabilistic dependences (correlations), and not physical or metaphysical relations. I might at least remark, without arguing for or elaborating on, that elsewhere I have tried to show that, given reasonable assumptions, a probabilistic dependence between the outcomes of an EPR/B experiment should be interpreted as a genuine (meta-)physical connection, so that a failure of the former to violate the Bell inequalities also indicates a failure of the latter to do so (Näger 2013). That analysis also reveals that the required probabilistic dependence on both settings most plausibly requires a causal connection from both settings to at least one of the outcomes.

After these preliminary remarks, I now turn to my argument against (H^{α}_{16}), which comes in two steps: I first show that if perfect correlations and perfect anti-correlations (and autonomy) hold, (H^{α}_{16}) is straightforwardly impossible. ('Straightforward' here means that the impossibility is not demonstrated via a Bell inequality, but in a more direct way.) This immediate inconsistency vanishes, when one relaxes perfect (anti-)correlations to nearly perfect (anti-)correlations (i.e. (anti-)correlations that show small deviations from perfectness). In this case, however, I demonstrate, second, that (H^{α}_{16}) and autonomy imply Bell inequalities. This is the genuine strengthening of the Bell argument that I have announced. Since the inequalities are empirically violated this also establishes an inconsistency (though a less direct one).

Let me start by putting my first claim in a precise form:

> **Lemma 1**: Autonomy, perfect correlations, perfect anti-correlations and (H^{α}_{16}) are an inconsistent set: (A) $\wedge$ (PCorr) $\wedge$ (PAcorr) $\rightarrow \neg$(H^{α}_{16})

(The proof of this lemma can be found in the appendix.)

Lemma 1 makes the surprising assertion that, given autonomy, purely outcome dependent theories are logically impossible if (for certain measurement settings) perfect correlations and (for certain other measurement settings) perfect anti-correlations hold. As can be seen in the proof of the lemma, the conflict does not require to formulate a Bell inequality: the inconsistency can be established in a much more direct way. Hence, in this case it does not even make sense to try to derive a Bell inequality, since the assumptions that would be needed for the derivation are already inconsistent. For this reason, lemma 1 is not in a literal sense a strengthening of the Bell argument. But since the aim of Bell's argument is to exclude certain theories of the micro-realm one might say that it is an *amendment* to the argument which strengthens its conclusion. The strengthening consists in the fact that lemma 1 precludes certain theories, namely purely outcome dependent ones, that usual Bell arguments do not rule out. What makes this result so remarkable is that the excluded theories are precisely those that have long been regarded to be the received view about quantum non-locality.

A defender of outcome dependence might try to avoid the conflict by asserting that perfect (anti-)correlations are not empirically confirmed. In fact, real experiments fail to yield the perfect (anti-)correlations that quantum theory predicts and that extrapolate the measured $\cos^2$-behavior (or $\sin^2$-behavior, respectively; cf. (Corr)) for non-parallel and non-perpendicular settings. Rather, the experiments show a certain deviation from perfect (anti-)correlations, such that perfect (anti-)correlations cannot be said to be empirically confirmed beyond doubt. Though it might seem reasonable to assume that they nevertheless do hold (because the experimental deviations from perfectness might be attributed to measurement errors and non-ideal detectors), it has become usual in the discussion about Bell's theorem to avoid the strong assumption of perfectness: either one does not make any reference to the correlations at parallel (or perpendicular) settings, or one assumes only *nearly* perfect correlations (nPCorr) (e.g. for parallel settings) and *nearly* perfect anti-correlations (nPACorr) (e.g. for perpendicular settings). Here we shall take the latter route and make the widely accepted assumption of nearly perfect (anti-)correlations. Relaxing the perfect (anti-)correlations, a direct inconsistency similar to the one stated by lemma 1 does not follow any more (autonomy, nearly perfect correlations, nearly perfect anti-correlations and (H^{α}_{16}) are not an inconsistent set). Instead, in this case one can prove the following claim:

> **Lemma 2**: Given autonomy, nearly perfect correlations and nearly perfect anti-correlations, (H^{α}_{16}) implies Bell inequalities: (A)$\wedge$(nPCorr)$\wedge$(nPACorr)$\wedge$ (H^{α}_{16}) $\rightarrow$ (BI)

(For the proof of the lemma see the appendix.)

While this claim does not establish a straightforward inconsistency as the one given strictly perfect (anti-)correlations (cf. lemma 1), it is clear that, via the Bell argument, lemma 2 can be extended to argue for the inconsistency of (H^{α}_{16}) with autonomy, nearly perfect (anti-)correlations and the empirically confirmed EPR/B correlations. In this way, lemma 2 allows for a literal strengthening of Bell's theorem: it allows to modify premise (P4) of the Bell argument to say that both local as well as purely outcome dependent theories imply Bell inequalities. As local theories, purely outcome dependent theories do not produce correlations that are strong enough to violate Bell inequalities. Accordingly, the conclusion of the argument changes to preclude *more* theories than has been believed so far. Besides the local theories it also eliminates those non-local theories which assume an outcome to be dependent (functionally or probabilistically) not only on the local variables but also on the other, distant outcome.

There is a discrepancy between the original Bell argument and lemma 2, which hints to another aspect in which the latter helps to strengthen the former: while the original argument assumes *strictly* perfect correlations, lemma 2 only presumes *nearly* perfect correlations and anti-correlations. In fact, the proof of lemma 2 which shows how to derive Bell inequalities from outcome dependent theories and nearly perfect (anti-)correlations (and autonomy), can easily be adjusted to derive Bell inequalities from local theories and nearly perfect (anti-)correlations (and autonomy). Then, it is clear that one can relax premise (P3) to say that EPR/B correlations involve nearly perfect correlations as well as nearly perfect anti-correlations (instead of strictly perfect correlations). So the proof of lemma 2 also demonstrates the remarkable fact that *one can derive a Wigner-Bell inequality without strictly perfect correlations* (which were so far regarded to be a necessary assumption for deriving that type of Bell inequality).

The Bell inequality that follows by this new kind of prove is a generalized Wigner-Bell inequality,

$$P(\alpha_-\beta_+|a_1b_3) - 2\epsilon - \epsilon^2 \leq \frac{P(\alpha_-\beta_+|a_1b_2) + P(\alpha_-\beta_+|a_2b_3)}{(1-\epsilon^2)}, \tag{2}$$

that differs from a usual Wigner-Bell inequality

$$P(\alpha_-\beta_+|a_1b_3) \leq P(\alpha_-\beta_+|a_1b_2) + P(\alpha_-\beta_+|a_2b_3) \tag{3}$$

by certain correction terms involving a parameter $0 < \epsilon \ll 1$, which is a measure for the deviation from perfect correlations and perfect anti-correlations. (Precisely, ϵ^3 is the maximal fraction of photons deviating from perfect correlations or anti-correlations; see the proof of lemma 2.) It is easy to see that in the border case $\epsilon \to 0$ the generalized Wigner-Bell inequality agrees with the usual one. One can further show (see the proof of lemma 2) that the generalized inequality is violated by the usual statistics of EPR/B experiments, if at least 99.989% of the runs with parallel settings as well as those with perpendicular settings turn out to be perfectly correlated and perfectly anti-correlated, respectively. This defines the above condition of nearly perfect (anti-)correlations more precisely: only in worlds where the fraction of perfectly (anti-)correlated runs exceeds the indicated threshold, purely outcome dependent theories are ruled out.

This quantitative limit reveals a final resort for the defender of pure outcome dependence: she might hint to the fact that in actual experiments far less than 99.989% of the entangled objects show perfect (anti-)correlations. This indeed shows that the question whether purely outcome dependent theories can hold or not is not yet decided empirically beyond doubt. Let me stress, however, that the main aim in this paper is not to decide this empirical and quantitative question, but the conceptual and qualitative one, namely whether it is possible to amend Bell's argument for a stronger conclusion, ruling out even certain non-local theories.

That said, I can add that I think that there are good reasons not to take the mentioned empirical discrepancy to undermine the argument against purely outcome dependent theories. First, the derivation of the inequality (2) uses certain rather rough estimations, which contribute to the fact that the degree of perfectness that is required for a violation to take place is high. Improved future derivations, which include more precise (and expectedly more complicated) estimations, might lower that degree considerably. Second, the past has shown that experimental physicists have continuously been increasing the fraction of measured perfectly (anti-)correlated pairs of entangled objects, by using more and more sophisticated experimental techniques. So it is to be expected that the empirically confirmed degree of perfect correlation will increase in the future as well. Finally, quantum mechanics predicts perfect correlations and at present there is no further, independent evidence (besides the fact that experiments do not yield strictly perfect (anti-)correlations) to doubt that quantum mechanics is wrong; for this reason, it seems reasonable to assume that the deviation from perfectness in experiments is due to experimental imperfections.

Whether these arguments against the empirical discrepancy are conclusive or not: if my mathematical proofs are correct, the clear result of this section is that, given autonomy, purely outcome dependent theories cannot be adequate theories of the quantum realm if either *strictly* perfect (anti-)correlations or *nearly* perfect (anti-)correlations with a fraction of (dis-)agreement larger than 99.989% hold.

4 Generalization: A comprehensive scheme of possible theories

Strengthening an argument it is desirable to make it as strong as possible. We shall now generalize the stronger Bell argument that we have just presented so as to rule out *all* theories that can be ruled out by this type of argument. In order to capture all theories we shall proceed systematically and list a scheme of all logically possible theories, for each of which we check whether it is consistent under the given assumptions, and, if it is, whether it implies Bell inequalities or not. Note that this list will also contain theories that do not seem physically plausible. It is important, however, to include these theories into our investigation because in the end we aim to show that we have provided the strongest possible argument on a qualitative level (see section 7).

As we have said in the last section, local factorization and (H_{16}^{α}) are particular product forms of the hidden joint probability. In general, according to the product rule of probability theory, any hidden joint probability can equivalently be written as a product,

$$\begin{aligned} P(\alpha\beta|ab\lambda) &= P(\alpha|\beta ba\lambda)P(\beta|ab\lambda) && (4)\\ &= P(\beta|\alpha ab\lambda)P(\alpha|ba\lambda). && (5)\end{aligned}$$

Since there are two such general product forms, one whose first factor is a conditional probability of $\boldsymbol{\alpha}$ and one whose first factor is a conditional probability of $\boldsymbol{\beta}$, for the time being, let us restrict our considerations to the product form (4), until in the next section we shall transfer the results to the other form (5).

We stress that the product form (4) of the hidden joint probability holds in general, i.e. for all probability distributions. According to probability distributions with appropriate independences, however, the factors on the right-hand side of the equation reduce in that certain variables in the conditionals can be left out. If, for instance, outcome independence holds, $\boldsymbol{\beta}$ can disappear from the first factor, and the joint probability is said to 'factorise'. *Local* factorisation further requires that the distant settings in both factors disappear, i.e. that so called parameter independence holds. Prima facie, any combination of variables in the two conditionals in (4) seems to constitute a distinct product form of the hidden joint probability. Restricting ourselves to *irreducibly hidden* joint probabilities, i.e. requiring $\boldsymbol{\lambda}$ to appear in both factors, there are $2^5 = 32$ *combinatorially* possible forms (for any of the three variables in the first conditional and any of the two variables in the second conditional *besides* $\boldsymbol{\lambda}$ can or cannot appear). Table 1 shows these conceivable forms which I label by (H_{1}^{α}) to (H_{32}^{α}) (the superscript α is due to the fact that we have used (4) instead of (5)).

Table 1: Classes of probability distributions

	I		II	III	IV			V	VI		VII	VIII	IX
	(H_i^α):	$P(\alpha\beta\|ab\lambda) = \dots$									PCorr	nPCorr	
	i	$P(\alpha\|$	β	b	a	$\lambda)$	$\cdot P(\beta\|$	a	b	$\lambda)$	$\square$(BI)	$\square$(BI)	Notes
strong non-locality$^\alpha$	1		1	1	1			1	1		0	0	
	2		1	1	1			1	0		0	0	
	3		1	1	1			0	1		0	0	$\mathrm{QM_p}$
	4		1	1	0			1	1		—	0	
	5		1	0	1			1	1		—	0	
	6		0	1	1			1	1		0	0	$\mathrm{Bohm_s}$
	7		1	1	1			0	0		0	0	$\mathrm{QM_m}$
	8		0	1	1			1	0		0	0	
	9		0	1	1			0	1		0	0	$\mathrm{Bohm}_{\beta<a}$
	10		1	0	0			1	1		—	0	
	11		0	1	0			1	1		0	0	
	12		0	0	1			1	1		0	0	$\mathrm{Bohm}_{\alpha<b}$
	13		0	1	1			0	0		0	0	
	14		0	0	0			1	1		0	0	
weak non-locality$^\alpha$	15		1	1	0			1	0		—	1	
	16		1	0	1			0	1		—	1	pure outc. dep.
	17		1	0	1			1	0		—	—	
	18		1	1	0			0	1		—	—	
	19		1	1	0			0	0		—	—	
	20		1	0	1			0	0		—	—	
	21		1	0	0			1	0		—	—	
	22		0	1	0			1	0		1	1	
	23		0	0	1			1	0		—	—	
	24		1	0	0			0	1		—	—	
	25		0	1	0			0	1		—	—	
	26		1	0	0			0	0		—	—	
	27		0	1	0			0	0		—	—	
	28		0	0	0			1	0		—	—	
locality$^\alpha$	29		0	0	1			0	1		1	1	local factoriz.
	30		0	0	1			0	0		—	—	
	31		0	0	0			0	1		—	—	
	32		0	0	0			0	0		—	—	

The specific product form of the hidden joint probability is *an essential feature* of the probability distributions of EPR/B experiments. For, as we shall see, it not only determines whether a probability distribution can violate Bell inequalities but also carries unambiguous information about which variables of the experiment are probabilistically independent of another. Therefore, it is natural to use the product form of the hidden joint probability in order to classify the probability distributions. We can say that each product form of the hidden joint probability constitutes a *class of probability distributions* in the sense that probability distributions with the same form (but different numerical weights of the factors) belong to the same class. In order to make the assignment of probability distributions to classes unambiguous let us require that each probability distribution belongs only to that class which corresponds to its *simplest* product form, i.e. to the form with the minimal number of variables appearing in the conditionals (according to the distribution in question).

This scheme of classes is comprehensive: Any probability distribution of the EPR/B experiment must belong to one of these 32 classes. In this systematic overview, the class constituted by local factorisation is (H_{29}^{α}), and it now also becomes clear why the product form of purely outcome dependent theories has been tagged (H_{16}^{α}) in the last section. Furthermore, if we allow that there might be no hidden states $\boldsymbol{\lambda}$, the quantum mechanical distribution as well as the empirical distribution (which as far as we know coincide, but see our discussion of perfect (anti-)correlations in the last section) belong to class (H_{7}^{α}) (if the photon state $\boldsymbol{\psi}$ is *maximally* entangled, noted by 'QM_{m}') or to (H_{3}^{α}), respectively (if $\boldsymbol{\psi}$ is *partially* entangled, noted by 'QM_{p}').[7] The de-Broglie-Bohm theory falls under class (H_{6}^{α}), when both settings are chosen before the detector at the respective other side has registered, i.e. $t(a) < t(\beta)$ and $t(b) < t(\alpha)$;[8] we label the corresponding probability distribution by '$\mathrm{Bohm}_{\mathrm{s}}$' (the index standing for *s*ymmetrical time ordering). Otherwise, when the $\boldsymbol{\beta}$-measurement completes before $\boldsymbol{a}$ has been set to its final state (labelled by '$\mathrm{Bohm}_{\beta<a}$'), the theory falls in class (H_{9}^{α}); and when the $\boldsymbol{\alpha}$-measurement is over before $\boldsymbol{b}$ has been chosen (labelled by '$\mathrm{Bohm}_{\alpha<b}$'), we have class (H_{12}^{α}). Similarly, any other theory of the quantum realm has its unique place in one of the classes.

One crucial advantage of such an abstract classification is that it simplifies matters insofar we can now derive features of *classes* of probability distributions and can be sure that these features hold for all members of a class, i.e. for all theories whose probability distributions fall under the class in question. The feature that we are most interested in is, of course, which of these classes (given autonomy) are consistent with the empirical probability distribution of EPR/B experiments. As in the previous section

[7] The typical case for EPR/B experiments is to prepare a maximally entangled quantum state (e.g. $|\psi\rangle = \sqrt{p}|+\rangle|+\rangle + \sqrt{1-p}|-\rangle|-\rangle$ with $p = \frac{1}{2}$), because one wants to have a maximal violation of the Bell inequalities. The slightest deviation from maximal entanglement ($p \neq \frac{1}{2}$), however, breaks the symmetry of the state. The probability distribution of such *partially* entangled states shows an additional probabilistic dependence on the local setting in the second factor; hence, they fall in class (H_{3}^{α}). For an overview of the dependences and independences in the quantum mechanical probability distribution of maximally and partially entangled states see Näger (2015, table 1).

[8] Such temporal ordering between space-like separated events is, of course, only possible when there is a preferred frame of reference, which Bohm's theory presupposes.

we discern two cases according to whether *strictly* perfect (anti-)correlations or *nearly* perfect (anti-)correlations hold.

4.1 Strictly perfect (anti-)correlations

For the case of strictly perfect (anti-)correlations, the following theorems hold:

> **Theorem 1.1**: Autonomy, perfect correlations, perfect anti-correlations and a class of probability distributions (H_i^α) form an inconsistent set if and only if (i) the product form of (H_i^α) involves at most one of the settings or (ii) the product form of (H_i^α) involves both settings but its first factor involves the distant outcome and at most one setting.

