A new framework for old Bell inequalities

Helv Phys Acta
Vol. 66 (1993)

0018-0238/93/080858-17$1.50+0.20/0

A new framework for old Bell inequalities

By B.S. Tsirelson

School of Mathematics, Tel Aviv University,
Tel Aviv 69978, Israel, e-mail: tsirel@math.tau.ac.il

Abstract. Three topics are shown to be closely connected, one belonging to the foundation of quantum theory (Bell-type inequalities), the second to statistical physics (inequalities for partition functions), and the third to probability theory (inequalities for one-dependent processes and two-block factors). To this end, Bell-type inequalities are reformulated for a new space-time arrangement.

1 Introduction

The traditional framework [1] for famous Bell inequalities is a correlation experiment on two noninteracting subsystems of a composite physical system. A correlation function

$$\langle A_k B_l \rangle$$

depends on two local parameters k, l; the first parameter k determines a local observable A_k on the first subsystem, while l determines B_l on the second subsystem. Supposing the observables to be two-valued ($A_k = \pm 1$, $B_l = \pm 1$), we have, for example,

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \le R \,, \tag{1.1}$$

where $R = 2$ for classical systems [2], when all four observables commute, while $R = 2\sqrt{2}$ for quantum systems [3], when $A_k B_l = B_l A_k$, but in general $A_1 A_2 \neq A_2 A_1$, $B_1 B_2 \neq B_2 B_1$.

This traditional framework is depicted by Fig. 1(a). The outcome A_k results from the

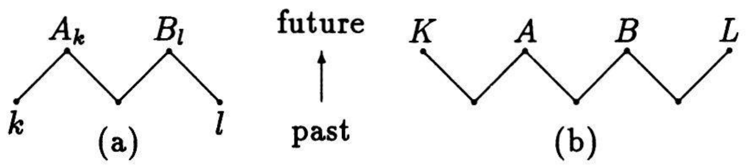

Figure 1: Graphs for the two frameworks: old (a) and new (b).

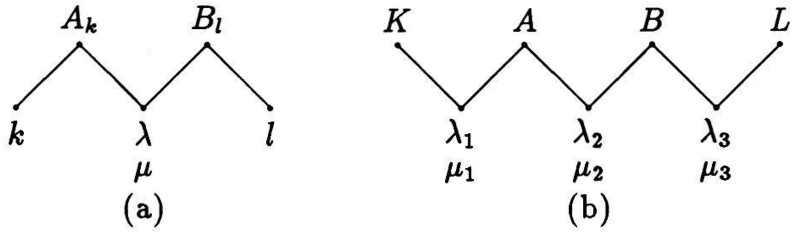

Figure 2: The two frameworks for classical systems.

classical input k and the first subsystem; B_l — from l and the second subsystem. The correlation function is defined as $c_{kl} = \langle A_k B_l \rangle$. The new framework presented here is depicted by Fig. 1(b). Four outcomes K, A, B, L are considered; their joint probability distribution determines a conditional expectation $c_{kl} = \mathbf{E}(AB|k,l)$ playing the role of the correlation function.

The proposed framework enables us to find the unexpected connections of Bell inequalities to some topics of statistical physics and probability theory, and to shed additional light on the problem of "free will" in choosing k, l.

In order to explain the difference between the two frameworks, the well-known proof of the classical CHSH inequality (Eq. (1.1) with $R = 2$) within the traditional framework will be repeated in brief, and a similar inequality will then be proved within the new framework.

2 The classical inequality within the old framework

A two-valued observable A_k is supposed to be a function of k and of a classical state λ of the whole system (a set of classical variables, possibly hidden); and similarly for B_l:

$$A_k(\lambda) = \pm 1 \,, \quad B_l(\lambda) = \pm 1 \,.$$

See Fig. 2(a). A statistical distribution of outcomes is supposed to result from a classical probability distribution μ for λ:

$$\langle A_k B_l \rangle = \int A_k(\lambda) B_l(\lambda) \, \mu(d\lambda) \,. \tag{2.1}$$

But for each λ we have

$$A_1(\lambda)B_1(\lambda) + A_1(\lambda)B_2(\lambda) + A_2(\lambda)B_1(\lambda) - A_2(\lambda)B_2(\lambda) \le 2 \tag{2.2}$$