> **Corollary 1.1**: A class (H_i^α) is consistent with autonomy, perfect correlations and perfect anti-correlations if and only if (¬i) the product form of (H_i^α) involves both settings and (¬ii) in case the distant outcome appears in the first factor of (H_i^α)'s product form, also both settings appear in that factor.

> **Theorem 1.2**: Given autonomy, perfect correlations and perfect anti-correlations a consistent class (i.e. a class that fulfills (¬i) and (¬ii)) implies Bell inequalities if and only if (iii) each factor of its product form involves at most one setting.

> **Corollary 1.2**: Given autonomy, perfect correlations and perfect anti-correlations a consistent class (i.e. a class that fulfills (¬i) and (¬ii)) does not imply Bell inequalities if and only if (¬iii) at least one factor of its product form involves both settings.

(The proofs of these theorems can be found in the appendix.)

The consequences of these claims for the status of the different classes are represented in column VII of table 1. The heading of the column, '$\Box$(BI)', means *necessarily, Bell inequalities hold*. So the column indicates whether a certain product form implies Bell inequalities ('1') or does not imply them ('0') (according to theorem 1.2 or corollary 1.2, respectively); it further indicates when this question does not make sense ('—') because a product form is inconsistent with the background assumptions autonomy and perfect (anti-)correlations (according to theorem 1.1). It is understood that classes that are marked by either '0' or '1' are consistent with the background assumptions (cf. corollary 1.1). Clearly, all classes that are marked either by '—' or '1' are impossible if autonomy and perfect (anti-)correlations hold: the former yield a direct contradiction with the background assumptions, while the latter contradict the empirical probability distribution via the Bell argument.

The inconsistent classes ('—') divide into two subgroups, corresponding to which condition for inconsistency, (i) or (ii) (cf. theorem 1.1), is fulfilled:

> Inconsistency due to condition (i): $\{(\mathrm{H}_{17}^\alpha), ..., (\mathrm{H}_{32}^\alpha)\}\backslash\{(\mathrm{H}_{22}^\alpha), (\mathrm{H}_{29}^\alpha)\}$

Inconsistency due to condition (ii): $\{(\mathrm{H}_4^\alpha), (\mathrm{H}_5^\alpha), (\mathrm{H}_{10}^\alpha), (\mathrm{H}_{15}^\alpha), (\mathrm{H}_{16}^\alpha)\}$

Positively, theorem 1.1 further says that all classes that are not inconsistent according to these criteria are consistent with the background assumptions: the criterion of consistency, (¬i) *and* (¬ii), as stated in corollary 1.1, is just the negation of the condition for inconsistency, (i) *or* (ii), in theorem 1.1.

We emphasize that the consistency and inconsistency claims of classes with the background assumptions have asymmetric consequences on the level of single probability distributions. On the one hand, a class being inconsistent with the background assumptions means that *every* probability distribution of that class forms an inconsistent set with the assumptions. It is the general product form defining the class which is in conflict with the assumptions, hence all members of the class are. The same, mutatis mutandis, however, is not true of the consistent classes. A class being consistent does *not* mean that *every* probability distribution of that class is consistent with the assumptions. Rather, by the laws of logic, it just means that *at least one* probability distribution of a class is consistent with the assumptions, showing that the general product form of that class is *not per se* in conflict with them. This is what consistency of a class means (when we define inconsistency in the natural way as just stated). This definition of consistency is perfectly compatible with the fact that there are distributions in a consistent class that are inconsistent with the assumptions due to their specific *numerical values*. For instance, one can easily imagine distributions falling under class (H_7^α) that, at parallel settings, involve correlations that are weaker than perfectness. These distributions are obviously not consistent with the background assumptions, although their general product form is. Hence, we have to keep in mind that being consistent with the background assumptions on the level of classes, which is the level the present analysis proceeds on, is just a necessary condition for the distributions in that class to be consistent.

Turning to theorem 1.2, all classes marked by '1', i.e. (H_{22}^α) and (H_{29}^α), can explicitly be shown to imply a Bell inequality. That (H_{29}^α), local factorization, implies the inequalities is well known, but that

$$P(\alpha\beta|ab\lambda) = P(\alpha|b\lambda)P(\beta|a\lambda), \tag{H_{22}^α}$$

a non-local class, does, has not been observed so far. That class is the symmetrical counterpart to local factorization, compared to which the settings are swapped, such that each outcome depends on its distant setting. For this reason the derivation of the Bell inequalities runs very similarly as for local factorization (just swap the settings in the original proof).

On the other hand, the theorem also says that any consistent class that violates (iii), can be shown *not* to imply the Bell inequalities. Here we have a similar asymmetry between the level of classes and that of distributions as in the case of (in-)consistency. Since a class implying Bell inequalities (given the background assumptions) means that *every* probability distribution having the product form in question obeys the inequalities, the claim that a class does *not* imply the inequalities (given the background assumptions) denotes the fact that *there is at least one probability distribution in that class that violates* *the inequalities.* Therefore, not implying Bell inequalities emphatically does *not* mean that *every* probability distribution in a class violates the inequalities. For this reason, given just the product form of one of the classes violating (iii) one cannot decide whether Bell inequalities hold; whether they do in these cases depends on the *numerical* features of the probability distribution in question. In this sense, one might reasonably say that probability distributions of these classes *can* violate Bell inequalities. So far for the meaning of implying and not implying the Bell inequalities. Let us now turn to the criterion which demarcates the two cases.

Condition (iii), that, in order to imply Bell inequalities, a consistent class may not involve more than one setting in each factor of its product form, is the essential characteristic (in terms of the product form) to tell apart classes marked by '1' from those marked by '0'. This criterion differs considerably from the usual message that local theories imply Bell inequalities (and non-local ones do not). In order to understand its content, let us partition the classes into three groups, depending on which variables appear in their constituting product forms:

Local$^\alpha$ classes: (H^α_{29})–(H^α_{32})
Each factor only contains time-like (or light-like) separated variables.

Weakly non-local$^\alpha$ classes: (H^α_{15})–(H^α_{28})
At least one of the factors involves space-like separated variables, but none of the factors involves both settings.

Strongly non-local$^\alpha$ classes: (H^α_{1})–(H^α_{14})
At least one of the factors involves both settings. (¬iii)

With these new concepts we can summarize theorems 1.1 and 1.2 as saying that *given autonomy, perfect correlations and perfect anti-correlations, every local*$^\alpha$ *and weakly non-local*$^\alpha$ *class either is inconsistent with autonomy and the perfect (anti-)correlations or (if it is consistent) obeys Bell inequalities. Certain strongly non-local*$^\alpha$ *classes are inconsistent with autonomy and the perfect (anti-)correlations as well; however, all consistent strongly non-local*$^\alpha$ *classes do not imply, i.e. can violate, the Bell inequalities.* (Strongly non-local$^\alpha$ classes are by definition just those classes that fulfill criterion (¬iii) not to imply the Bell inequalities.)

What does this result mean? On the one hand, it sounds familiar that local classes are impossible in the given situation. Local classes involve only time-like (or light-like) separated variables in the factors of their hidden joint probability, and local factorization, which is well-known to imply Bell inequalities, is the paradigm of product forms constituting these classes. Theorem 1.1 just adds the further claim that given autonomy and perfect (anti-)correlations all other local classes are directly inconsistent.

The surprising consequence of theorems 1.1 and 1.2 rather is that *even certain non-local classes are ruled out.* Every class in the group of weakly non-local$^\alpha$ classes is forbidden. Most of the classes in that group are directly inconsistent with the assumptions of autonomy and perfect (anti-)correlations, including (as we have shown in the previous section) the purely outcome dependent class (H^{α}_{16}). That purely outcome dependent theories are not even available when perfect (anti-)correlations (and autonomy) hold, is, as we have already remarked, a central result of this investigation, because it belongs to the group of classes that has evolved as the received view of what quantum non-locality amounts to on a probabilistic level. In fact, there is only one weakly non-local$^{\alpha}$ class, which is consistent with the background assumptions autonomy and perfect (anti-)correlations, viz. (H^{α}_{22}). Whether this class is physically plausible or not: the fact that it implies the inequalities proves the important insight that *Bell inequalities are not a locality condition* (because there is a class obeying Bell inequalities that is non-local).

Instead of locality, the hallmark of theories implying Bell inequalities rather is, as theorem 1.2 states, that they may not involve more than one setting in each factor of their product form. The negation of this condition, that at least one factor contains both settings, is exactly the defining feature of strongly non-local$^{\alpha}$ classes. That is why the above partition is so natural. However, this does not mean that all strongly non-local$^{\alpha}$ classes are allowed in the given situation; for some of them—(H^{α}_{4}), (H^{α}_{5}) and (H^{α}_{10})—are inconsistent with autonomy and perfect (anti-)correlations ('—'). This demonstrates that the criteria for being consistent with these background assumptions, (¬i) *and* (¬ii), and for not implying Bell inequalities, (¬iii), which is equivalent to being strongly non-local$^{\alpha}$, are not disjunct. A theory can be strongly non-local$^{\alpha}$ and still violate (¬ii) (all strongly non-local$^{\alpha}$ classes marked by '—'), e.g. (H^{α}_{4}); and there are theories fulfilling (¬i) and (¬ii) but fail to be strongly non-local$^{\alpha}$, e.g. (H^{α}_{22}). A successful theory must belong to one of the classes that takes both hurdles, and those are the ones marked by '0' in column VII of table 1.

4.2 Nearly perfect (anti-)correlations

Let us now relax the assumption of strictly perfect (anti-)correlations to nearly perfect (anti-)correlations and observe how that changes the situation. In this case, the following theorems can be proven:

> **Theorem 2.1**: Autonomy, nearly perfect correlations, nearly perfect anti-correlations, and a class of probability distributions (H^{α}_{i}) form an inconsistent set if and only if (i) the product form of (H^{α}_{i}) involves at most one of the settings.
>
> **Corollary 2.1**: A class (H^{α}_{i}) is consistent with autonomy, nearly perfect correlations and nearly perfect anti-correlations if and only if (¬i) the product form of (H^{α}_{i}) involves both settings.
>
> **Theorem 2.2**: Given autonomy, nearly perfect correlations and nearly perfect anti-correlations each consistent class (i.e. each class that fulfills (¬i)) implies Bell inequalities if and only if (iii) each factor of its product form involves at most one setting.
>
> **Corollary 2.2**: Given autonomy, nearly perfect correlations and nearly perfect anti-correlations each consistent class (i.e. each class that fulfills

(¬i)) does not imply Bell inequalities if and only if (¬iii) at least one factor of its product form involves both settings.

(The proofs of the theorems can be found in the appendix.)

The consequences of these claims are represented in column VIII of table 1. Since nearly perfect (anti-)correlations are a considerably weaker requirement than that of strictly perfect ones, one essential change that occurs in these theorems compared to the former is that the conditions for consistency with the background assumptions (autonomy and nearly perfect (anti-)correlations in the new case) are considerably weaker as well: theorem 2.1 just requires condition (¬i) but not condition (¬ii). As a consequence, all classes that have been ruled out by condition (¬ii) (in theorem 1.1), viz. (H^α_4), (H^α_5), (H^α_{10}), (H^α_{15}) and (H^α_{16}), are now consistent with the new, less strict background assumptions.

Especially purely outcome dependent theories defined by

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta a\lambda)P(\beta|b\lambda) \tag{H^α_{16}}$$

cease to be directly inconsistent with the background assumptions. However, since the criterion for implying Bell inequalities stays essentially unchanged (requirement (iii) still holds),[9] outcome dependent theories imply Bell inequalities (see our discussion in section 3), so they are still forbidden. It is just that the reason why they are forbidden changes. Similar facts are true for the symmetrical counterpart to purely outcome dependent theories,

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta b\lambda)P(\beta|a\lambda), \tag{H^α_{15}}$$

which differs from (H^α_{16}) in that the settings are swapped between the factors, such that each outcome depends on the distant (instead of on the local) setting. In effect, also in the new situation it is still true that *all local$^\alpha$ and weakly non-local$^\alpha$ classes are forbidden*.

Concerning the strongly non-local$^\alpha$ classes, however, the situation changes. Formerly, certain strongly non-local$^\alpha$ classes, (H^α_4), (H^α_5) and (H^α_{10}), were forbidden because they were inconsistent with the background assumptions. Relaxing the background assumptions, we have already said that they become consistent. But unlike the weakly non-local$^\alpha$ classes that have become consistent, (H^α_{15}) and (H^α_{16}), these strongly non-local$^\alpha$ classes do not imply Bell inequalities, because they clearly do not fulfill condition (iii); by weakening the background assumptions, these classes cease to be ruled out by the theorems. As a consequence, all strongly non-local$^\alpha$ classes are now consistent with the background assumptions and do not imply Bell inequalities. The reason for this new situation is that abandoning criterion (¬ii) for consistency, the remaining criterion, (¬i), is entailed by the criterion for not implying Bell inequalities, viz. to be strongly non-local$^\alpha$ (so the two criteria are not logical independent any more).

[9] There is just the slight difference that now both nearly perfect correlations *and* nearly perfect anti-correlations are required, whereas according to theorem 1.2 the anti-correlations were not needed for the derivation.

In sum, the result is that *given autonomy, nearly perfect correlations and nearly perfect anti-correlations, every local$^{\alpha}$ and weakly non-local$^{\alpha}$ class either is inconsistent with autonomy and the perfect (anti-)correlations or (if it is consistent) obeys Bell inequalities. In contrast, all strongly non-local$^{\alpha}$ classes are consistent with autonomy and the perfect (anti-)correlations and do not imply Bell inequalities.* Unlike the case with strictly perfect correlations, there are no forbidden strongly non-local$^{\alpha}$ theories, which amounts to a slight modification of the set of precluded classes.

The main messages, however, have not changed: as opposed to what the standard discussion suggests, it is *not* true that local factorisation (and the other local product forms) are the *only* product forms which are forbidden by the empirical statistics of EPR/B experiments (if autonomy holds). Rather, we have found that 18 (21 in the case of strictly perfect (anti-)correlations) of the 32 logically possible classes are forbidden, among them 14 (17 in the case of strictly perfect (anti-)correlations) *non-local* classes. Some of these non-local classes are forbidden because they are directly inconsistent with the assumptions autonomy and (nearly) perfect (anti-)correlations. Others are forbidden because they imply Bell inequalities. This latter fact has two important consequences. First, it makes explicit that Bell inequalities are not a locality condition. Neither, second, is locality a necessary condition for deriving Bell inequalities. The criterion to imply the inequalities (if autonomy and (nearly) perfect anti-correlations hold) rather is a different one, which has not to do with the locality/non-locality divide: Bell inequalities are implied by each probability distribution whose product form involves at most one setting in each of its factors. So according to a probability distribution the outcomes might depend on their distant setting as well as on each other, (H^{α}_{15}), and still Bell inequalities follow. As a consequence, if one searches for theories which conform to the empirical fact that (nearly) perfect correlations hold and Bell inequalities are violated they can only be among the strongly non-local$^{\alpha}$ ones (which are defined to involve both settings in at least one factor). Contrary to the view suggested by Bell's original theorem it cannot be a weakly non-local$^{\alpha}$ class.

5 Strengthening Bell's theorem

It is clear that each set of theorems (1.1 and 1.2 as well as 2.1 and 2.2) can be used to strengthen Bell's theorem. On the other hand, it is not clear which of these available new arguments should be considered to be the strongest. (The first set results in an argument that, compared to the argument resulting from the second set, requires the stronger assumption of strictly perfect correlations (weakening the argument), but allows for a stronger conclusion, because it rules out even some of the strongly non-local$^{\alpha}$ classes). Here we restrict our discussion to the argument resulting from the second set, because it avoids the controversial assumption of strictly perfect (anti-)correlations. (The argument from the first set can be formulated mutatis mutandis.)

(P1) There are EPR/B correlations: (Corr)

(P2) EPR/B correlations violate Bell inequalities: (Corr) $\rightarrow \neg$(BI)

(P3′) EPR/B correlations include nearly perfect correlations and nearly perfect anti-correlations: (Corr) $\rightarrow$ (nPCorr) $\wedge$ (nPACorr)

(P6) Those local$^{\alpha}$ and weakly non-local$^{\alpha}$ classes that involve at most one setting in their product form are inconsistent with autonomy, nearly perfect correlations and nearly perfect anti-correlations:

$$(\mathrm{A}) \wedge (\mathrm{nPCorr}) \wedge (\mathrm{nPACorr}) \rightarrow \bigwedge_{\substack{i=17..32\\ \setminus\{22,29\}}} \neg(\mathrm{H}_i^{\alpha})$$

(P4′) Bell inequalities can be derived from autonomy, nearly perfect correlations, nearly perfect anti-correlations and any local$^{\alpha}$ or weakly non-local$^{\alpha}$ class of probability distributions that involves both settings in its product form:

$$\Big[(\mathrm{A}) \wedge (\mathrm{nPCorr}) \wedge (\mathrm{nPACorr}) \wedge \Big(\bigvee_{\substack{i=15,16,\\ 22,29}} (\mathrm{H}_i^{\alpha})\Big)\Big] \rightarrow (\mathrm{BI})$$

(P5) Autonomy holds: (A)

(C1′) Both local$^{\alpha}$ and weakly non-local$^{\alpha}$ classes fail:

$$\Big(\bigwedge_{i=15}^{32} \neg(\mathrm{H}_i^{\alpha})\Big)$$

Compared to the original Bell argument (section 2) there are three substantial changes, which strengthen the argument. A first change concerns the fact that everywhere in the argument we have relaxed controversial strictly perfect correlations to uncontroversial nearly perfect correlations (in premisses (P3) and (P4) of the original argument). This is a strengthening in the sense that the argument makes weaker assumptions. At the same places in the argument where nearly perfect correlations occur we have additionally introduced nearly perfect (anti-)correlations. This might seem as a weakening of the argument; in fact, however, it is a neutral move, because it is uncontroversial that the nearly perfect anti-correlations follow from the EPR/B correlations (as the nearly perfect correlations do; see premise (P3′)), and these EPR/B correlations have already been assumed in the original argument (premise (P1)).

A second strengthening of the argument stems from introducing a completely new premise (P6), which states the content of theorem 2.1, that certain classes are not compatible with autonomy, nearly perfect correlations and perfect anti-correlations. Given that autonomy and perfect (anti-)correlations are assumed anyway (or derive from assumptions), it is clear that these classes will be ruled out by the overall argument. In this sense, (P6) provides a genuine strengthening of the conclusion of the theorem. Deriving a direct contradiction between the background assumptions and certain classes without involving a Bell inequality, premise (P6) has no counterpart in the original argument and rather has the status of an amendment—however, an amendment that naturally fits in. Note that assuming the additional premise (P6) does not weaken the argument because it can be proven mathematically (see the proof of theorem 2.1).

A third modification, indeed the central strengthening, consists in the adaption of premise (P4) to theorem 2.2, which says that one can derive Bell inequalities not only from local factorization but from all those local$^\alpha$ and weakly non-local$^\alpha$ classes that are consistent given autonomy and perfect (anti-)correlations. Accordingly, we have replaced local factorisation in the antecedent by the disjunction of these product forms. This makes the antecedent of (P4′) weaker than that in (P4) and, hence, the argument stronger. Since the overall Bell argument is a modus tollens argument to the negation of that premise, this modification also strengthens the conclusion of the theorem.

Making these changes has a considerable effect on the overall Bell argument. Instead of the standard conclusion (C1), that the violation implies the failure of local factorisation, by the modified argument we arrive at the essentially stronger conclusion (C1′). While the original result, the failure of local factorisation, implied that all local$^\alpha$ classes fail (because the other local classes are specializations of local factorisation), the new result additionally excludes all weakly non-local$^\alpha$ classes.

6 Further strengthening by a complementary partition

Our considerations leading to this new result of the Bell argument rest on the fact that we have found alternatives to local factorisation from writing the hidden joint probability according to the product rule (4) and conceiving different possible product forms (table 1). However, we can as well write the hidden joint probability according to the second product rule (5), and similar arguments as above lead us to a similar table as table 1, whose classes, (H_1^β)–(H_{32}^β), differ to those in table 1 in that the outcomes and the settings are swapped. For instance, class (H_{16}^β) is defined by the product form $P(\alpha\beta|ab\lambda) = P(\beta|\alpha b\lambda)P(\alpha|a\lambda)$ in contrast to (H_{16}^α), which is constituted by $P(\alpha\beta|ab\lambda) = P(\alpha|\beta a\lambda)P(\beta|b\lambda)$. Note that this new classification is a *different* partition of the possible probability distributions, which reasonably might be called *complementary partition*. Any probability distribution must fall in exactly one of the classes (H_1^α)–(H_{32}^α) *and* in exactly one of the classes (H_1^β)–(H_{32}^β). Analogously to theorem 1 one can prove for the new partition that (given autonomy and nearly perfect (anti-) correlations) also each local$^\beta$ and weakly non-local$^\beta$ class either is inconsistent or implies Bell inequalities, so that we can reformulate (P6) and (P4′) as:

(P6′) Those local$^\alpha$, weakly non-local$^\alpha$, local$^\beta$ and weakly non-local$^\beta$ classes that involve at most one setting in their product form are inconsistent with autonomy, nearly perfect correlations and nearly perfect anti-correlations:

$$(\mathrm{A}) \wedge (\mathrm{nPCorr}) \wedge (\mathrm{nPACorr}) \rightarrow \bigwedge_{\substack{i=17..32\\ \setminus\{22,29\}}} \neg(\mathrm{H}_i^\alpha) \quad \wedge \quad \bigwedge_{\substack{i=17..32\\ \setminus\{22,29\}}} \neg(\mathrm{H}_i^\beta)$$

(P4″) Bell inequalities can be derived from autonomy, nearly perfect correlations, nearly perfect anti-correlations and any local$^\alpha$, weakly non-local$^\alpha$, local$^\beta$ or weakly non-local$^\beta$ class of probability distributions that involves both settings in its product form:

$$\Big[(\mathrm{A}) \wedge (\mathrm{nPCorr}) \wedge (\mathrm{nPACorr}) \wedge \Big(\bigvee_{\substack{i=15,16,\\22,29}} (\mathrm{H}_i^\alpha) \quad \vee \bigvee_{\substack{i=15,16,\\22,29}} (\mathrm{H}_i^\beta)\Big)\Big] \rightarrow (\mathrm{BI})$$

With these new premises we can formulate an even stronger Bell argument from (P1), (P2), (P3′), (P6′), (P4″) and (P5) to

(C1″) All local$^\alpha$, weakly non-local$^\alpha$, local$^\beta$ and weakly non-local$^\beta$ classes fail:

$$\left(\bigwedge_{i=15}^{32} \neg(\mathrm{H}_i^\alpha) \wedge \bigwedge_{i=15}^{32} \neg(\mathrm{H}_i^\beta)\right)$$

This is the conclusion of the new stronger Bell argument. It takes the usual result from *any* kind of non-locality (the mere failure of local factorisation) to a more *specific* one (namely exclusive the weakly non-local$^\alpha$ and weakly non-local$^\beta$ classes). Stating which classes are excluded, the result formulated here is a negative one. But it is easy to turn it into a positive formulation: since our scheme of logically possible classes is comprehensive, the failure of all local$^\alpha$ and weakly non-local$^\alpha$ classes is equivalent to the fact that one of the strongly non-local$^\alpha$ classes, (H_1^α)–(H_{14}^α), holds. Analogously, if a probability distribution is neither local$^\beta$ nor weakly non-local$^\beta$ it must be strongly non-local$^\beta$, i.e. belong to one of the classes (H_1^β)–(H_{14}^β). Therefore, equivalently to (C1″) we can say:

(C1‴) One of the strongly non-local$^\alpha$ classes and one of the strongly non-local$^\beta$ classes has to hold.

$$\left(\bigvee_{i=1}^{14} (\mathrm{H}_i^\alpha) \wedge \bigvee_{i=1}^{14} (\mathrm{H}_i^\beta)\right)$$

This is the positive conclusion of the stronger Bell argument in terms of classes.

Finally, we can formulate the same result in terms of which features the hidden joint probability must have. Let us define the following concept:

> **Probabilistic Bell contextuality (PBC)** holds if and only if according to both product forms of the hidden joint probability $P(\alpha\beta|ab\lambda)$ at least one of the outcomes depends probabilistically on both settings.

Then, equivalently to (C1″) or (C1‴), we can say:

(C1⁗) Probabilistic Bell Contextuality holds.

(C1″), (C1‴) and (C1⁗) are equivalent conclusions of the stronger Bell argument.

7 Impossibility of stronger consequences

These conclusions of the new Bell argument are considerably stronger than those of previous versions. Here, we would like to stress the fact that, if our considerations have been correct and the typical background assumptions hold (autonomy and nearly perfect (anti-)correlations), by our systematic approach we can even be sure that these conclusions are the *strongest possible consequences of the violation of Bell inequalities on a qualitative probabilistic level.* What does this mean and how can we argue for this claim?

The claim is meant to say that it is impossible to strengthen Bell's argument in such a way as to rule out more classes of probability distributions than we have ruled out here. Note that this is not to say that certain classes might not be ruled out due to other criteria, maybe due to their incompatibility with relativity or the like. Since the classes are defined by the probabilistic dependences and independences of the respective product form and do not refer to any quantitative aspects of the probability distributions, we call our result a 'qualitative probabilistic' one.

Our considerations in this paper have two important features that preclude future strengthenings of the argument to rule out more classes. First, the central methodological procedure of our argument was to consider *all logically possible* classes of probability distributions. Hence, any probability distribution that conceivably might describe an EPR/B experiment must fall under one of the classes in our systematic overview (cf. table 1). For this reason, we can be sure that we have not overlooked any probability distribution for the EPR/B experiment. There simply are no probability distributions left that might bring in some surprise; we have captured them all.

A second important feature is that our argument provides sufficient *and* necessary conditions for classes to imply Bell inequalities. By stating that local classes imply Bell inequalities, former arguments typically have only provided sufficient criteria. This left open the possibility that there are further classes implying the inequalities—and, indeed, here we have found that many non-local classes, viz. the weakly non-local ones, do as well. On the other hand, by explicitly showing that the remaining classes, the strongly non-local classes, can violate the inequalities (see the proofs of theorems 1.2 and 2.2, where we have constructed explicit examples of distributions in those classes that violate the inequalities), we have precluded that future arguments might show one of the strongly non-local classes to imply the inequalities as well. And if this argument, that proceeds on the qualitative probabilistic level of the classes and their product forms, is correct, and the background assumptions we have presupposed hold, we cannot entail a stronger claim on that level than that local and weakly non-local classes imply Bell inequalities while strongly non-local classes can violate them.

The latter claim also reveals a certain limitation of the argument presented here. It emphatically does not say that strongly non-local classes violate Bell inequalities; it only says that strongly non-local classes *can* violate Bell inequalities, meaning that some of the strongly non-local distributions do violate the inequalities while others do not. In fact, one can explicitly find examples for probability distributions in each of the strongly non-local$^\alpha$ classes (H_1^α)–(H_{14}^α) (as well as in the strongly non-local$^\beta$ classes (H_1^β)–(H_{14}^β))

which *obey* Bell inequalities—and these distributions clearly could be ruled out by more precise arguments. However, belonging to the same class, discerning strongly non-local classes which violate the inequalities from those that obey them clearly can*not* be made on a *qualitative* probabilistic level. Any improvement of the argument must refer to the specific *numerical values* of the probability distribution in question, so there is no general claim that can be made on the basis of the mere product form; the product form of any strongly non-local class alone does not determine whether Bell inequalities hold or fail.

It follows that the consequence of my stronger Bell argument, that the quantum world can only be described correctly by a theory falling under a strongly non-local class, is only a *necessary* condition for violating Bell inequalities; it is not a sufficient one. (Note the difference between conditions for *violating* Bell inequalities and conditions for *not implying* them; we have provided necessary and sufficient conditions for the latter but only necessary ones for the former.) Sufficient criteria to violate Bell inequalities would have to involve conditions for the *strength* of the correlations. A common measure for how strong a correlation is, is mutual information, so information theoretic works which derive numerical values for how much mutual information has to be given in order to violate Bell inequalities, provide an answer to that question (cf. Maudlin 1994, ch. 6 and Pawlowski et al. 2010). These are important works, which can further sharpen our concept of quantum non-locality following from EPR/B experiments. Such quantitative improvements, however, do not count against my claim here that the conclusion of my new stronger Bell argument captures the strongest possible consequences of the violation of Bell inequalities on a qualitative probabilistic level.

8 Discussion

In this paper we have presented a considerably stronger version of Bell's theorem. The new argument rests on the insight that the members of a large range of non-local theories, which we have called *weakly non-local*, either are inconsistent with autonomy and nearly perfect correlations or imply Bell inequalities (as do local theories). Consequently, the empirical violation of the inequalities does rule out local theories (which is well known from the original argument) *and* these weakly non-local ones (which is the central result of this paper). Showing that the violation of Bell inequalities excludes more theories than the standard Bell argument suggests, the new argument is a considerable strengthening of the original one. It is clear that a strengthening of such a fundamental and influential theorem as Bell's comes with a plethora of new consequences, and, obviously we cannot discuss all of them here. Therefore, we restrict ourselves to four rather immediate and central consequences and one comment.

(1) The new result reveals that *the usual concept of quantum non-locality*, which follows from the standard Bell argument, *is inappropriately weak*. For the latter states a failure of the local factorisation condition

$$P(\alpha\beta|ab\lambda) = P(\alpha|a\lambda)P(\beta|b\lambda), \tag{H^{α}_{29}}$$

which suggest that one could have *any* non-local dependence in the product form, i.e. either a dependence on the distant outcome (in the first factor) or on the distant setting (in the first or second factor). Although the argument is logically correct, its conclusion is not an appropriate characterization of quantum non-locality. Capturing all non-local classes it includes classes which we have found to be compatible with Bell inequalities (weakly non-local classes). For instance, neither is it possible to violate Bell inequalities if there is a dependence on the respective distant setting in each factor, as in the product form

$$P(\alpha\beta|ab\lambda) = P(\alpha|b\lambda)P(\beta|a\lambda); \tag{H^{α}_{22}}$$

nor is this possible, if a dependence on the distant outcome holds as in the product form

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta a\lambda)P(\beta|b\lambda). \tag{H^{α}_{16}}$$

Both constitute weakly non-local classes and imply Bell inequalities.

Concerning the former of these product forms, (H^{α}_{22}), one might take the stance that it is not physically plausible anyway (because it does not involve a dependence on the local settings), and, similarly, many of the logically possible classes that we have taken into account (cf. table 1) might not be very appealing from a physical point of view. (I mention once again that we have not investigated them because we find them all physically plausible, but because, we have been in need of an exhaustive list of logical possibilities, in order to prove that we have drawn the strongest possible consequences; see (4) below.) But the same probably cannot be said of the latter product form, (H^{α}_{16}). It involves dependences on the local settings and a non-local dependence between the outcomes, and therefore belongs to the group of outcome dependent theories, which has found many supporters in the discussion about the consequences of Bell's theorem (cf. Jarrett (1984) and Shimony (1984, 1986) who have introduced the view). Many seem to have found the view attractive that (i) the non-locality following most plausibly from the violation of the Bell inequalities on a probabilistic level essentially consists in a statistical dependence between the outcomes of EPR/B measurements, and that (ii) physically the non-locality is realized by a non-local physical connection between the outcomes (a so called non-separability, according to most authors). Especially, it has been said that (iii) quantum mechanics is an instance of this view.

If, however, what I have argued for in this paper, is correct, claim (i) of this position cannot be true. For we have shown that theories from class (H^{α}_{16}), which involve a non-local dependence between the outcomes, nevertheless imply Bell inequalities. The inevitable conclusion is that *a probabilistic dependence between the outcomes is too weak to violate Bell inequalities.* Proponents of the outcome dependence view are not right in citing outcome dependence when it comes to the question why Bell inequalities are violated. A dependence between the outcomes, as our treatment of (H^{α}_{16}) shows, cannot explain the violation of Bell inequalities. This is one of the central results of this paper.

Focussing on the immediate probabilistic consequences of the Bell argument, in this paper we have not touched claim (ii) that the physical nature of quantum non-locality is a relation of non-separability between the outcomes. It is true, since probabilistic outcome dependence cannot account for a violation of the inequalities, it might seem tempting to immediately conclude that also (ii) must be wrong. But inferring physical or metaphysical relations from probabilistic facts requires careful analysis, since the transition is well known to be vulnerable to fallacies ('correlation is not causation'). For this reason, establishing the right kind of (meta-)physical connection would have required a further lengthy analysis—so here we have to remain tacit on this question. Having said this, it might be interesting to remark that there seem to be good arguments that the current result, that a probabilistic dependence between the outcomes is too weak to explain a violation of the Bell inequalities, most plausibly entails that also a physical connection between the outcomes is not strong enough to account for a violation (Näger 2013).

What about quantum mechanics in this new picture? Quantum mechanics is well-known to be outcome dependent *and* to correctly reproduce the EPR/B correlations—so how does this fit with the present results? An answer can be seen by realizing that my result does not mean that all outcome dependent theories (in the probabilistic sense) imply the inequalities; it just says that a probabilistic dependence between the outcomes *per se* does not suffice to explain a violation. It follows that quantum mechanics cannot only involve a probabilistic dependence between the outcomes, and especially it cannot belong to class (H_{16}^{α}). Rather, the quantum mechanical product form (here: for maximally entangled states),

$$P(\alpha\beta|ab) = P(\alpha|\beta ab)P(\beta) = P(\beta|\alpha ab)P(\alpha) \tag{6}$$

additionally involves a dependence on the distant setting in each first factor—and it is the dependence on both settings in these factors (rather than the dependence between the outcomes), which is crucial for violating the Bell inequalities. This is a second central result of this paper, so let me state it in an appropriately general form:

(2) The new result also provides us with *a positive characterization of quantum non-locality, which is tighter and more informative* than the original one. The necessary condition for being able to violate Bell inequalities we have derived is that *at least one of the factors in the product form involves both settings in its conditionals*, i.e. at least one of the outcomes must depend probabilistically (or functionally, respectively) on both settings. Without such a dependence between an outcome and both settings Bell inequalities cannot be violated. We have called such product forms *strongly non-local*, and their property of depending on both settings *probabilistic Bell contextuality.* It is obvious that the quantum mechanical distribution (6) constitutes an example of such a form, rather than of (H_{16}^{α}).

So what is crucial for being able to violate Bell inequalities is the dependence on both settings. Apart from that, an outcome might depend on the distant outcome or not; whether it does is irrelevant for this purpose.[10] In this sense, some form of dependence

[10] A dependence on the distant outcome, however, does matter when one considers not only the violation of Bell inequalities but the exact quantitative reproduction of EPR/B correlations. Pawlowski et al.

on the distant setting is required (though not necessarily the kind of dependence that is usually called parameter dependence). As a consequence, one might still have a theory that is outcome dependent, but the outcome dependence is not the critical non-local dependence. A full list of logically possible product forms that can violate Bell inequalities can be found in table 1 under the label 'strongly non-local classes' (H_1^α–H_{14}^α). Note that also this positive result is meant in a purely probabilistic sense: in this paper we have only established that at least one of the outcomes must depend *probabilistically* on both settings (but see Näger (2013) for arguments to infer causal dependences from these probabilistic dependences).