(a finite number of possible combinations of ± 1's can be exhausted "by hand"). By integrating Eq. (2.2) we obtain the well-known CHSH inequality [2]

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \le 2 \, . \tag{2.3}$$

3 The classical inequality within the new framework

Three local classical states $\lambda_1, \lambda_2, \lambda_3$ are supposed to exist, and four observables K, A, B, L are supposed to be some functions of $\lambda_1, \lambda_2, \lambda_3$, their dependence being restricted by the considered graph, see Fig. 2(b):

$$\begin{aligned} A(\lambda_1, \lambda_2) &= \pm 1 \, ; \qquad & K(\lambda_1) &= 1 \text{ or } 2 \, ; \\ B(\lambda_2, \lambda_3) &= \pm 1 \, ; \qquad & L(\lambda_3) &= 1 \text{ or } 2 \, . \end{aligned}$$

A statistical distribution of outcomes is supposed to result from some classical probability distributions μ_1, μ_2, μ_3 for $\lambda_1, \lambda_2, \lambda_3$, respectively, with $\lambda_1, \lambda_2, \lambda_3$ assumed statistically independent. The correlation function is defined as

$$c_{kl} = \mathbf{E}(AB | K = k, L = l) = \frac{\int\limits_{\Lambda_1(k)} \mu_1(d\lambda_1) \int\limits_{\Lambda_3(l)} \mu_3(d\lambda_3) \int \mu_2(d\lambda_2) A(\lambda_1, \lambda_2) B(\lambda_2, \lambda_3)}{\mu_1(\Lambda_1(k)) \cdot \mu_3(\Lambda_3(l))} \tag{3.1}$$

where

$$\Lambda_1(k) = \{\lambda_1 : K(\lambda_1) = k\} \, , \quad \Lambda_3(l) = \{\lambda_3 : L(\lambda_3) = l\} \, .$$

The following new version of the CHSH inequality will be proved:

$$c_{11} + c_{12} + c_{21} - c_{22} \le 2 \, . \tag{3.2}$$

To this end, define functions A_1, A_2, B_1, B_2 of λ_2 as follows:

$$A_k(\lambda_2) = \frac{\int\limits_{\Lambda_1(k)} \mu_1(d\lambda_1) A(\lambda_1, \lambda_2)}{\mu_1(\Lambda_1(k))} \, , \tag{3.3}$$

and similarly for B_l. Clearly

$$c_{kl} = \int A_k(\lambda_2) B_l(\lambda_2) \, \mu_2(d\lambda_2) \, , \tag{3.4}$$

which leads to Eq. (3.2), just as Eq. (2.1) leads to Eq. (2.3). Although A_k and B_l are now no longer two-valued, but only satisfy

$$|A_k(\lambda_2)| \le 1 \, , \quad |B_l(\lambda_2)| \le 1 \, ,$$

it is well-known (and easy to see) that this is not an obstacle. A further proof, dispensing with such an enlargement of spectra of A_k, B_l, will be given in Sect. 8.

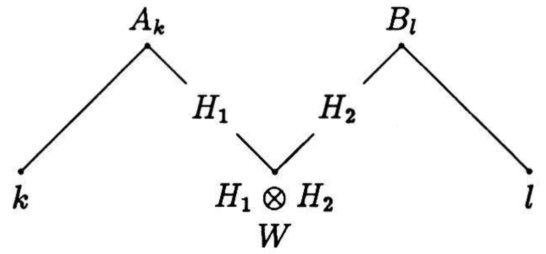

Figure 3: The old framework for quantum systems.

4 The quantum inequality within the old framework

A two-valued observable A_k is now supposed to be an operator on a Hilbert space H_1 describing the first subsystem of a composite quantum system; and similarly for B_l, H_2:

$$\begin{aligned} A_k : H_1 \to H_1 \ , \qquad & A_k^2 = 1 \ ; \\ B_l : H_2 \to H_2 \ , \qquad & B_l^2 = 1 \ . \end{aligned}$$

See Fig. 3. A statistical distribution of outcomes is supposed to be determined by a density matrix W on the space $H = H_1 \otimes H_2$ describing the whole system:

$$\langle A_k B_l \rangle = \mathrm{Tr}((A_k \otimes B_l)W) \ . \tag{4.1}$$

Being arbitrary, W may correspond, in particular, to an entangled state vector.