We should mention here that there is another approach to quantum non-locality whose result seems to converge with ours. Maudlin (1994, ch. 6; cf. also a recent refinement by Pawlowski et al. 2010) examines the quantum non-locality not via Bell's theorem, but directly investigates the EPR/B correlations by information theoretic methods. He proves that at least one of the outcomes must depend on *information* about both settings. Since (Shannon mutual) information implies correlation, it seems that Maudlin's claim is—at least roughly—in accordance with our results. On the one hand, this is good news because two different methods yielding the same results are good evidence for the stability of a claim.

On the other hand, I stress that there are at least four non-trivial differences between Maudlin's approach and the one we have presented in this paper. First, each proposal has its own, very different methodology. Maudlin analyses the correlations information theoretically and does not connect his considerations to Bell's argument. In contrast, our approach here is in continuity with Bell's thoughts, which have started and shaped the discussion; it develops and strengthens the method that is most common in the debate, viz. the access via Bell inequalities. Second, the information theoretic approaches are stronger in that they provide the amount of information (the quantitative strength of the correlations) that is required in order to reproduce EPR/B correlations. This is an important result providing sufficient criteria for reproducing the correlations. In another sense, however, third, the information theoretic considerations are also weaker than our results in this paper. Maudlin only claims a dependence between an outcome and both settings, whereas we have presented a detailed list of allowed classes with explicit probabilistic expressions, which are open to further probabilistic analysis. Such analyses are relevant in order to make a precise claim about *conditional on which other variables* a dependence of an outcome on its distant setting holds. Another point, fourth, which our argument here reveals, but Maudlin's is tacit on, is given by the following consequence of our approach:

(3) It is a crucial result of our analysis that we have been able to show that the argument we have presented has the *strongest possible conclusion on a qualitative level* (which only takes into account dependences and independences rather than numerical strengths of correlations). This precludes speculations whether the argument could be

(2010) have shown that there must be information about the distant outcome and that information can either be available by a direct correlation (as in the case of quantum mechanics) or be revealed by a hidden variable (which, however, is not available in the case of quantum mechanics).

made even stronger. Our argument yields the strongest possible necessary conditions for violating Bell inequalities on a qualitative level. It has been essential for arriving at this result to have the complete list of logically possible classes (see table 1), because in this way we could be sure not to have neglected possible classes of theories.

(4) We should mention that our considerations also shed new light on Bell inequalities and their meaning: our result shows that *Bell inequalities are not locality conditions* in the sense that, if a probability distribution obeys a Bell inequality, it must be local. In the discussion, Bell inequalities are so closely linked to locality that one could have this impression. Of course, Bell's argument never really justified that view, for the logic of the standard Bell argument is that local factorisation (given autonomy and perfect (anti-)correlations) is merely *sufficient* (and not necessary) for Bell inequalities. Maybe the association between Bell inequalities and locality might have arisen from the fact that up to now local factorisation has been the *only* product form which has been shown to imply Bell inequalities. Given only this information, it was at least possible (though unproven) that the holding of Bell inequalities implies locality. However, since we have shown that some weakly non-local classes in general imply Bell inequalities and since the simulations show that even some strongly non-local distributions can conform to Bell inequalities, it has become explicit that this is not true. *Not all* probability distributions obeying Bell inequalities are *local*.

(5) Finally, we may ask, *why these stronger consequences of the Bell argument*, that we have derived in this paper, *have been overlooked so far.* Obviously, it has wrongly been assumed that local factorisation is the *only* basis to derive Bell inequalities, and the main reason for neglecting other product forms of hidden joint probabilities might have been the fact that, originally, Bell inequalities were derived to capture consequences of a *local* worldview. The question that shaped Bell's original work clearly was Einstein's search for a local hidden variable theory and his main result was that such a theory is impossible: locality has consequences which are in conflict with the quantum mechanical distribution—one cannot have a local hidden variable theory which yields the same predictions as quantum mechanics. Given this historical background, the idea to derive Bell inequalities from non-local assumptions maybe was beyond interest because the conflict with locality was considered to be the crucial point; or maybe it was neglected because Bell inequalities were so tightly associated with locality that a derivation from non-locality sounded totally implausible. Systematically, however, since today it is clear that the quantum mechanical distribution is empirically correct and Bell inequalities are violated, it is desirable to draw as strong consequences as possible from the argument, which requires to check without prejudice whether some non-local classes allow a derivation of Bell inequalities as well. That this is indeed the case is the result of this paper.

Acknowledgements

I would like to thank Frank Arntzenius, Tjorven Hetzger, Gábor Hofer-Szabó, Meinard Kuhlmann, Wayne Myrvold, Thorben Petersen, Manfred Stöckler, Nicola Vona, Adrian

Wüthrich and audiences at the conference 'Philosophy of Physics in Germany' (Hanover, 2010) and at the '14th Congress of Logic, Methodology and Philosophy of Science' (Nancy, 2011) for helpful comments and discussion. I am also grateful to a referee for valuable comments. This paper is a partial result of a research project funded by the Deutsche Forschungsgemeinschaft (DFG).

References

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

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

Bell, J. S. (1971). Introduction to the hidden-variable question. In B. d'Espagnat (Ed.), *Foundations of Quantum Mechanics: Proceedings of the International School of Physics 'Enrico Fermi', Course XLIX*, pp. 171–181. (Reprinted in Bell 1987b). New York: Academic Press.

Bell, J. S. (1975). The theory of local beables. *TH-2053-CERN* (Reprinted in Bell 1987b).

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

Clauser, J. F. and M. A. Horne (1974). Experimental consequences of objective local theories. *Physical Review D 10*(2), 526–535.

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

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

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

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

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

Maudlin, T. (1994). *Quantum Non-locality and Relativity: Metaphysical Intimations of Modern Physics.* Oxford: Blackwell.

Näger, P. M. (2013). Causal graphs for EPR experiments. *Preprint.* http://philsci-archive.pitt.edu/9915/

Näger, P. M. (2015). The causal problem of entanglement. *Synthese*. Advance online publication. doi: 10.1007/s11229-015-0668-6.

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

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

San Pedro, I. (2012). Causation, measurement relevance and no-conspiracy in EPR. *European Journal for Philosophy of Science 2*, 137–156.

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

Shimony, A. (1986). Events and processes in the quantum world. In R. Penrose and C. J. Isham (Eds.), *Quantum Concepts in Space and Time*, pp. 182–203. Oxford: Clarendon Press.

Sutherland, R. I. (1983). Bell's theorem and backwards-in-time causality. *International Journal of Theoretcial Physics 22*(4), 377–384.

Ursin, R., Tiefenbacher, F., Schmitt-Manderbach, T., Weier, H., Scheidl, T., Lindenthal, M., Blauensteiner, B., Jennewein, T., Perdigues, J., Trojek, P., Omer, B., Furst, M., Meyenburg, M., Rarity, J., Sodnik, Z., Barbieri, C., Weinfurter, H., and Zeilinger, Anton (2007). Entanglement-based quantum communication over 144 km. *Nature Physics 3*(7), 481–486.

van Fraassen, B. C. (1989). The charybdis of realism: Epistemological implications of Bell's inequality. In J. T. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, pp. 97–113. (Reprinted with additional comments from Synthese 52, 25–38). Notre Dame: University of Notre Dame Press.

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

Wigner, E. P. (1970). On hidden variables and quantum mechanical probabilities. *American Journal of Physics 38*, 1005–1009.

Appendix

Proof of lemma 1

We proceed by reductio. By autonomy and (H^{α}_{16}) we rewrite the conditions for perfect correlations and for perfect anti-correlations:

$$P(\alpha_\pm\beta_\mp|a_ib_i) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\mp a_i\lambda)P(\beta_\mp|b_i\lambda) \tag{7}$$

$$P(\alpha_\pm\beta_\mp|a_{i_\perp}b_{i_\perp}) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda)P(\beta_\mp|b_{i_\perp}\lambda) \tag{8}$$

$$P(\alpha_\pm\beta_\pm|a_ib_{i_\perp}) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\pm a_i\lambda)P(\beta_\pm|b_{i_\perp}\lambda) \tag{9}$$

$$P(\alpha_\pm\beta_\pm|a_{i_\perp}b_i) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\pm a_{i_\perp}\lambda)P(\beta_\pm|b_i\lambda) \tag{10}$$

Since probabilities are non-negative (and since without loss of generality we can assume $P(\lambda) > 0$ for all λ), at least one of the two remaining factors in each summand must be zero, i.e. for all values i and λ we must have:

$$\begin{array}{rlcll}
 & [\ P(\alpha_+|\beta_- a_i\lambda) = 0 & \vee & P(\beta_-|b_i\lambda) = 0\ ] & (11)\\
\wedge & [\ P(\alpha_-|\beta_+ a_i\lambda) = 0 & \vee & P(\beta_+|b_i\lambda) = 0\ ] & (12)\\
\wedge & [\ P(\alpha_+|\beta_- a_{i_\perp}\lambda) = 0 & \vee & P(\beta_-|b_{i_\perp}\lambda) = 0\ ] & (13)\\
\wedge & [\ P(\alpha_-|\beta_+ a_{i_\perp}\lambda) = 0 & \vee & P(\beta_+|b_{i_\perp}\lambda) = 0\ ] & (14)\\
\wedge & [\ P(\alpha_+|\beta_+ a_i\lambda) = 0 & \vee & P(\beta_+|b_{i_\perp}\lambda) = 0\ ] & (15)\\
\wedge & [\ P(\alpha_-|\beta_- a_i\lambda) = 0 & \vee & P(\beta_-|b_{i_\perp}\lambda) = 0\ ] & (16)\\
\wedge & [\ P(\alpha_+|\beta_+ a_{i_\perp}\lambda) = 0 & \vee & P(\beta_+|b_i\lambda) = 0\ ] & (17)\\
\wedge & [\ P(\alpha_-|\beta_- a_{i_\perp}\lambda) = 0 & \vee & P(\beta_-|b_i\lambda) = 0\ ] & (18)
\end{array}$$

From these conditions one can infer that all involved probabilities must be 0 or 1 (determinism). More precisely, for every i and λ one of the following two cases holds:

Case 1: $P(\alpha_+|\beta_- a_i \lambda) = 0$

$$
\begin{array}{rrrr}
\overset{(CE)}{\Rightarrow} & P(\alpha_-|\beta_- a_i \lambda) = 1 & \overset{(16)}{\Rightarrow} & P(\beta_-|b_{i_\perp} \lambda) = 0 \\
\overset{(CE)}{\Rightarrow} & P(\beta_+|b_{i_\perp} \lambda) = 1 & \overset{(14)}{\underset{(15)}{\Rightarrow}} & P(\alpha_-|\beta_+ a_{i_\perp} \lambda) = 0 \\
 & & & \wedge\ P(\alpha_+|\beta_+ a_i \lambda) = 0 \\
\overset{(CE)}{\Rightarrow} & P(\alpha_+|\beta_+ a_{i_\perp} \lambda) = 1 & \overset{(17)}{\underset{(12)}{\Rightarrow}} & P(\beta_+|b_i \lambda) = 0 \\
 & \wedge\ P(\alpha_-|\beta_+ a_i \lambda) = 1 & & \\
\overset{(CE)}{\Rightarrow} & P(\beta_-|b_i \lambda) = 1 & \overset{(18)}{\Rightarrow} & P(\alpha_-|\beta_- a_{i_\perp} \lambda) = 0 \\
\overset{(CE)}{\Rightarrow} & P(\alpha_+|\beta_- a_{i_\perp} \lambda) = 1 & &
\end{array}
$$

NB: (CE) stands for the following theorem of probability theory: $P(A|B) + P(\bar{A}|B) = 1$.

Case 2: $P(\alpha_+|\beta_- a_i \lambda) > 0$

$$
\begin{array}{rrrr}
 & & \overset{(11)}{\Rightarrow} & P(\beta_-|b_i \lambda) = 0 \\
\overset{(CE)}{\Rightarrow} & P(\beta_+|b_i \lambda) = 1 & \overset{(12)}{\underset{(17)}{\Rightarrow}} & P(\alpha_-|\beta_+ a_i \lambda) = 0 \\
 & & & \wedge\ P(\alpha_+|\beta_+ a_{i_\perp} \lambda) = 0 \\
\overset{(CE)}{\Rightarrow} & P(\alpha_+|\beta_+ a_i \lambda) = 1 & \overset{(15)}{\underset{(14)}{\Rightarrow}} & P(\beta_+|b_{i_\perp} \lambda) = 0 \\
 & \wedge\ P(\alpha_-|\beta_+ a_{i_\perp} \lambda) = 1 & & \\
\overset{(CE)}{\Rightarrow} & P(\beta_-|b_{i_\perp} \lambda) = 1 & \overset{(13)}{\underset{(16)}{\Rightarrow}} & P(\alpha_+|\beta_- a_{i_\perp} \lambda) = 0 \\
 & & & \wedge\ P(\alpha_-|\beta_- a_i \lambda) = 0 \\
\overset{(CE)}{\Rightarrow} & P(\alpha_-|\beta_- a_{i_\perp} \lambda) = 1 & & \\
 & \wedge\ P(\alpha_+|\beta_- a_i \lambda) = 1 & &
\end{array}
$$

Since in each case we have

$$P(\alpha_+|\beta_+ a_i \lambda) = P(\alpha_+|\beta_- a_i \lambda) \tag{19}$$
$$P(\alpha_-|\beta_+ a_i \lambda) = P(\alpha_-|\beta_- a_i \lambda) \tag{20}$$
$$P(\alpha_+|\beta_+ a_{i_\perp} \lambda) = P(\alpha_+|\beta_- a_{i_\perp} \lambda) \tag{21}$$
$$P(\alpha_-|\beta_+ a_{i_\perp} \lambda) = P(\alpha_-|\beta_- a_{i_\perp} \lambda) \tag{22}$$

it is true that

$$\forall \alpha, \beta, a, \lambda : \quad P(\alpha|\beta a \lambda) = P(\alpha|a\lambda). \tag{23}$$

By this statistical independence the product form (H_{16}^{α})

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta a \lambda) P(\beta|b\lambda) \tag{24}$$

loses its dependence on the outcome β in the first factor, i.e. it reads

$$P(\alpha\beta|ab\lambda) = P(\alpha|a\lambda) P(\beta|b\lambda). \tag{25}$$

This, however, is the well known local product form (local factorization), contradicting the assumption that we have the non-local product form (H_{16}^{α}).

Note that this proof makes essential use of the perfect correlations and perfect antiocorrelations (7–10), i.e. the probabilities $P(\alpha_\pm\beta_\mp|a_i b_i)$, $P(\alpha_\pm\beta_\mp|a_{i_\perp} b_{i_\perp})$, $P(\alpha_\pm\beta_\pm|a_i b_{i_\perp})$ and $P(\alpha_\pm\beta_\pm|a_{i_\perp} b_i)$ have to be strictly 0. If these conditions are only slightly relaxed, i.e. if any of these probabilities takes on a positive value, even if very small, the conclusion does not follow. q.e.d.