The following operator inequality is well-known [3]:

$$\| A_1 \otimes B_1 + A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2 \| \le 2\sqrt{2} \ . \tag{4.2}$$

Multiplying by W and taking the trace, we obtain [3]

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \le 2\sqrt{2} \ . \tag{4.3}$$

5 The quantum inequality within the new framework

Three non-interacting and non-correlated quantum systems are supposed to exist at the beginning. Each system decays into two subsystems (generally, correlated), and some pairs of subsystems then merge according to the considered graph, see Fig. 4. Four observables K, A, B, L are supposed to be operators on corresponding spaces:

$$\begin{aligned} K : H_1 \to H_1 \ ; \qquad & A : H_2 \otimes H_3 \to H_2 \otimes H_3 \ ; \\ L : H_6 \to H_6 \ ; \qquad & B : H_4 \otimes H_5 \to H_4 \otimes H_5 \ . \end{aligned}$$

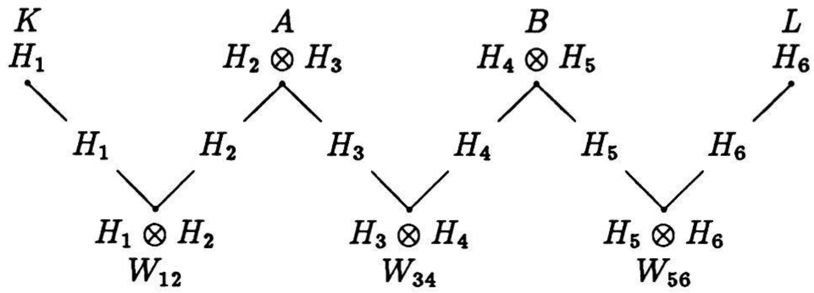

Figure 4: The new framework for quantum systems.

They are supposed two-valued:

$$\mathrm{Spec}(A) \subset \{-1,+1\}\ ; \qquad \mathrm{Spec}(K) \subset \{1,2\}\ ;$$
$$\mathrm{Spec}(B) \subset \{-1,+1\}\ ; \qquad \mathrm{Spec}(L) \subset \{1,2\}\ .$$

The initial state is described by three density matrices: W_{12} on $H_1 \otimes H_2$, W_{34} on $H_3 \otimes H_4$, and W_{56} on $H_5 \otimes H_6$. A statistical distribution of outcomes is determined by the standard formalism in the space

$$H = H_1 \otimes H_2 \otimes H_3 \otimes H_4 \otimes H_5 \otimes H_6\ .$$

That is,

$$c_{kl} = \mathbf{E}(AB|K=k, L=l) = \frac{\mathrm{Tr}((1_k(K) \otimes A \otimes B \otimes 1_l(L))W)}{\mathrm{Tr}((1_k(K) \otimes 1_l(L))W)}\ ; \tag{5.1}$$

here $1_k(K)$ is the projection operator onto the eigenspace of K, corresponding to its eigenvalue k, and $W = W_{12} \otimes W_{34} \otimes W_{56}$.

Clearly, Eq. (5.1) is a quantum counterpart of Eq. (3.1). The following inequality (the quantum counterpart of Eq. (3.2)) will be proved:

$$c_{11} + c_{12} + c_{21} - c_{22} \le 2\sqrt{2}\ . \tag{5.2}$$

To this end, introduce a quantum counterpart of Eq. (3.3) as follows. Define a Hermitian operator A_k on H_3 by the condition

$$\mathrm{Tr}(A_k W_3) = \frac{\mathrm{Tr}((1_k(K) \otimes A)(W_{12} \otimes W_3))}{\mathrm{Tr}(1_k(K)W_{12})} \tag{5.3}$$

for any density matrix W_3 on H_3. Of course, we suppose that $\mathrm{Tr}(1_k(K)W_{12}) > 0$. The existence of such A_k follows from standard arguments; in particular, for a finite dimension it follows immediately from the obvious fact that the right-hand side of Eq. (5.3) is a real-valued linear functional of W_3. Being the conditional expectation of A, this functional lies within $[-1,+1]$ for any W_3; hence, $\|A_k\| \le 1$. Similarly, this holds for the operators B_l on H_4.

Now we need the equality

$$c_{kl} = \mathrm{Tr}((A_k \otimes B_l)W_{34}) \,. \tag{5.4}$$

It is evident when W_{34} is a product $W_3 \otimes W_4$; but how do we generalize Eq. (5.4) for an arbitrary W_{34}? Due to linearity![1] Both sides of Eq. (5.4) may be considered linear functionals of W_{34}. Hence, their difference may be written as $\mathrm{Tr}(CW_{34})$ with some Hermitian C. We have $\mathrm{Tr}(C(W_3 \otimes W_4)) = 0$ for any W_3, W_4. Taking them to be one-dimensional, we obtain $\langle \psi_3 \otimes \psi_4 | C | \psi_3 \otimes \psi_4 \rangle = 0$ for any vectors $\psi_3 \in H_3, \psi_4 \in H_4$; this is possible only for $C = 0$.

Obtained Eq. (5.4) leads to Eq. (5.2), just as Eq. (4.1) leads to Eq. (4.3), since Eq. (4.2) remains true when the condition $A_k^2 = 1$, $B_l^2 = 1$ is replaced with $\|A_k\| \le 1$, $\|B_l\| \le 1$. Another proof, dispensing with the enlargement of spectra of A_k, B_l, will be given in Sect. 8.

6 Violating the classical inequality

It is well-known [2] that the classical CHSH inequality (2.3) can be violated by means of a pair of spin-1/2 particles in the singlet state, and the quantum inequality (4.3) can be turned into an equality. It will be shown here that the same holds within the new framework: the quantum inequality (5.2) can be turned into an equality, thus violating the classical inequality (3.2).

The famous spin singlet state

$$\psi = \frac{1}{\sqrt{2}} |\downarrow\uparrow\rangle - \frac{1}{\sqrt{2}} |\uparrow\downarrow\rangle \tag{6.1}$$

has the following correlation property:

$$\langle \sigma_\alpha \otimes \sigma_\beta \rangle_\psi = -\cos(\alpha - \beta) \,,$$

where $\sigma_\alpha = \sigma_x \cos\alpha + \sigma_y \sin\alpha$. Taking $A_1 = \sigma_0$, $A_2 = \sigma_{\pi/2}$, $B_1 = -\sigma_{\pi/4}$, $B_2 = -\sigma_{-\pi/4}$, we obtain

$$\langle A_1 B_1 \rangle_\psi + \langle A_1 B_2 \rangle_\psi + \langle A_2 B_1 \rangle_\psi - \langle A_2 B_2 \rangle_\psi = 2\sqrt{2} \,. \tag{6.2}$$

Now consider an experiment of the kind shown in Fig. 4. Take $W_{12} = W_{34} = W_{56} = |\psi\rangle\langle\psi|$ with ψ from Eq. (6.1), and let

$$K = \frac{3+\sigma}{2} \,, \qquad A = \frac{1+\sigma}{2} \otimes A_1 + \frac{1-\sigma}{2} \otimes A_2 \,,$$

$$L = \frac{3+\sigma}{2} \,, \qquad B = B_1 \otimes \frac{1+\sigma}{2} + B_2 \otimes \frac{1-\sigma}{2} \,;$$

[1] Of course, we cannot approximate an arbitrary W_{34} by means of linear combinations of products $W_3 \otimes W_4$ with positive coefficients. But here negative coefficients are acceptable, too. This is why the approximation is possible.