Proof of lemma 2

By autonomy and (H_{16}^{α}) we rewrite (some of) the conditions for nearly perfect correlations and for nearly perfect anti-correlations:

$$P(\alpha_\pm\beta_\mp|a_i b_i) = \delta_{ii} \quad = \sum_\lambda P(\lambda) P(\alpha_\pm|\beta_\mp a_i \lambda) P(\beta_\mp|b_i\lambda) \tag{26}$$
$$P(\alpha_\pm\beta_\pm|a_i b_{i_\perp}) = \delta_{ii_\perp} = \sum_\lambda P(\lambda) P(\alpha_\pm|\beta_\pm a_i \lambda) P(\beta_\pm|b_{i_\perp}\lambda), \tag{27}$$

where δ_{ii} and $\delta_{ii_\perp}$ are positive and small. Since probabilities are non-negative, all summands are non-negative; so each summand must be less or equal than the total value of the sum:

$$P(\lambda) P(\alpha_\pm|\beta_\mp a_i \lambda) P(\beta_\mp|b_i\lambda) \leq \delta_{ii} \tag{28}$$
$$P(\lambda) P(\alpha_\pm|\beta_\pm a_i \lambda) P(\beta_\pm|b_{i_\perp}\lambda) \leq \delta_{ii_\perp} \tag{29}$$

In order to facilitate the following considerations, let us define

$$\epsilon := \max_{i=1,2,3}(\sqrt[3]{\delta_{ii}}, \sqrt[3]{\delta_{ii_\perp}}) \tag{30}$$

where $i = 1, 2, 3$ represent three distinct measurement directions. We can then write:

$$P(\lambda)P(\alpha_\pm|\beta_\mp a_i\lambda)P(\beta_\mp|b_i\lambda) \leq \epsilon^3 \tag{31}$$

$$P(\lambda)P(\alpha_\pm|\beta_\pm a_i\lambda)P(\beta_\pm|b_{i_\perp}\lambda) \leq \epsilon^3. \tag{32}$$

Since a product of three non-negative factors is never smaller than the cube root of its smallest factor, each product must contain (at least) one factor that is less or equal than ϵ, i.e. for all values i and λ we must have:

$$\big[P(\lambda) \leq \epsilon \quad \vee \quad P(\alpha_+|\beta_- a_i\lambda) \leq \epsilon \quad \vee \quad P(\beta_-|b_i\lambda) \leq \epsilon\big] \tag{33}$$

$$\wedge \quad \big[P(\lambda) \leq \epsilon \quad \vee \quad P(\alpha_-|\beta_+ a_i\lambda) \leq \epsilon \quad \vee \quad P(\beta_+|b_i\lambda) \leq \epsilon\big] \tag{34}$$

$$\wedge \quad \big[P(\lambda) \leq \epsilon \quad \vee \quad P(\alpha_+|\beta_+ a_i\lambda) \leq \epsilon \quad \vee \quad P(\beta_+|b_{i_\perp}\lambda) \leq \epsilon\big] \tag{35}$$

$$\wedge \quad \big[P(\lambda) \leq \epsilon \quad \vee \quad P(\alpha_-|\beta_- a_i\lambda) \leq \epsilon \quad \vee \quad P(\beta_-|b_{i_\perp}\lambda) \leq \epsilon\big] \tag{36}$$

There are three cases that solve these conditions:

Case 1: $P(\lambda) > \epsilon \ \wedge \ P(\alpha_+|\beta_- a_i\lambda) \leq \epsilon$

$$\overset{\text{(CE)}}{\Rightarrow} P(\alpha_-|\beta_- a_i\lambda) > 1-\epsilon \qquad \overset{(36)}{\Rightarrow} P(\beta_-|b_{i_\perp}\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\beta_+|b_{i_\perp}\lambda) > 1-\epsilon \qquad \overset{(35)}{\Rightarrow} P(\alpha_+|\beta_+ a_i\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\alpha_-|\beta_+ a_i\lambda) > 1-\epsilon \qquad \overset{(34)}{\Rightarrow} P(\beta_+|b_i\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\beta_-|b_i\lambda) > 1-\epsilon$$

Case 2: $P(\lambda) > \epsilon \ \wedge \ P(\alpha_+|\beta_- a_i\lambda) > \epsilon$

$$\overset{(33)}{\Rightarrow} P(\beta_-|b_i\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\beta_+|b_i\lambda) > 1-\epsilon \qquad \overset{(34)}{\Rightarrow} P(\alpha_-|\beta_+ a_i\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\alpha_+|\beta_+ a_i\lambda) > 1-\epsilon \qquad \overset{(35)}{\Rightarrow} P(\beta_+|b_{i_\perp}\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\beta_-|b_{i_\perp}\lambda) > 1-\epsilon \qquad \overset{(36)}{\Rightarrow} P(\alpha_-|\beta_- a_i\lambda) \leq \epsilon$$

$$\overset{\text{(CE)}}{\Rightarrow} P(\alpha_+|\beta_- a_i\lambda) > 1-\epsilon$$

Case 3: $P(\lambda) \leq \epsilon$

(no particular restrictions for other probabilities)

The three cases are disjunct and define a partition of the values of $\boldsymbol{\lambda}$:

$$\Lambda_1(i) := \{\lambda | P(\lambda) > \epsilon \ \wedge \ P(\alpha_+ | \beta_- a_i \lambda) \leq \epsilon)\}$$

$$\Lambda_2(i) := \{\lambda | P(\lambda) > \epsilon \ \wedge \ P(\alpha_+ | \beta_- a_i \lambda) \geq 1 - \epsilon)\}$$

$$\Lambda_3(i) := \{\lambda | P(\lambda) \leq \epsilon\} = \Lambda_3$$

Note that each value i defines a different partition, but that $\Lambda_3(i) = \Lambda_3$ is independent of i.

We can use the fact that the $\boldsymbol{\lambda}$-partitions depend on just *one* setting i to estimate values for the hidden joint probability $P(\alpha\beta|ab\lambda)$ for *any* choice of measurement directions $a_i b_j$ by forming intersections of partitions for different settings (see table 2). Note that the table only comprises five of the nine combinatorially possible cases; the ignored cases are empty sets ($\Lambda_1(i) \wedge \Lambda_3 = \emptyset$ because $\Lambda_1(i)$ requires $P(\lambda) > \epsilon$, whereas Λ_3 implies $P(\lambda) \leq \epsilon$; and analogously $\Lambda_2(i) \wedge \Lambda_3 = \emptyset$, $\Lambda_3 \wedge \Lambda_1(j) = \emptyset$, $\Lambda_3 \wedge \Lambda_2(j) = \emptyset$). The last column is defined as $\Lambda_3(i) \cap \Lambda_3(j) = \Lambda_3$, and the label 'n.r.' means 'no restriction', i.e. the value of the hidden joint probability is not confined to any specific interval; rather, in this set it is the case that $P(\lambda) \leq \epsilon$.

Table 2: Values of the hidden joint probability

	$\lambda \in$				
	$\Lambda_1(i) \cap \Lambda_1(j)$	$\Lambda_1(i) \cap \Lambda_2(j)$	$\Lambda_2(i) \cap \Lambda_1(j)$	$\Lambda_2(i) \cap \Lambda_2(j)$	Λ_3
$P(\alpha_+\beta_+|a_ib_j\lambda)$	$\leq \epsilon^2$	$\leq \epsilon$	$\leq \epsilon$	$> (1-\epsilon)^2$	n.r.
$P(\alpha_+\beta_-|a_ib_j\lambda)$	$\leq \epsilon$	$\leq \epsilon^2$	$> (1-\epsilon)^2$	$\leq \epsilon$	n.r.
$P(\alpha_-\beta_+|a_ib_j\lambda)$	$\leq \epsilon$	$> (1-\epsilon)^2$	$\leq \epsilon^2$	$\leq \epsilon$	n.r.
$P(\alpha_-\beta_-|a_ib_j\lambda)$	$> (1-\epsilon)^2$	$\leq \epsilon$	$\leq \epsilon$	$\leq \epsilon^2$	n.r.

Given the estimations for the hidden joint probability in table 2 one can derive a generalised Wigner-Bell inequality. Consider the inequality

$$P(X \cap \overline{Z}) \leq P(X \cap \overline{Y}) + P(Y \cap \overline{Z}), \tag{37}$$

which in general holds for any events X, Y, Z of a measurable space, as can easily be seen by rewriting the involved probabilities:

$$P(X \cap \overline{Z}) = P(X \cap Y \cap \overline{Z}) + P(X \cap \overline{Y} \cap \overline{Z}) \tag{38}$$
$$P(X \cap \overline{Y}) = P(X \cap \overline{Y} \cap Z) + P(X \cap \overline{Y} \cap \overline{Z}) \tag{39}$$
$$P(Y \cap \overline{Z}) = P(X \cap Y \cap \overline{Z}) + P(\overline{X} \cap Y \cap \overline{Z}) \tag{40}$$

Assuming $X = \Lambda_1(1) \cup \Lambda_3$, $Y = \Lambda_1(2) \cup \Lambda_3$ and $Z = \Lambda_1(3)$ gives the inequality

$$P([\Lambda_1(1) \cup \Lambda_3] \cap \overline{\Lambda_1(3)}) \leq P([\Lambda_1(1) \cup \Lambda_3] \cap \overline{[\Lambda_1(2) \cup \Lambda_3]}) + P([\Lambda_1(2) \cup \Lambda_3] \cap \overline{\Lambda_1(3)}). \tag{41}$$

We calculate the sets involved in the inequality:

$$\begin{aligned}
[\Lambda_1(i) \cup \Lambda_3] \cap \overline{\Lambda_1(j)} &= [\Lambda_1(i) \cup \Lambda_3] \cap [\Lambda_2(j) \cup \Lambda_3] = \\
&= [\Lambda_1(i) \cap \Lambda_2(j)] \cup \underbrace{[\Lambda_1(i) \cap \Lambda_3]}_{\emptyset} \cup \underbrace{[\Lambda_3 \cap \Lambda_2(j)]}_{\emptyset} \cup \underbrace{[\Lambda_3 \cap \Lambda_3]}_{\Lambda_3} \\
&= [\Lambda_1(i) \cap \Lambda_2(j)] \cup \Lambda_3
\end{aligned} \tag{42}$$

$$\begin{aligned}
[\Lambda_1(i) \cup \Lambda_3] \cap \overline{[\Lambda_1(j) \cup \Lambda_3]} &= [\Lambda_1(i) \cup \Lambda_3] \cap \Lambda_2(j) \\
&= [\Lambda_1(i) \cap \Lambda_2(j)] \cup \underbrace{[\Lambda_3 \cap \Lambda_2(j)]}_{\emptyset} \\
&= \Lambda_1(i) \cap \Lambda_2(j)
\end{aligned} \tag{43}$$

If we further define the shorthand

$$\Lambda_{kl}(i,j) := \Lambda_k(i) \cap \Lambda_l(j), \tag{44}$$

we can rewrite inequality (41) as

$$P(\Lambda_{12}(1,3) \cup \Lambda_3) \leq P(\Lambda_{12}(1,2)) + P(\Lambda_{12}(2,3) \cup \Lambda_3). \tag{45}$$

This inequality can be transformed to yield a generalized Wigner-Bell inequality. We have to rewrite the inequality such that it only involves empirically accessible probabilities, i.e. probabilities that do not involve the hidden state $\boldsymbol{\lambda}$, and this can be done by using the estimates for the hidden joint probability from table 2. Especially, we have to find a lower estimate for the left hand side of the inequality and an upper estimate for its right hand side. We start by deriving the former:

$$\begin{aligned}
P(\Lambda_{12}(1,3) \cup \Lambda_3) &\overset{(\sigma\text{-additivity})}{=} \sum_{\lambda \in \Lambda_{12}(1,3) \cup \Lambda_3} P(\lambda) \\
&\geq \sum_{\lambda \in \Lambda_{12}(1,3) \cup \Lambda_3} P(\lambda) P(\alpha_- \beta_+ | a_1 b_3 \lambda) \\
&= \sum_{\lambda \in \Lambda} P(\lambda) P(\alpha_- \beta_+ | a_1 b_3 \lambda) - \sum_{\lambda \in \Lambda \backslash [\Lambda_{12}(1,3) \cup \Lambda_3]} P(\lambda) P(\alpha_- \beta_+ | a_1 b_3 \lambda) \\
&= P(\alpha_- \beta_+ | a_1 b_3) - \sum_{\lambda \in \Lambda_{11}(1,3)} P(\lambda) P(\alpha_- \beta_+ | a_1 b_3 \lambda) \\
&\quad - \sum_{\lambda \in \Lambda_{21}(1,3)} P(\lambda) P(\alpha_- \beta_+ | a_1 b_3 \lambda) - \sum_{\lambda \in \Lambda_{22}(1,3)} P(\lambda) P(\alpha_- \beta_+ | a_1 b_3 \lambda) \\
&\overset{(\text{table } 2)}{\geq} P(\alpha_- \beta_+ | a_1 b_3) - \epsilon \sum_{\lambda \in \Lambda_{11}(1,3)} P(\lambda) - \epsilon^2 \sum_{\lambda \in \Lambda_{21}(1,3)} P(\lambda) - \epsilon \sum_{\lambda \in \Lambda_{22}(1,3)} P(\lambda) \\
&\geq P(\alpha_- \beta_+ | a_1 b_3) - 2\epsilon - \epsilon^2
\end{aligned} \tag{46}$$

An upper estimate for the right hand side of (45) can be calculated as follows:

$$
\begin{aligned}
&P(\Lambda_{12}(1,2)) + P(\Lambda_{12}(2,3)\cup\Lambda_3) = && (47)\\
&\overset{(\sigma\text{-additivity})}{=} \sum_{\lambda\,\in\,\Lambda_{12}(1,2)} P(\lambda) + \sum_{\lambda\,\in\,\Lambda_{12}(2,3)\cup\Lambda_3} P(\lambda) \\
&\leq \sum_{\lambda\,\in\,\Lambda_{12}(1,2)} P(\lambda)\frac{P(\alpha_-\beta_+|a_1b_2\lambda)}{(1-\epsilon)^2} + \sum_{\lambda\,\in\,\Lambda_{12}(2,3)\cup\Lambda_3} P(\lambda)\frac{P(\alpha_-\beta_+|a_2b_3\lambda)}{(1-\epsilon)^2} \\
&\leq \sum_{\lambda\,\in\,\Lambda} P(\lambda)\frac{P(\alpha_-\beta_+|a_1b_2\lambda)}{(1-\epsilon)^2} + \sum_{\lambda\,\in\,\Lambda} P(\lambda)\frac{P(\alpha_-\beta_+|a_2b_3\lambda)}{(1-\epsilon)^2} \\
&= \frac{P(\alpha_-\beta_+|a_1b_2\lambda) + P(\alpha_-\beta_+|a_2b_3\lambda)}{(1-\epsilon)^2} && (48)
\end{aligned}
$$

The resulting inequality

$$P(\alpha_-\beta_+|a_1b_3) - 2\epsilon - \epsilon^2 \leq \frac{P(\alpha_-\beta_+|a_1b_2\lambda) + P(\alpha_-\beta_+|a_2b_3\lambda)}{(1-\epsilon)^2} \tag{49}$$

is the Wigner-Bell inequality we have been looking for. It generalizes usual Wigner-Bell inequalities such as

$$P(\alpha_-\beta_+|a_1b_3) \leq P(\alpha_-\beta_+|a_1b_2\lambda) + P(\alpha_-\beta_+|a_2b_3\lambda) \tag{50}$$

in that it introduces correction terms with the parameter ϵ. It is an inequality of fourth order in ϵ and one can check numerically that it is violated by the empirical measurement results

$$P(\alpha_-\beta_+|a_1b_3) = 0.375, \qquad P(\alpha_-\beta_+|a_1b_2) = 0.125, \qquad P(\alpha_-\beta_+|a_2b_3) = 0.125, \tag{51}$$

(which are a maximal violation of the usual Wigner-Bell inequality and occur e.g. for the measurement settings being chosen as $1 = 0°$, $2 = 30°$, $3 = 60°$ given the quantum state ψ_0), if

$$0 < \epsilon < 0.048328 \tag{52}$$

Hence, the maximal deviation of perfect correlations for the generalized Wigner-Bell inequality still to be violated is

$$\delta = (\epsilon_{\max})^3 = 0.048328^3 = 1.1280 \cdot 10^{-4}, \tag{53}$$

i.e. at least 99.989% of the photons must be perfectly correlated and anti-correlated.

q.e.d.

Proof of theorem 1.1

We split the theorem up into three partial claims:

Claim 1: Autonomy, perfect correlations, perfect anti-correlations and a class of probability distributions (H_i^α) form an inconsistent set if (i) the product form of (H_i^α) involves at most one of the settings.

Claim 2: Autonomy, perfect correlations, perfect anti-correlations and a class of probability distributions (H_i^α) form an inconsistent set if (ii) the product form of (H_i^α) involves both settings but its first factor involves the distant outcome and at most one setting.

Claim 3: A class (H_i^α) is consistent with autonomy, perfect correlations and perfect anti-correlations if ($\neg$i) the product form of (H_i^α) involves both settings and ($\neg$ii) in case the distant outcome appears in the first factor of (H_i^α)'s product form, also both settings appear in that factor.

<u>Proof of claim 1</u>

Condition (i), that the product form involves at most one of the settings, is fulfilled by the classes $\{(\mathrm{H}_{17}^\alpha), \ldots, (\mathrm{H}_{32}^\alpha)\} \backslash \{(\mathrm{H}_{22}^\alpha), (\mathrm{H}_{29}^\alpha)\}$. Here we have to show the inconsistency of these classes with the set of assumptions autonomy, perfect correlations and perfect anti-correlations.

Consider, for instance,

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta a\lambda)P(\beta|a\lambda) = P(\alpha\beta|a\lambda), \tag{H_{17}^α}$$

which fails to involve the setting $\boldsymbol{b}$. It is easy to show that this product form can neither account for the perfect correlations nor for the perfect anti-correlations. The perfect correlations read:

$$P(\alpha_\pm\beta_\pm|a_ib_i) = \frac{1}{2} \qquad\qquad P(\alpha_\pm\beta_\pm|a_ib_{i_\perp}) = 0. \tag{54}$$

Now, the value of these empirical probabilities depends crucially on the value of the setting $\boldsymbol{b}$. However, one can demonstrate without much effort that (H_{17}^α)'s failure to involve the setting $\boldsymbol{b}$ on a hidden level, extends to the empirical level, if one assumes autonomy:

$$\begin{aligned} P(\alpha\beta|ab) = \sum_\lambda P(\lambda|ab)P(\alpha\beta|ab\lambda) &\overset{(\mathrm{A})}{=} \sum_\lambda P(\lambda|ab')P(\alpha\beta|ab\lambda) = \\ &\overset{(\mathrm{H}_{17}^\alpha)}{=} \sum_\lambda P(\lambda|ab')P(\alpha\beta|ab'\lambda) = P(\alpha\beta|ab') \end{aligned} \tag{55}$$

This implies that according to (H_{17}^α) all empirical probabilities $P(\alpha\beta|ab)$ that only differ by their value for the setting $\boldsymbol{b}$ must equal another—which obviously contradicts (54). For the same reason, (H_{17}^α) contradicts the perfect anti-correlations

$$P(\alpha_\pm\beta_\mp|a_ib_{i_\perp}) = \frac{1}{2} \qquad\qquad P(\alpha_\pm\beta_\mp|a_ib_i) = 0. \tag{56}$$

In the same way, all other product forms that do not involve the setting $\boldsymbol{b}$ are in conflict with the perfect correlations (54) and perfect anti-correlations (56), and, similarly, all product forms that fail to involve the setting $\boldsymbol{a}$ are in conflict with the perfect correlations

$$P(\alpha_\pm\beta_\pm|a_ib_i) = \frac{1}{2} \qquad\qquad P(\alpha_\pm\beta_\pm|a_{i_\perp}b_i) = 0. \tag{57}$$

or the perfect anti-correlations

$$P(\alpha_\pm\beta_\mp|a_{i_\perp}b_i) = \frac{1}{2} \qquad\qquad P(\alpha_\pm\beta_\mp|a_ib_i) = 0. \tag{58}$$

<u>Proof of claim 2</u>

Condition (ii), that the product form involves both settings but its first factor involves the distant outcome and at most one setting, is fulfilled by the product forms (H_4^α), (H_5^α), (H_{10}^α), (H_{15}^α) and (H_{16}^α). Here we have to show the inconsistency of these classes with the set of assumptions autonomy, perfect correlations and perfect anti-correlations.