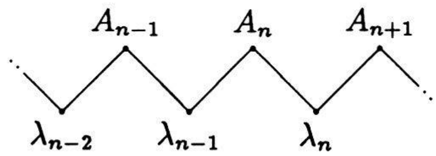

Figure 5: The framework investigated in probability theory.

any σ_α may be chosen for σ, for example, $\sigma = \sigma_0 = \sigma_x$. So, first and third singlets (W_{12} and W_{56}) are used simply as classical signals, while the second singlet (W_{34}) implements quantum correlations. The first pair (W_{12}) may be thought of as being either ψ_1 or ψ_2, where $\langle \sigma \otimes 1 \rangle_{\psi_1} = -1$, $\langle 1 \otimes \sigma \rangle_{\psi_1} = +1$, $\langle \sigma \otimes 1 \rangle_{\psi_2} = +1$, $\langle 1 \otimes \sigma \rangle_{\psi_2} = -1$; indeed, only these commuting operators are used on $H_1 \otimes H_2$ to form K, A. In the case of ψ_1 we have $K = (3-1)/2 = 1$, and A amounts to $\frac{1+1}{2} \otimes A_1 + \frac{1-1}{2} \otimes A_2 = A_1$; in the case of ψ_2 we obtain $K = 2$ and $A = A_2$. The same holds for L and B. Clearly, the correlation function coincides with that used in Eq. (6.2); in particular,

$$c_{11} + c_{12} + c_{21} - c_{22} = 2\sqrt{2}\,.$$

7 A connection to probability theory

The inequality (3.2) constrains a joint probability distribution for a quadruple (K, A, B, L) of random variables, provided that the distribution emerges from some distribution of independent random variables $\lambda_1, \lambda_2, \lambda_3$. These $\lambda_1, \lambda_2, \lambda_3$ may be called "hidden" in contrast to the "observable" variables K, A, B, L. In accordance with the given graph (see Fig. 2(b)), each hidden variable influences two adjacent observable variables.

Interestingly, the problem of finding constraints, resulting from the existence of such hidden variables, is studied in probability theory [4, 5, 6], but its connection to Bell-type inequalities is recognised for the first time.

A stationary random sequence $\{A_n\}$ is called a *two-block factor*, if it can be represented in the form [4]

$$A_n = f(\lambda_{n-1}, \lambda_n) \tag{7.1}$$

via a sequence $\{\lambda_n\}$ of independent identically distributed random variables (see Fig. 5). A clear restriction to such $\{A_n\}$ is the fact that A_{n+1} is independent of A_{n-1}, and moreover, the sequence $\{\ldots, A_{n-2}, A_{n-1}\}$ and the sequence $\{A_{n+1}, A_{n+2}, \ldots\}$ are (statistically) independent for each n. This property is known as *one-dependence.* The main result of Ref. [4] is the existence of a non-evident constraint for all two-valued stationary two-block factors. In other words, there exists a two-valued stationary one-dependent sequence that is not a two-block factor.

We will see that a close result can be obtained by means of inequality (3.2). First, Eq. (7.1) may be generalized for the non-homogeneous (=non-stationary) case:

$$A_n = f_n(\lambda_{n-1}, \lambda_n) \; ; \tag{7.2}$$

that is, f_n may now depend on n, as well as distributions for λ_n and A_n. The domain of n may be finite, $n \in \{1, 2, \dots, N\}$, or infinite. Accept Eq. (7.2) as a definition of a two-block factor, while Eq. (7.1) — of a homogeneous (or stationary) two-block factor.

The conditions imposed on the quadruple (K, A, B, L) in Sect. 3 mean exactly that it is a two-block factor of length $N = 4$, with two values. Hence, Eq. (3.2) constrains any two-valued two-block factor of length 4. When a given two-block factor $\{A_n\}$ is not two-valued, and/or of length > 4, we may still apply Eq. (3.2) to any quadruple of the form

$$g_1(A_n),\; g_2(A_{n+1}),\; g_3(A_{n+2}),\; g_4(A_{n+3}) \;, \tag{7.3}$$

where g_1, g_2, g_3, g_4 are arbitrary two-valued functions. Indeed, the quadruple (7.3) is again a two-block factor.

Unfortunately, explicit inequalities are cumbersome. For a two-block factor (A_1, A_2, A_3, A_4) with two values 0 and 1 we have, for example,

$$\begin{aligned} c_{01} &= \mathbf{E}((2A_2 - 1)(2A_3 - 1) | A_1 = 0, A_4 = 1) = \frac{\langle (1 - A_1)(2A_2 - 1)(2A_3 - 1)A_4 \rangle}{\langle (1 - A_1)A_4 \rangle} \\ &= \frac{1}{1 - \langle A_1 \rangle} \cdot \frac{1}{\langle A_4 \rangle} \cdot (-4\langle A_1 A_2 A_3 A_4 \rangle + 2\langle A_1 A_2 \rangle \langle A_4 \rangle + 2\langle A_1 \rangle \langle A_3 A_4 \rangle - \langle A_1 \rangle \langle A_4 \rangle \\ &\qquad + 4\langle A_2 A_3 A_4 \rangle - 2\langle A_2 \rangle \langle A_4 \rangle - 2\langle A_3 A_4 \rangle + \langle A_4 \rangle) \; ; \end{aligned} \tag{7.4}$$

disconnected products are factorized due to one-dependence. Similarly c_{00}, c_{10}, and c_{11} have to be found, and substituted into inequality (3.2) or into the more general one:

$$|c_{00} + c_{01} + c_{10} + c_{11} - 2c_{kl}| \le 2 \,. \tag{7.5}$$

It would be difficult to find these inequalities without the mediation of Bell's inequality! It will be shown in Sect. 8 that the above inequalities are the best possible. They are true boundaries for the class of two-valued two-block factors within the including class of two-valued one-dependent processes described by means of 10 averaged connected products $\langle A_1 \rangle$, $\langle A_2 \rangle$, $\langle A_3 \rangle$, $\langle A_4 \rangle$, $\langle A_1 A_2 \rangle$, $\langle A_2 A_3 \rangle$, $\langle A_3 A_4 \rangle$, $\langle A_1 A_2 A_3 \rangle$, $\langle A_2 A_3 A_4 \rangle$, $\langle A_1 A_2 A_3 A_4 \rangle$.

The conditions imposed on the quadruple (K, A, B, L) in Sect. 5 may be generalized as follows. A random sequence will be called a *quantum two-block factor,* if its joint distribution can be represented as the joint distribution of a sequence $\{A_n\}$ of commuting observables in the situation shown in Fig. 6. That is, a sequence $\{H_n\}$ of Hilbert spaces has to be given, A_n acting on $H_{2n-1} \otimes H_{2n}$, and probabilities determined by the tensor product of some density matrices W_n acting on $H_{2n} \otimes H_{2n+1}$. Due to locality, we avoid dealing with infinite tensor products; it is enough to consider joint distributions for all finite regions.

Sect. 6 gives an example of a quantum two-block factor (K, A, B, L) which is not a classical two-block factor. (After introducing the term "quantum two-block factor" it is

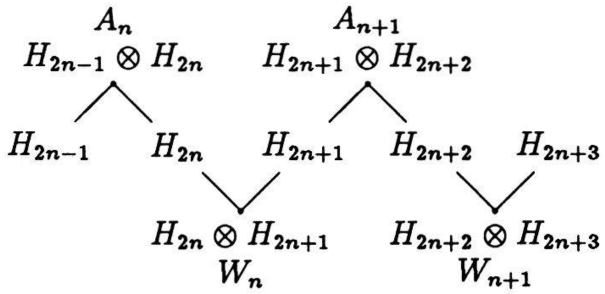

Figure 6: A quantum two-block factor.

X_{n-4}	X_{n-3}	X_{n-2}	X_{n-1}	X_n	X_{n+1}
K_{n-4}	K_{n-3}	K_{n-2}	K_{n-1}	K_n	K_{n+1}
A_{n-3}	A_{n-2}	A_{n-1}	A_n	A_{n+1}	A_{n+2}
B_{n-2}	B_{n-1}	B_n	B_{n+1}	B_{n+2}	B_{n+3}
L_{n-1}	L_n	L_{n+1}	L_{n+2}	L_{n+3}	L_{n+4}

Figure 7: A homogeneous example of a quantum two-block factor.

natural to add the word "classical" to the old term.) A homogeneous example can be obtained as follows. Take an infinite sequence of independent quadruples (K_n, A_n, B_n, L_n) for $n = \ldots -2, -1, 0, 1, 2 \ldots$, each quadruple being a copy of the above-mentioned (K, A, B, L), and define a multi-component random sequence $\{X_n\}$ as follows (see Fig. 7):

$$X_n = (K_n, A_{n+1}, B_{n+2}, L_{n+3}) \, .$$

Alternatively, the four two-valued components can be converted into one 16-valued component, for example,

$$X_n = 8(K_n - 1) + 4 \cdot \frac{1 + A_{n+1}}{2} + 2 \cdot \frac{1 + B_{n+2}}{2} + L_{n+3} \, . \tag{7.6}$$

It is easy to see that $\{X_n\}$ is indeed a quantum two-block factor, but not a classical two-block factor, since it violates the constraint Eqs. (3.2) and (7.3).