By lemma 1 we have already proven that (H_{16}^α)

$$P(\alpha\beta|ab\lambda) = P(\alpha|\beta a\lambda)P(\beta|b\lambda) \tag{59}$$

forms an inconsistent set with autonomy, perfect correlations and perfect anti-correlations. Mutatis mutandis, also the classes (H_{10}^α) and (H_{15}^α) lead to a similar inconsistency. In each case the product form looses its dependence on the distant outcome in the first factor, i.e. (H_{10}^α) reduces to (H_{14}^α), whereas (H_{15}^α) reduces to (H_{22}^α).

The proofs against the classes (H_4^α) and (H_5^α) work in a similar way, but require a little more care due to an additional case differentiation. Let me shortly demonstrate this for class (H_5^α). As for (H_{16}^α) one starts with expressing the perfect (anti-)correlations in terms of the product form,

$$P(\alpha_\pm\beta_\mp|a_ib_i) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\mp a_i\lambda)P(\beta_\mp|a_ib_i\lambda) \tag{60}$$

$$P(\alpha_\pm\beta_\mp|a_{i_\perp}b_{i_\perp}) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda)P(\beta_\mp|a_{i_\perp}b_{i_\perp}\lambda) \tag{61}$$

$$P(\alpha_\pm\beta_\pm|a_ib_{i_\perp}) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\pm a_i\lambda)P(\beta_\pm|a_ib_{i_\perp}\lambda) \tag{62}$$

$$P(\alpha_\pm\beta_\pm|a_{i_\perp}b_i) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|\beta_\pm a_{i_\perp}\lambda)P(\beta_\pm|a_{i_\perp}b_i\lambda). \tag{63}$$

In the case of (H_{16}^α) there were two cases, defined by $P(\alpha_\pm|\beta_\mp a_i\lambda) = 0$ or $P(\alpha_\pm|\beta_\mp a_i\lambda) > 0$, respectively, and all other probabilities followed from each of these defining probabilities. In the present case, however, when, accordingly, we assume $P(\alpha_\pm|\beta_\mp a_i\lambda) = 0$ or $P(\alpha_\pm|\beta_\mp a_i\lambda) > 0$, respectively, only the factors of the hidden joint probabilities on the right hand side of equations (60) and (62) are implied, i.e. the probabilities in (61) and (63) remain undetermined by these assumptions (due to the fact that there are *two* settings in the second factor of the product form). The latter probabilities have to be determined by further assumptions, e.g. by setting $P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda) = 0$ or $P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda) > 0$, respectively. These assumptions introduce two new cases, that are logically independent of the former two. In total, this makes four cases (instead of two):

$$P(\alpha_\pm|\beta_\mp a_i\lambda) = 0 \quad \wedge \quad P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda) = 0 \tag{64}$$

$$P(\alpha_\pm|\beta_\mp a_i\lambda) = 0 \quad \wedge \quad P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda) > 0 \tag{65}$$

$$P(\alpha_\pm|\beta_\mp a_i\lambda) > 0 \quad \wedge \quad P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda) = 0 \tag{66}$$

$$P(\alpha_\pm|\beta_\mp a_i\lambda) > 0 \quad \wedge \quad P(\alpha_\pm|\beta_\mp a_{i_\perp}\lambda) > 0 \tag{67}$$

While this renders the proof slightly more complex, the crucial fact to mention here is that in all four cases we have

$$\forall \alpha, \beta, a, \lambda : \quad P(\alpha|\beta a\lambda) = P(\alpha|a\lambda), \tag{68}$$

i.e. (H_5^α) reduces to (H_{12}^α). Similarly, one can show that (H_4^α) reduces to (H_{11}^α).

Proof of claim 3

Condition $(\neg\mathrm{i})$ *and* $(\neg\mathrm{ii})$, that the product form involves both settings and in case the distant outcome appears in the first factor, also both settings appear in that factor, is fulfilled by the product forms $\{(\mathrm{H}_1^\alpha), \ldots, (\mathrm{H}_{14}^\alpha)\}\backslash\{(\mathrm{H}_4^\alpha), (\mathrm{H}_5^\alpha), (\mathrm{H}_{10}^\alpha)\}$, (H_{22}^α) and (H_{29}^α). Here we have to show the consistency of these classes with the set of assumptions autonomy, perfect correlations and perfect anti-correlations.

Since a class being inconsistent with certain assumptions means that every distribution of a class contradicts the assumptions, a class being consistent means that there is at least one probability distribution in that class which is compatible with the assumptions. Hence, what we need for each of these classes in order to show their consistency with the background assumptions, is one example of a probability distribution belonging to that class that respects the background assumptions. In fact, such examples are easy to construct. Let me demonstrate the procedure with one of the weakest classes in that group, (H_{29}^α), whose product form is local factorization.

Requiring just *any* example we can presuppose a minimal setup, i.e. the hidden variable as well as each setting can be assumed to have only two possible values: $\boldsymbol{\lambda} = \lambda_1, \lambda_2$, $\boldsymbol{a} = a_i, a_{i_\perp}$ and $\boldsymbol{b} = b_i, b_{i_\perp}$ with $a_i = b_i$ and $a_{i_\perp} = b_{i_\perp}$. We start by writing down the perfect correlations and perfect anti-correlations, and express the probabilities on the empirical level by the probabilities on the hidden level using the product form and autonomy:

$$P(\alpha_\pm\beta_\mp|a_ib_i) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|a_i\lambda)P(\beta_\mp|b_i\lambda) \tag{69}$$
$$P(\alpha_\pm\beta_\mp|a_{i_\perp}b_{i_\perp}) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|a_{i_\perp}\lambda)P(\beta_\mp|b_{i_\perp}\lambda) \tag{70}$$
$$P(\alpha_\pm\beta_\pm|a_{i_\perp}b_i) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|a_{i_\perp}\lambda)P(\beta_\pm|b_i\lambda) \tag{71}$$
$$P(\alpha_\pm\beta_\pm|a_ib_{i_\perp}) = 0 = \sum_\lambda P(\lambda)P(\alpha_\pm|a_i\lambda)P(\beta_\pm|b_{i_\perp}\lambda). \tag{72}$$

$$P(\alpha_\pm\beta_\pm|a_ib_i) = \frac{1}{2} = \sum_\lambda P(\lambda)P(\alpha_\pm|a_i\lambda)P(\beta_\pm|b_i\lambda) \tag{73}$$
$$P(\alpha_\pm\beta_\pm|a_{i_\perp}b_{i_\perp}) = \frac{1}{2} = \sum_\lambda P(\lambda)P(\alpha_\pm|a_{i_\perp}\lambda)P(\beta_\pm|b_{i_\perp}\lambda) \tag{74}$$
$$P(\alpha_\pm\beta_\mp|a_{i_\perp}b_i) = \frac{1}{2} = \sum_\lambda P(\lambda)P(\alpha_\pm|a_{i_\perp}\lambda)P(\beta_\mp|b_i\lambda) \tag{75}$$
$$P(\alpha_\pm\beta_\mp|a_ib_{i_\perp}) = \frac{1}{2} = \sum_\lambda P(\lambda)P(\alpha_\pm|a_i\lambda)P(\beta_\mp|b_{i_\perp}\lambda). \tag{76}$$

Then choose a value for any of the probabilities on the right hand side that does not lead into inconsistencies, e.g.

$$P(\alpha_+|a_i\lambda_1) = 0. \tag{77}$$

By (69)–(72) this entails the following probabilities:

$$\overset{\text{(CE)}}{\Rightarrow} \quad P(\alpha_-|a_i\lambda_1) = 1 \quad \underset{(72)}{\overset{(69)}{\Rightarrow}} \quad P(\beta_+|b_i\lambda_1) = 0 \tag{78}$$
$$\wedge\ P(\beta_-|b_{i_\perp}\lambda_1) = 0 \tag{79}$$
$$\overset{\text{(CE)}}{\Rightarrow} \quad P(\beta_-|b_i\lambda_1) = 1 \quad \underset{(70)}{\overset{(71)}{\Rightarrow}} \quad P(\alpha_-|a_{i_\perp}\lambda_1) = 0 \tag{80}$$
$$\wedge\ P(\beta_+|b_{i_\perp}\lambda_1) = 1 \tag{81}$$
$$\overset{\text{(CE)}}{\Rightarrow} \quad P(\alpha_+|a_{i_\perp}\lambda_1) = 1 \tag{82}$$

Similarly, choose a value for the corresponding probability conditional on λ_2, e.g.

$$P(\alpha_+|a_i\lambda_2) = 1 \tag{83}$$

and draw the appropriate consequences:

$$\overset{(69)}{\underset{(72)}{\Rightarrow}} \quad P(\beta_-|b_i\lambda_2) = 0 \tag{84}$$

$$\wedge \ P(\beta_+|b_{i_\perp}\lambda_2) = 0 \tag{85}$$

$$\overset{(\mathrm{CE})}{\Rightarrow} \quad P(\beta_+|b_i\lambda_2) = 1 \quad \overset{(69),(72)}{\underset{(70),(71)}{\Rightarrow}} \quad P(\alpha_-|a_i\lambda_2) = 0 \tag{86}$$

$$\wedge \ P(\beta_-|b_{i_\perp}\lambda_2) = 1 \qquad P(\alpha_+|a_{i_\perp}\lambda_2) = 0 \tag{87}$$

$$\overset{(\mathrm{CE})}{\Rightarrow} \quad P(\alpha_-|a_{i_\perp}\lambda_2) = 1 \tag{88}$$

These probabilities determine the values of the hidden joint probabilities consistent with equations (69)–(72). Note that we have

$$\forall\alpha,\lambda: \quad P(\alpha|a_i\lambda) \neq P(\alpha|a_{i_\perp}\lambda) \qquad \forall\alpha, a: \quad P(\alpha|a\lambda_1) \neq P(\alpha|a\lambda_2) \tag{89}$$

$$\forall\beta,\lambda: \quad P(\beta|b_i\lambda) \neq P(\beta|b_{i_\perp}\lambda) \qquad \forall\beta, b: \quad P(\beta|b\lambda_1) \neq P(\beta|b\lambda_2), \tag{90}$$

which means that the product form does not reduce to any other product form (i.e. the product form is consistent with the assumptions so far).

Inserting the determined values of the hidden joint probability into equations (73)–(76) yields:

$$P(\lambda_1) = \frac{1}{2} \qquad P(\lambda_2) = \frac{1}{2} \tag{91}$$

Finally we can freely choose, say,

$$P(a_i) = \frac{1}{2} = P(a_{i_\perp}) \qquad P(b_i) = \frac{1}{2} = P(b_{i_\perp}) \tag{92}$$

such that by the formula

$$P(\alpha\beta ab\lambda) = P(\alpha|a\lambda)P(\beta|b\lambda)P(\lambda)P(a)P(b) \tag{93}$$

we arrive at the following probability distribution:

$$\begin{array}{rrrr}
P(\alpha_+\beta_+a_ib_i\lambda_1) = 0 & P(\alpha_+\beta_-a_ib_i\lambda_1) = 0 & P(\alpha_-\beta_+a_ib_i\lambda_1) = 0 & P(\alpha_-\beta_-a_ib_i\lambda_1) = \frac{1}{8} \\
P(\alpha_+\beta_+a_ib_{i_\perp}\lambda_1) = 0 & P(\alpha_+\beta_-a_ib_{i_\perp}\lambda_1) = 0 & P(\alpha_-\beta_+a_ib_{i_\perp}\lambda_1) = \frac{1}{8} & P(\alpha_-\beta_-a_ib_{i_\perp}\lambda_1) = 0 \\
P(\alpha_+\beta_+a_{i_\perp}b_i\lambda_1) = 0 & P(\alpha_+\beta_-a_{i_\perp}b_i\lambda_1) = \frac{1}{8} & P(\alpha_-\beta_+a_{i_\perp}b_i\lambda_1) = 0 & P(\alpha_-\beta_-a_{i_\perp}b_i\lambda_1) = 0 \\
P(\alpha_+\beta_+a_{i_\perp}b_{i_\perp}\lambda_1) = \frac{1}{8} & P(\alpha_+\beta_-a_{i_\perp}b_{i_\perp}\lambda_1) = 0 & P(\alpha_-\beta_+a_{i_\perp}b_{i_\perp}\lambda_1) = 0 & P(\alpha_-\beta_-a_{i_\perp}b_{i_\perp}\lambda_1) = 0 \\
P(\alpha_+\beta_+a_ib_i\lambda_2) = \frac{1}{8} & P(\alpha_+\beta_-a_ib_i\lambda_2) = 0 & P(\alpha_-\beta_+a_ib_i\lambda_2) = 0 & P(\alpha_-\beta_-a_ib_i\lambda_2) = 0 \\
P(\alpha_+\beta_+a_ib_{i_\perp}\lambda_2) = 0 & P(\alpha_+\beta_-a_ib_{i_\perp}\lambda_2) = \frac{1}{8} & P(\alpha_-\beta_+a_ib_{i_\perp}\lambda_2) = 0 & P(\alpha_-\beta_-a_ib_{i_\perp}\lambda_2) = 0 \\
P(\alpha_+\beta_+a_{i_\perp}b_i\lambda_2) = 0 & P(\alpha_+\beta_-a_{i_\perp}b_i\lambda_2) = 0 & P(\alpha_-\beta_+a_{i_\perp}b_i\lambda_2) = \frac{1}{8} & P(\alpha_-\beta_-a_{i_\perp}b_i\lambda_2) = 0 \\
P(\alpha_+\beta_+a_{i_\perp}b_{i_\perp}\lambda_2) = 0 & P(\alpha_+\beta_-a_{i_\perp}b_{i_\perp}\lambda_2) = 0 & P(\alpha_-\beta_+a_{i_\perp}b_{i_\perp}\lambda_2) = 0 & P(\alpha_-\beta_-a_{i_\perp}b_{i_\perp}\lambda_2) = \frac{1}{8}
\end{array}$$

This distribution is in accordance with the axioms of probability theory; by construction its hidden joint probability has the product form that is characteristic for class $(\mathrm{H}_{29}^{\alpha})$, and it reproduces the perfect correlations and anti-correlations. This explicit example shows that class $(\mathrm{H}_{29}^{\alpha})$ is consistent with the assumptions autonomy, perfect correlations and perfect (anti-)correlations.

In a similar way one can construct examples of probability distributions for the other classes fulfilling $(\neg\mathrm{i})$ and $(\neg\mathrm{ii})$. Since $(\mathrm{H}_{22}^{\alpha})$ is symmetric to $(\mathrm{H}_{29}^{\alpha})$ under interchanging the settings, it is clear that the constructed distribution for the latter class can easily be turned into an example for the former if in each total probability one swaps the values of the settings. Furthermore, it is straightforward to modify the construction such that it yields distributions for the classes $\{(\mathrm{H}_1^{\alpha}), \ldots, (\mathrm{H}_{14}^{\alpha})\} \backslash \{(\mathrm{H}_4^{\alpha}), (\mathrm{H}_5^{\alpha}), (\mathrm{H}_{10}^{\alpha})\}$. Note that in these classes there are more degrees of freedom than in the presented example, so one might freely choose more values of probabilities. This completes our proof of theorem 1.1.

q.e.d.

Proof of theorem 1.2

We split the theorem up into two partial claims:

Claim 1: Given autonomy, perfect correlations and perfect anti-correlations a consistent class (i.e. a class that fulfills $(\neg\mathrm{i})$ and $(\neg\mathrm{ii})$) implies Bell inequalities if (iii) each factor of its product form involves at most one setting.

Claim 2: Given autonomy, perfect correlations and perfect anti-correlations a consistent class (i.e. a class that fulfills $(\neg\mathrm{i})$ and $(\neg\mathrm{ii})$) does not imply Bell inequalities if $(\neg\mathrm{iii})$ at least one factor of its product form involves both settings.

Proof of claim 1

The set of classes fulfilling $(\neg\mathrm{i})$, $(\neg\mathrm{ii})$ and (iii) consists of (H_{22}) and (H_{29}). Here we have to show that, given autonomy, perfect correlations and perfect (anti-)correlations, each of these classes implies Bell inequalities.

By usual derivations of Wigner-Bell inequalities, it is well-known that local factorisation (H_{29}) implies Bell inequalities (given autonomy and perfect correlations; cf. premise (P4) of the Bell argument above). Now, it is easy to see that in a very similar way one can use $(\mathrm{H}_{22}^{\alpha})$ to derive Bell inequalities. For, as we have said, (H_{22}) differs from local factorisation only in that the settings in the product form are swapped: instead of a dependence of each outcome on the local settings each factor involves a dependence on the *distant* setting. Accordingly, the derivation from (H_{22}) results from the usual one by interchanging the settings in each expression.

Proof of claim 2

The classes fulfilling conditions (¬i) and (¬ii) while violating (iii) are $(\mathrm{H}_1^\alpha)..(\mathrm{H}_{14}^\alpha)\backslash\{(\mathrm{H}_4^\alpha),(\mathrm{H}_5^\alpha),(\mathrm{H}_{10}^\alpha)\}$. Here we have to show that given the background assumptions autonomy, perfect correlations and perfect anti-correlations, these classes do *not* imply the Bell inequalities. Since a class implying the inequalities means that all distributions of a class imply the inequalities, demonstrating that a class does not imply the inequalities amounts to showing that there is at least one distribution in that class that violates the inequalities. In other words, we have to show that there is at least one distribution for each class that fulfills autonomy, perfect correlations, perfect anti-correlations and violates the Bell inequalities.