So, the presented theory provides us with a "quantum" proof of the following result from purely "classical" probability theory: there exists a one-dependent stationary random sequence which is not a (classical) two-block factor. Note that the above-mentioned result of Ref. [4] is stronger, since a two-valued example is constructed there. No "quantum" proof of it is known. It is also unknown, whether any homogeneous one-dependent two-valued random sequence is a quantum two-block factor, or not. For the 16-valued case a counterexample can be constructed by using Eqs. (5.2) and (7.6).

8 Are the two frameworks equivalent?

New inequalities (3.2) and (5.2) were derived from old inequalities (2.3) and (4.3). The question arises, whether all "new type" constraints (ensuing from the new frameworks) are derivable from "old type" constraints, or not.

An affirmative answer will be given, and not only for the two-valued case but also for the general case. The answer will be formulated in terms of the following two definitions. For simplicity, we restrict ourselves to the discrete case, supposing all observables to be integer-valued.

Define an *old-type classical probability set* as a family $\{p_{kl}^{ab}\}$ of numbers given for all integers k, l, a, b, admitting the following representation:

$$p_{kl}^{ab} = \mu\{\lambda : A_k(\lambda) = a, B_l(\lambda) = b\} \tag{8.1}$$

with some functions A_k, B_l on some set Λ carrying a probability measure μ.

Define a *new-type classical probability set* as a family $\{p_{kabl}\}$ of numbers given for all integers k, l, a, b, admitting the following representation:

$$p_{kabl} = (\mu_1 \otimes \mu_2 \otimes \mu_3)\{(\lambda_1, \lambda_2, \lambda_3) : K(\lambda_1) = k, A(\lambda_1, \lambda_2) = a, B(\lambda_2, \lambda_3) = b, L(\lambda_3) = l\} \tag{8.2}$$

with some sets $\Lambda_1, \Lambda_2, \Lambda_3$, carrying some probability measures μ_1, μ_2, μ_3, respectively, and some functions K on Λ_1, A on $\Lambda_1 \times \Lambda_2$, B on $\Lambda_2 \times \Lambda_3$, and L on Λ_3.

Suppose we are given both an old-type classical probability set $\{p_{kl}^{ab}\}$ and a new-type classical probability set $\{p_{kabl}\}$. We will call them corresponding (to each other), if the identity

$$p_{kabl} = p'_k p''_l p_{kl}^{ab} \tag{8.3}$$

holds for some sequences $\{p'_k\}$, $\{p''_l\}$, each summing to 1. It follows from Eq. (8.3) that

$$p'_k = \sum_{a,b,l} p_{kabl} \ , \quad p''_l = \sum_{k,a,b} p_{kabl} \ , \tag{8.4}$$

and hence $\{p_{kabl}\}$ determines $\{p_{kl}^{ab}\}$ uniquely provided that all p'_k, p''_l obtained from Eq. (8.4) are non-zero. On the contrary, $\{p_{kl}^{ab}\}$ contains no information about $\{p'_k\}$, $\{p''_l\}$, and does not determine $\{p_{kabl}\}$.

Theorem 1 *Let $\{p_{kabl}\}$ be a new-type classical probability set. Form $\{p'_k\}$, $\{p''_l\}$ according to Eq. (8.4) and suppose they all differ from zero. Then the numbers*

$$p_{kl}^{ab} = \frac{p_{kabl}}{p'_k p''_l}$$

form an old-type classical probability set.

Theorem 2 *Let $\{p^{ab}_{kl}\}$ be an old-type classical probability set, and $\{p'_k\}$, $\{p''_l\}$ be two sequences of non-negative numbers such that $\sum p'_k = 1$, $\sum p''_l = 1$. Then the numbers*

$$p_{kabl} = p'_k p''_l p^{ab}_{kl}$$

form a new-type classical probability set.