One way to find such examples is to look at existing hidden-variable theories that successfully explain the statistics of EPR/B experiments. In our overview of the classes we have seen that the de-Broglie-Bohm theory falls under different classes depending on which temporal order the experiment has, (H_6^α), (H_9^α) or (H_{12}^α). For each of these classes, the probability distribution of the theory provides an example with the desired features. Moreover, the example for (H_9^α) can be turned into one for (H_8^α) by reversing the dependence on the settings. And similarly, the example for (H_{12}^α) can be turned into one for (H_{11}^α). Since (H_1^α), (H_2^α), (H_3^α) and (H_7^α) are stronger product forms (involve more dependences) than one or several of the classes (H_6^α), (H_8^α), (H_9^α), (H_{11}^α) or (H_{12}^α), by small modifications of the available examples one can construct examples for these classes as well.

It remains to find examples for classes (H_{13}^α) and (H_{14}^α). Since there are no theories available for these classes, here the construction has to be from scratch. Let me demonstrate how the construction works for class (H_{14}^α). We first of all take into account the perfect correlations and perfect anti-correlations. This goes, mutatis mutandis, very similar to finding a probability distribution from class (H_{29}^α) that is compatible with perfect (anti-)correlations (see proof of claim 3 in the proof of theorem 1.1). By similar equations to (69)–(72) (exchange the product form of (H_{29}^α) on the right hand side with the product form of (H_{14}^α)), for any i and λ there are two possible cases:

Case I:

$$P(\alpha_+|\lambda)=0 \qquad P(\alpha_-|\lambda)=1 \tag{94}$$

$$P(\beta_+|a_ib_i\lambda)=0 \qquad P(\beta_-|a_ib_i\lambda)=1 \qquad P(\beta_+|a_{i_\perp}b_{i_\perp}\lambda)=0 \qquad P(\beta_-|a_{i_\perp}b_{i_\perp}\lambda)=1 \tag{95}$$

$$P(\beta_+|a_ib_{i_\perp}\lambda)=1 \qquad P(\beta_-|a_ib_{i_\perp}\lambda)=0 \qquad P(\beta_+|a_{i_\perp}b_i\lambda)=1 \qquad P(\beta_-|a_{i_\perp}b_i\lambda)=0 \tag{96}$$

Case II: (replace all 0's in case I by 1 and vice versa)

Requiring just *any* example we can assume a toy model with only two possible hidden states ($\boldsymbol{\lambda}=1,2$). Then we might, for instance, choose case I for λ_1 and case II for λ_2 for all i's. Then, by equations similar to (73)–(76) it follows

$$P(\lambda_1)=\frac{1}{2} \qquad P(\lambda_2)=\frac{1}{2} \tag{97}$$

In this way we have accounted for the perfect correlations as well as for the perfect anti-correlations.

Now it remains to reproduce the EPR/B correlations for non-parallel and non-perpendicular settings. A minimal set of such probabilities, which can violate the Bell inequalities (both the usual ones as well as the Wigner-Bell inequalities), can be found if each of the settings $\boldsymbol{a}$ and $\boldsymbol{b}$ has two possible values, e.g. $a_1 = 0°$, $a_2 = 30°$, $b_1 = 30°$ and $b_2 = 60°$. Measuring the quantum state $\psi_0 = (|++\rangle + |--\rangle)/\sqrt{2}$ at these settings yields the following observable probabilities:

$$P(\alpha_\pm\beta_\pm|a_1b_1) = \tfrac{3}{8} \quad P(\alpha_\pm\beta_\mp|a_1b_1) = \tfrac{1}{8} \quad P(\alpha_\pm\beta_\pm|a_1b_2) = \tfrac{1}{8} \quad P(\alpha_\pm\beta_\mp|a_1b_2) = \tfrac{3}{8} \tag{98}$$
$$P(\alpha_\pm\beta_\pm|a_2b_1) = \tfrac{1}{2} \quad P(\alpha_\pm\beta_\mp|a_2b_1) = 0 \quad P(\alpha_\pm\beta_\pm|a_2b_2) = \tfrac{3}{8} \quad P(\alpha_\pm\beta_\mp|a_2b_2) = \tfrac{1}{8} \tag{99}$$

These are sixteen equations, and any of the probabilities on their left hand sides can be expressed by the product form of the hidden joint probability:

$$P(\alpha\beta|ab) = \sum_\lambda P(\lambda)P(\alpha|\lambda)P(\beta|ab\lambda) \tag{100}$$

$P(\lambda)$ and $P(\alpha|\lambda)$ are already completely determined by the perfect (anti-)correlations, $P(\beta|ab\lambda)$ partly so (namely only for the parallel settings $a_2 = b_1$):

$$P(\lambda_1) = \tfrac{1}{2} \quad P(\lambda_2) = \tfrac{1}{2} \tag{101}$$
$$P(\alpha_+|\lambda_1) = 0 \quad P(\alpha_-|\lambda_1) = 1 \quad P(\alpha_+|\lambda_2) = 1 \quad P(\alpha_-|\lambda_2) = 0 \tag{102}$$
$$P(\beta_+|a_2b_1\lambda_1) = 0 \quad P(\beta_-|a_2b_1\lambda_1) = 1 \quad P(\beta_+|a_2b_1\lambda_2) = 1 \quad P(\beta_-|a_2b_1\lambda_2) = 0 \tag{103}$$

Inserting these values in the respective equations yields the following consistent values for the missing probabilities $P(\beta|ab\lambda)$:

$$P(\beta_+|a_1b_1\lambda_1) = \tfrac{1}{4} \quad P(\beta_-|a_1b_1\lambda_1) = \tfrac{3}{4} \quad P(\beta_+|a_1b_1\lambda_2) = \tfrac{3}{4} \quad P(\beta_-|a_1b_1\lambda_2) = \tfrac{1}{4} \tag{104}$$
$$P(\beta_+|a_1b_2\lambda_1) = \tfrac{3}{4} \quad P(\beta_-|a_1b_2\lambda_1) = \tfrac{1}{4} \quad P(\beta_+|a_1b_2\lambda_2) = \tfrac{1}{4} \quad P(\beta_-|a_1b_1\lambda_2) = \tfrac{3}{4} \tag{105}$$
$$P(\beta_+|a_2b_2\lambda_1) = \tfrac{1}{4} \quad P(\beta_-|a_2b_2\lambda_1) = \tfrac{3}{4} \quad P(\beta_+|a_2b_2\lambda_2) = \tfrac{3}{4} \quad P(\beta_-|a_2b_2\lambda_2) = \tfrac{1}{4} \tag{106}$$

Finally, choosing, say,

$$P(a_i) = \frac{1}{2} \quad P(a_{i_\perp}) = \frac{1}{2} \quad P(b_i) = \frac{1}{2} \quad P(b_{i_\perp}) = \frac{1}{2} \tag{107}$$

the formula

$$P(\alpha\beta ab\lambda) = P(\alpha|\lambda)P(\beta|ab\lambda)P(\lambda)P(a)P(b) \tag{108}$$

entails the following total probabilities, which constitute the searched for probability distribution:

$$\begin{array}{llll}
P(\alpha_+\beta_+a_1b_1\lambda_1)=0 & P(\alpha_+\beta_-a_1b_1\lambda_1)=0 & P(\alpha_-\beta_+a_1b_1\lambda_1)=\frac{1}{32} & P(\alpha_-\beta_-a_1b_1\lambda_1)=\frac{3}{32}\\
P(\alpha_+\beta_+a_1b_2\lambda_1)=0 & P(\alpha_+\beta_-a_1b_2\lambda_1)=0 & P(\alpha_-\beta_+a_1b_2\lambda_1)=\frac{3}{32} & P(\alpha_-\beta_-a_1b_2\lambda_1)=\frac{1}{32}\\
P(\alpha_+\beta_+a_2b_1\lambda_1)=0 & P(\alpha_+\beta_-a_2b_1\lambda_1)=0 & P(\alpha_-\beta_+a_2b_1\lambda_1)=0 & P(\alpha_-\beta_-a_2b_1\lambda_1)=\frac{1}{8}\\
P(\alpha_+\beta_+a_2b_2\lambda_1)=0 & P(\alpha_+\beta_-a_2b_2\lambda_1)=0 & P(\alpha_-\beta_+a_2b_2\lambda_1)=\frac{1}{32} & P(\alpha_-\beta_-a_2b_2\lambda_1)=\frac{3}{32}\\
P(\alpha_+\beta_+a_1b_1\lambda_2)=\frac{3}{32} & P(\alpha_+\beta_-a_1b_1\lambda_2)=\frac{1}{32} & P(\alpha_-\beta_+a_1b_1\lambda_2)=0 & P(\alpha_-\beta_-a_1b_1\lambda_2)=0\\
P(\alpha_+\beta_+a_1b_2\lambda_2)=\frac{1}{32} & P(\alpha_+\beta_-a_1b_2\lambda_2)=\frac{3}{32} & P(\alpha_-\beta_+a_1b_2\lambda_2)=0 & P(\alpha_-\beta_-a_1b_2\lambda_2)=0\\
P(\alpha_+\beta_+a_2b_1\lambda_2)=\frac{1}{8} & P(\alpha_+\beta_-a_2b_1\lambda_2)=0 & P(\alpha_-\beta_+a_2b_1\lambda_2)=0 & P(\alpha_-\beta_-a_2b_1\lambda_2)=0\\
P(\alpha_+\beta_+a_2b_2\lambda_2)=\frac{3}{32} & P(\alpha_+\beta_-a_2b_2\lambda_2)=\frac{1}{32} & P(\alpha_-\beta_+a_2b_2\lambda_2)=0 & P(\alpha_-\beta_-a_2b_2\lambda_2)=0
\end{array}$$

Note that here we have not explicitly noted the probabilities for parallel or perpendicular settings, but by constructing the distribution in the indicated way we have implicitly taken account of the perfect (anti-)correlations at these settings and it is straight forward to extent the distribution to include these settings as well (the distribution just becomes much longer, when for each measurement setting at one side one includes a parallel and a perpendicular setting at the other side).

This completes our construction of a distribution from class (H^α_{14}) which respects, autonomy, perfect correlations, perfect anti-correlations and violates the Bell inequalities. In a similar way, one can construct an example for class (H^α_{13}), which differs from (H^α_{14}) just in that the dependence on both settings is not in the second but in the first factor of its product form. q.e.d.

Proof of theorem 2.1

We split the theorem up into two partial claims:

Claim 1: Autonomy, nearly perfect correlations, nearly perfect anti-correlations, and a class of probability distributions (H^α_i) form an inconsistent set if (i) the product form of (H^α_i) involves at most one of the settings.

Claim 2: A class (H^α_i) is consistent with autonomy, nearly perfect correlations and nearly perfect anti-correlations if ($\neg$i) the product form of (H^α_i) involves both settings.

<u>Proof of claim 1</u>

Condition (i), that the product form involves at most one of the settings, is fulfilled by the classes $\{(\mathrm{H}^\alpha_{17}),\ldots,(\mathrm{H}^\alpha_{32})\}\backslash\{(\mathrm{H}^\alpha_{22}),(\mathrm{H}^\alpha_{29})\}$. Here we have to show the inconsistency of these classes with the set of assumptions autonomy, nearly perfect correlations and nearly perfect anti-correlations.

The proof runs very similar to our demonstration of claim 1 in the proof of theorem 1.1. On the one hand, the nearly perfect correlations and anti-correlations involve dependences on each of the settings, e.g. the nearly perfect correlations

$$P(\alpha_\pm\beta_\pm|a_ib_i) = \frac{1}{2} - \delta_{ii} \qquad\qquad P(\alpha_\pm\beta_\pm|a_ib_{i_\perp}) = \delta_{ii_\perp} \tag{109}$$

reveal a dependence on the setting $\boldsymbol{b}$, while e.g. the nearly perfect correlations

$$P(\alpha_\pm\beta_\pm|a_ib_i) = \frac{1}{2} - \delta_{ii} \qquad\qquad P(\alpha_\pm\beta_\pm|a_{i_\perp}b_i) = \delta_{i_\perp i} \tag{110}$$

show a dependence on the setting $\boldsymbol{a}$. On the other hand, any hidden joint probability that does not involve the setting $\boldsymbol{b}$ (i.e. is independent of $\boldsymbol{b}$), cannot account for changing values in the empirical joint probability with changing values of $\boldsymbol{b}$ (cf. (55)); so it necessarily contradicts the set of equations (109). And similarly, hidden joint probabilities that are independent of $\boldsymbol{a}$ contradict the set of equations (110).

Note that condition (ii) from theorem 1.1 is not a criterion for inconsistency according to theorem 2.1, because the inconsistency in question essentially relies on *strictly* perfect (anti-)correlations, which are not assumed in theorem 2.1.

Proof of claim 2

The classes fulfilling criterion (¬i) to involve both settings in their product forms are (H_1^α)...(H_{16}^α), (H_{22}^α) and (H_{29}^α). Here we have to show the consistency of these classes with the set of assumptions autonomy, nearly perfect correlations and nearly perfect anti-correlations.

As in the proof of claim 3 in the proof of theorem 1.1 one can demonstrate the present claim by providing an example of a probability distribution for each class that is consistent with these assumptions. Since nearly perfect correlations are a weaker requirement than strictly perfect ones, it is clear that for all classes which we have shown to be consistent with the latter—viz. $(\mathrm{H}_1^\alpha)..(\mathrm{H}_{14}^\alpha)\backslash\{(\mathrm{H}_4^\alpha),(\mathrm{H}_5^\alpha),(\mathrm{H}_{10}^\alpha)\}$—are also consistent with the former. Therefore, what still needs to be proven here is that autonomy and nearly perfect (anti-)correlations are consistent with those classes fulfilling criterion (¬i) that are inconsistent with the strictly perfect ones (because they fulfill (ii)). The classes in question are (H_4^α), (H_5^α), (H_{10}^α), (H_{15}^α) and (H_{16}^α).

Again, the best way to find such examples is by constructing them such that the conditions are fulfilled. Here we show how to construct a distribution for class (H_{10}^α). The starting point are the equations for nearly perfect (anti-)correlations:

$$\begin{array}{llll}
P(\alpha_\pm\beta_\mp|a_ib_i) = \delta_{ii} & P(\alpha_\pm\beta_\mp|a_{i_\perp}b_{i_\perp}) = \delta_{i_\perp i_\perp} & P(\alpha_\pm\beta_\pm|a_{i_\perp}b_i) = \delta_{i_\perp i} & P(\alpha_\pm\beta_\pm|a_ib_{i_\perp}) = \delta_{ii_\perp} \\
P(\alpha_\pm\beta_\pm|a_ib_i) = \frac{1}{2} - \delta_{ii} & P(\alpha_\pm\beta_\pm|a_{i_\perp}b_{i_\perp}) = \frac{1}{2} - \delta_{i_\perp i_\perp} & P(\alpha_\pm\beta_\mp|a_{i_\perp}b_i) = \frac{1}{2} - \delta_{i_\perp i} & P(\alpha_\pm\beta_\mp|a_ib_{i_\perp}) = \frac{1}{2} - \delta_{ii_\perp}
\end{array}$$

Replacing the empirical probability on the left hand side of each equation by an equivalent expression involving hidden probabilities of the product form,

$$P(\alpha\beta|ab) = \sum_\lambda P(\lambda)P(\alpha|\beta\lambda)P(\beta|ab\lambda), \tag{111}$$

yields a set of equations, whose solutions determine probability distributions with the required features.

The δ's in these equations indicate the deviation from strictly perfect correlations. One might use realistic empirical values for them but since the task here is a merely conceptual one, one might as well just stipulate any small, positive values. Due to the lacking perfectness, the resulting set of equations is more complicated than that in theorem 1.2, and solutions are best determined by appropriate computer algorithms. Here, we shall present a solution for the special case

$$\delta_{ii} = \delta_{i_\perp i_\perp} = \delta_{i_\perp i} = \delta_{ii_\perp} =: \delta, \tag{112}$$

which reads:

$$P(\lambda_1) = \tfrac{1}{2} \qquad P(\lambda_2) = \tfrac{1}{2} \tag{113}$$

$$P(\alpha_+|\beta_+\lambda_1) = 0 \qquad P(\alpha_-|\beta_+\lambda_1) = 1 \qquad P(\alpha_+|\beta_+\lambda_2) = 1-2\delta \qquad P(\alpha_-|\beta_+\lambda_2) = 2\delta \tag{114}$$

$$P(\alpha_+|\beta_-\lambda_1) = 2\delta \qquad P(\alpha_-|\beta_-\lambda_1) = 1-2\delta \qquad P(\alpha_+|\beta_-\lambda_2) = 1 \qquad P(\alpha_-|\beta_-\lambda_2) = 0 \tag{115}$$

$$P(\beta_+|a_ib_i\lambda_1) = 0 \qquad P(\beta_-|a_ib_i\lambda_1) = 1 \qquad P(\beta_+|a_ib_i\lambda_2) = 1 \qquad P(\beta_-|a_ib_i\lambda_2) = 0 \tag{116}$$

$$P(\beta_+|a_ib_{i_\perp}\lambda_1) = \tfrac{4\delta-1}{2\delta-1} \qquad P(\beta_-|a_ib_{i_\perp}\lambda_1) = \tfrac{2\delta}{1-2\delta} \qquad P(\beta_+|a_ib_{i_\perp}\lambda_2) = \tfrac{2\delta}{1-2\delta} \qquad P(\beta_-|a_ib_{i_\perp}\lambda_2) = \tfrac{4\delta-1}{2\delta-1} \tag{117}$$

$$P(\beta_+|a_{i_\perp}b_i\lambda_1) = \tfrac{4\delta-1}{2\delta-1} \qquad P(\beta_-|a_{i_\perp}b_i\lambda_1) = \tfrac{2\delta}{1-2\delta} \qquad P(\beta_+|a_{i_\perp}b_i\lambda_2) = \tfrac{2\delta}{1-2\delta} \qquad P(\beta_-|a_{i_\perp}b_{i_\perp}\lambda_2) = \tfrac{4\delta-1}{2\delta-1} \tag{118}$$

$$P(\beta_+|a_{i_\perp}b_{i_\perp}\lambda_1) = 0 \qquad P(\beta_-|a_{i_\perp}b_{i_\perp}\lambda_1) = 1 \qquad P(\beta_+|a_{i_\perp}b_{i_\perp}\lambda_2) = 1 \qquad P(\beta_-|a_{i_\perp}b_{i_\perp}\lambda_2) = 0 \tag{119}$$

Note that according to this solution all dependences of the product form $(\mathrm{H}_{10}^{\alpha})$ are preserved, because, for instance, we have

$$P(\alpha_+|\beta_+\lambda_1) \neq P(\alpha_+|\beta_-\lambda_1) \qquad P(\alpha_+|\beta_+\lambda_1) \neq P(\alpha_+|\beta_+\lambda_2) \tag{120}$$

$$P(\beta_+|a_ib_i\lambda_1) \neq P(\beta_+|a_{i_\perp}b_i\lambda_1) \qquad P(\beta_+|a_ib_i\lambda_1) \neq P(\beta_+|a_ib_{i_\perp}\lambda_1) \tag{121}$$

$$P(\beta_+|a_ib_i\lambda_1) \neq P(\beta_+|a_ib_i\lambda_2) \tag{122}$$

Finally, when we further assume, say,

$$P(a_i) = \frac{1}{2} \qquad P(a_{i_\perp}) = \frac{1}{2} \qquad P(b_i) = \frac{1}{2} \qquad P(b_{i_\perp}) = \frac{1}{2} \tag{123}$$

by the equation

$$P(\alpha\beta ab\lambda) = P(\alpha|\beta\lambda)P(\beta|ab\lambda)P(\lambda)P(a)P(b) \tag{124}$$

the results so far determine the values of the total probability distribution:

$$P(\alpha_+\beta_+a_ib_i\lambda_1) = 0 \qquad P(\alpha_-\beta_+a_ib_i\lambda_1) = 0 \tag{125}$$
$$P(\alpha_+\beta_+a_ib_{i_\perp}\lambda_1) = 0 \qquad P(\alpha_-\beta_+a_ib_{i_\perp}\lambda_1) = \tfrac{1-4\delta}{8(1-2\delta)} \tag{126}$$
$$P(\alpha_+\beta_+a_{i_\perp}b_i\lambda_1) = 0 \qquad P(\alpha_-\beta_+a_{i_\perp}b_i\lambda_1) = \tfrac{1-4\delta}{8(1-2\delta)} \tag{127}$$
$$P(\alpha_+\beta_+a_{i_\perp}b_{i_\perp}\lambda_1) = 0 \qquad P(\alpha_-\beta_+a_{i_\perp}b_{i_\perp}\lambda_1) = 0 \tag{128}$$

$$P(\alpha_+\beta_-a_ib_i\lambda_1) = \tfrac{\delta}{4} \qquad P(\alpha_-\beta_-a_ib_i\lambda_1) = \tfrac{1}{8}(1-2\delta) \tag{129}$$
$$P(\alpha_+\beta_-a_ib_{i_\perp}\lambda_1) = \tfrac{\delta^2}{2(1-2\delta)} \qquad P(\alpha_-\beta_-a_ib_{i_\perp}\lambda_1) = \tfrac{\delta}{4} \tag{130}$$
$$P(\alpha_+\beta_-a_{i_\perp}b_i\lambda_1) = \tfrac{\delta^2}{2(1-2\delta)} \qquad P(\alpha_-\beta_-a_{i_\perp}b_i\lambda_1) = \tfrac{\delta}{4} \tag{131}$$
$$P(\alpha_+\beta_-a_{i_\perp}b_{i_\perp}\lambda_1) = \tfrac{\delta}{4} \qquad P(\alpha_-\beta_-a_{i_\perp}b_{i_\perp}\lambda_1) = \tfrac{1}{8}(1-2\delta) \tag{132}$$

$$P(\alpha_+\beta_+a_ib_i\lambda_2) = \tfrac{1}{8}(1-2\delta) \qquad P(\alpha_-\beta_+a_ib_i\lambda_2) = \tfrac{\delta}{4} \tag{133}$$
$$P(\alpha_+\beta_+a_ib_{i_\perp}\lambda_2) = \tfrac{\delta}{4} \qquad P(\alpha_-\beta_+a_ib_{i_\perp}\lambda_2) = \tfrac{\delta^2}{2(1-2\delta)} \tag{134}$$
$$P(\alpha_+\beta_+a_{i_\perp}b_i\lambda_2) = \tfrac{\delta}{4} \qquad P(\alpha_-\beta_+a_{i_\perp}b_i\lambda_2) = \tfrac{\delta^2}{2(1-2\delta)} \tag{135}$$
$$P(\alpha_+\beta_+a_{i_\perp}b_{i_\perp}\lambda_2) = \tfrac{1}{8}(1-2\delta) \qquad P(\alpha_-\beta_+a_{i_\perp}b_{i_\perp}\lambda_2) = \tfrac{\delta}{4} \tag{136}$$

$$P(\alpha_+\beta_-a_ib_i\lambda_2) = 0 \qquad P(\alpha_-\beta_-a_ib_i\lambda_2) = 0 \tag{137}$$
$$P(\alpha_+\beta_-a_ib_{i_\perp}\lambda_2) = \tfrac{1-4\delta}{8(1-2\delta)} \qquad P(\alpha_-\beta_-a_ib_{i_\perp}\lambda_2) = 0 \tag{138}$$
$$P(\alpha_+\beta_-a_{i_\perp}b_i\lambda_2) = \tfrac{1-4\delta}{8(1-2\delta)} \qquad P(\alpha_-\beta_-a_{i_\perp}b_i\lambda_2) = 0 \tag{139}$$
$$P(\alpha_+\beta_-a_{i_\perp}b_{i_\perp}\lambda_2) = 0 \qquad P(\alpha_-\beta_-a_{i_\perp}b_{i_\perp}\lambda_2) = 0 \tag{140}$$

By construction this distribution has the product form that is characteristic for class (H^α_{10}), and it involves autonomy, nearly perfect correlations for parallel settings and nearly perfect anti-correlations for perpendicular settings. This explicitly shows class (H^α_{14}) to be consistent with these assumptions. In a similar way, one can find examples for classes (H^α_{4}), (H^α_{5}), (H^α_{15}) and (H^α_{16}) consistent with the mentioned assumptions.

q.e.d.

Proof of theorem 2.2

We split the theorem up into two partial claims:

Claim 1: Given autonomy, nearly perfect correlations and nearly perfect anti-correlations a consistent class (i.e. a class that fulfills (¬i)) implies Bell inequalities if (iii) each factor of its product form involves at most one setting.

Claim 2: Given autonomy, nearly perfect correlations and nearly perfect anti-correlations a consistent class (i.e. a class that fulfills (¬i)) does not imply Bell inequalities if (¬iii) at least one factor of its product form involves both settings.

Proof of claim 1

The set of classes fulfilling (¬i) and (iii) consists of (H_{15}), (H_{16}), (H_{22}) and (H_{29}). It has to be shown that given autonomy, nearly perfect correlations and nearly perfect anti-correlations, each of these class implies Bell inequalities.

In lemma 2 we have already demonstrated that under these conditions (H^{α}_{16}) implies Bell inequalities. Since (H^{α}_{15}) only differs from (H^{α}_{16}) in that the settings are swapped in the product form, mutatis mutandis also (H^{α}_{15}) implies the inequalities. Finally, since local factorization (H^{α}_{29}) is a weaker product form than (H^{α}_{16}), and since (H^{α}_{22}) is a weaker form than (H^{α}_{15}), it is clear that also these other two product forms imply the Bell inequalities in the given circumstances.

Note that though it might seem obvious that local factorization implies the inequalities, it is a non-trivial claim that it does imply the Wigner-Bell inequalities with only *nearly* perfect (anti-)correlations, because usual derivations so far did have to assume strictly perfect correlations; however, our derivation with (H^{α}_{16}) can easily be adopted to derive the inequalities from local factorization and just the nearly perfect (anti-)correlations. q.e.d.

Proof of claim 2

The classes fulfilling condition (¬i) while violating (iii) are (H^{α}_{1})...(H^{α}_{14}). Here we have to show that given the background assumptions autonomy, nearly perfect correlations and nearly perfect anti-correlations, these classes do *not* imply the Bell inequalities. This amounts to showing that there is at least one distribution for each class that fulfills the background assumptions and violates the Bell inequalities.

We know already from theorem 1.2 that the classes $(\mathrm{H}^{\alpha}_{1})..(\mathrm{H}^{\alpha}_{14})\backslash\{(\mathrm{H}^{\alpha}_{4}),(\mathrm{H}^{\alpha}_{5}),(\mathrm{H}^{\alpha}_{10})\}$ can violate the inequalities given the assumptions of autonomy and *strictly* perfect correlations. Since the latter are a stronger condition than *nearly* perfect correlations, it is clear that these classes can violate the Bell inequalities also in the present case. It remains to show that the classes $(\mathrm{H}^{\alpha}_{4}),(\mathrm{H}^{\alpha}_{5}),(\mathrm{H}^{\alpha}_{10})$ can violate the inequalities under the given assumptions. Here we explicitly construct an example for class (H^{α}_{10}).

In the proof of claim 2 of theorem 2.1 we have constructed a toy example of a probability distribution for this class that is compatible with autonomy and nearly perfect (anti-)correlations. When, for any setting i, we use the resulting probabilities (113)–(119) we can be sure that the distribution we are about to construct is consistent with the nearly perfect (anti-)correlations. What remains to be done is to reproduce the EPR/B correlations for non-parallel and non-perpendicular settings. We again choose the settings $a_1 = 0°$, $a_2 = 30°$, $b_1 = 30°$ and $b_2 = 60°$ as well as the quantum state $\psi = (|++\rangle + |--\rangle)/\sqrt{2}$. Then the observable probabilities read:

$$P(\alpha_\pm\beta_\pm|a_1b_1) = \tfrac{3}{8} \quad P(\alpha_\pm\beta_\mp|a_1b_1) = \tfrac{1}{8} \quad P(\alpha_\pm\beta_\pm|a_1b_2) = \tfrac{1}{8} \quad P(\alpha_\pm\beta_\mp|a_1b_2) = \tfrac{3}{8} \tag{141}$$
$$P(\alpha_\pm\beta_\pm|a_2b_1) = \tfrac{1}{2} - \delta \quad P(\alpha_\pm\beta_\mp|a_2b_1) = \delta \quad P(\alpha_\pm\beta_\pm|a_2b_2) = \tfrac{3}{8} \quad P(\alpha_\pm\beta_\mp|a_2b_2) = \tfrac{1}{8} \tag{142}$$

(Note the difference to the probabilities with the same settings and quantum state in (98)–(99), which involve strictly perfect anti-correlations for parallel settings $a_2 = b_1$

($P(\alpha_\pm\beta_\pm|a_2b_1) = \frac{1}{2}$ and $P(\alpha_\pm\beta_\mp|a_2b_1) = 0$) instead of nearly perfect ones ($P(\alpha_\pm\beta_\pm|a_2b_1) = \frac{1}{2} - \delta$ and $P(\alpha_\pm\beta_\mp|a_2b_1) = \delta$).
These are sixteen equations, and any of the probabilities on their left hand sides can be expressed by the product form of the hidden joint probability:

$$P(\alpha\beta|ab) = \sum_\lambda P(\lambda)P(\alpha|\beta\lambda)P(\beta|ab\lambda) \tag{143}$$

$P(\lambda)$ and $P(\alpha|\beta\lambda)$ are completely determined by the requirements of the perfect (anti-)correlations (113)–(115), $P(\beta|ab\lambda)$ partly so (namely only for the parallel settings, (116)).

Inserting these predetermined probabilities into equations (141)–(142) yields the following consistent values for the missing probabilities $P(\beta|ab\lambda)$:

$$P(\beta_+|a_1b_1\lambda_1) = \tfrac{1-8\delta}{4(1-2\delta)} \quad P(\beta_-|a_1b_1\lambda_1) = \tfrac{3}{4(1-2\delta)} \quad P(\beta_+|a_1b_1\lambda_2) = \tfrac{3}{4(1-2\delta)} \quad P(\beta_-|a_1b_1\lambda_2) = \tfrac{1-8\delta}{4(1-2\delta)} \tag{144}$$

$$P(\beta_+|a_1b_2\lambda_1) = \tfrac{3-8\delta}{4(1-2\delta)} \quad P(\beta_-|a_1b_2\lambda_1) = \tfrac{1}{4(1-2\delta)} \quad P(\beta_+|a_1b_2\lambda_2) = \tfrac{1}{4(1-2\delta)} \quad P(\beta_-|a_1b_1\lambda_2) = \tfrac{3-8\delta}{4(1-2\delta)} \tag{145}$$

$$P(\beta_+|a_2b_2\lambda_1) = \tfrac{1-8\delta}{4(1-2\delta)} \quad P(\beta_-|a_2b_2\lambda_1) = \tfrac{3}{4(1-2\delta)} \quad P(\beta_+|a_2b_2\lambda_2) = \tfrac{3}{4(1-2\delta)} \quad P(\beta_-|a_2b_2\lambda_2) = \tfrac{1-8\delta}{4(1-2\delta)} \tag{146}$$

Finally, choosing, say,

$$P(a_i) = \frac{1}{2} \qquad P(a_{i_\perp}) = \frac{1}{2} \qquad P(b_i) = \frac{1}{2} \qquad P(b_{i_\perp}) = \frac{1}{2} \tag{147}$$

the formula

$$P(\alpha\beta ab\lambda) = P(\alpha|\lambda)P(\beta|ab\lambda)P(\lambda)P(a)P(b) \tag{148}$$

entails the following total probabilities:

$P(\alpha_+\beta_+a_1b_1\lambda_1) = 0$	$P(\alpha_+\beta_-a_1b_1\lambda_1) = \frac{3\delta}{16(1-2\delta)}$	$P(\alpha_-\beta_+a_1b_1\lambda_1) = \frac{1-8\delta}{32(1-2\delta)}$	$P(\alpha_-\beta_-a_1b_1\lambda_1) = \frac{3}{32}$
$P(\alpha_+\beta_+a_1b_2\lambda_1) = 0$	$P(\alpha_+\beta_-a_1b_2\lambda_1) = \frac{\delta}{16(1-2\delta)}$	$P(\alpha_-\beta_+a_1b_2\lambda_1) = \frac{3-8\delta}{32(1-2\delta)}$	$P(\alpha_-\beta_-a_1b_2\lambda_1) = \frac{1}{32}$
$P(\alpha_+\beta_+a_2b_1\lambda_1) = 0$	$P(\alpha_+\beta_-a_2b_1\lambda_1) = \frac{\delta}{4}$	$P(\alpha_-\beta_+a_2b_1\lambda_1) = 0$	$P(\alpha_-\beta_-a_2b_1\lambda_1) = \frac{1-2\delta}{8}$
$P(\alpha_+\beta_+a_2b_2\lambda_1) = 0$	$P(\alpha_+\beta_-a_2b_2\lambda_1) = \frac{3\delta}{16(1-2\delta)}$	$P(\alpha_-\beta_+a_2b_2\lambda_1) = \frac{1-8\delta}{32(1-2\delta)}$	$P(\alpha_-\beta_-a_2b_2\lambda_1) = \frac{3}{32}$
$P(\alpha_+\beta_+a_1b_1\lambda_2) = \frac{3}{32}$	$P(\alpha_+\beta_-a_1b_1\lambda_2) = \frac{1-8\delta}{32(1-2\delta)}$	$P(\alpha_-\beta_+a_1b_1\lambda_2) = \frac{3\delta}{16(1-2\delta)}$	$P(\alpha_-\beta_-a_1b_1\lambda_2) = 0$
$P(\alpha_+\beta_+a_1b_2\lambda_2) = \frac{1}{32}$	$P(\alpha_+\beta_-a_1b_2\lambda_2) = \frac{3-8\delta}{32(1-2\delta)}$	$P(\alpha_-\beta_+a_1b_2\lambda_2) = \frac{\delta}{16(1-2\delta)}$	$P(\alpha_-\beta_-a_1b_2\lambda_2) = 0$
$P(\alpha_+\beta_+a_2b_1\lambda_2) = \frac{1-2\delta}{8}$	$P(\alpha_+\beta_-a_2b_1\lambda_2) = 0$	$P(\alpha_-\beta_+a_2b_1\lambda_2) = \frac{\delta}{4}$	$P(\alpha_-\beta_-a_2b_1\lambda_2) = 0$
$P(\alpha_+\beta_+a_2b_2\lambda_2) = \frac{3}{32}$	$P(\alpha_+\beta_-a_2b_2\lambda_2) = \frac{1-8\delta}{32(1-2\delta)}$	$P(\alpha_-\beta_+a_2b_2\lambda_2) = \frac{3\delta}{16(1-2\delta)}$	$P(\alpha_-\beta_-a_2b_2\lambda_2) = 0$

This completes our construction of a distribution from class (H_{10}^{α}) which respects, autonomy, nearly perfect correlations, nearly perfect anti-correlations and violates the Bell inequalities. Similarly, one can construct examples of distributions for class (H_4^{α}) and (H_5^{α}). q.e.d.