Proof of Theorem 2 is straightforward: take $\lambda_1 = k$, $\lambda_3 = l$, $\lambda_2 = \lambda$, $A(\lambda_1, \lambda_2) = A_k(\lambda)$, and $B(\lambda_2, \lambda_3) = B_l(\lambda)$.

Proof of Theorem 1. As in Sect. 3, introduce $\Lambda_1(k) = \{\lambda_1 : K(\lambda_1) = k\}$, then $\mu_1(\Lambda_1(k)) = p'_k$. Fix some nonatomic probability space (Ω_1, P_1), and for each k choose a map $\xi'_k : \Omega_1 \to \Lambda_1$ such that

$$P_1(\xi'_k \in \Delta) = \frac{1}{p'_k}\mu_1(\Delta \cap \Lambda_1(k))$$

for any $\Delta \subset \Lambda_1$. Similarly construct $\xi''_l : \Omega_3 \to \Lambda_3$. Finally, take the space $\Lambda = \Omega_1 \times \Lambda_2 \times \Omega_3$ with the product measure $\mu = P_1 \otimes \mu_2 \otimes P_3$, and for any $\lambda = (\omega_1, \lambda_2, \omega_3) \in \Lambda$ define

$$A_k(\lambda) = A(\xi'_k(\omega_1), \lambda_2)\,, \quad B_l(\lambda) = B(\lambda_2, \xi''_l(\omega_3))\,.$$

Then

$$\begin{aligned}
&\mu\{\lambda : A_k(\lambda) = a, B_l(\lambda) = b\} = \int \mu_2(d\lambda_2) P_1(A(\xi'_k, \lambda_2) = a) P_3(B(\lambda_2, \xi''_l) = b) \\
&= \int \mu_2(d\lambda_2) \frac{1}{p'_k}\mu_1\{\lambda_1 : K(\lambda_1) = k,\ A(\lambda_1, \lambda_2) = a\} \frac{1}{p''_l}\mu_3\{\lambda_3 : L(\lambda_3) = l,\ B(\lambda_2, \lambda_3) = b\} \\
&= \frac{1}{p'_k p''_l}(\mu_1 \otimes \mu_2 \otimes \mu_3)\{(\lambda_1, \lambda_2, \lambda_3) : K(\lambda_1) = k,\ A(\lambda_1, \lambda_2) = a,\ B(\lambda_2, \lambda_3) = b,\ L(\lambda_3) = l\} \\
&= \frac{p_{kabl}}{p'_k p''_l} = p^{ab}_{kl}\,,
\end{aligned}$$

q.e.d.

So, the two frameworks are equivalent for the classical theory. Returning to the two-valued case we conclude, that a necessary and sufficient condition is obtained for a quadruple (A_1, A_2, A_3, A_4) of two-valued random variables to be a (classical) two-block factor (defined by Eq. (7.2)). Indeed, the last means that their joint distribution $\{p_{kabl}\}$ is a new-type classical probability set. Due to Theorems 1,2 it is necessary and sufficient that corresponding $\{p^{ab}_{kl}\}$ is an old-type classical probability set. (The quadruple (A_1, A_2, A_3, A_4) is supposed one-dependent. The case, when at least one of p'_k, p''_l vanishes, is omitted because of its triviality.) Now inequalities (7.5) form a necessary and sufficient condition, as is well-known [7]. It follows that inequalities (7.4–7.5) are the best possible.

Note that the given proof of necessity provides us with one more proof of inequality (3.2), as was promised in Sect. 3.

A similar question for a sequence $(A_1, \ldots, A_n)$ with $n > 4$ remains open.

Are the two frameworks equivalent for quantum theory, too? Yes, they are. Define old-type and new-type quantum probability sets by replacing Eq. (8.1) with

$$p_{kl}^{ab} = \operatorname{Tr}((1_a(A_k) \otimes 1_b(B_l))W) ,$$

and Eq. (8.2) with

$$\begin{aligned} p_{kabl} &= \operatorname{Tr}((1_k(K) \otimes 1_a(A) \otimes 1_b(B) \otimes 1_l(L))W) , \\ W &= W_{12} \otimes W_{34} \otimes W_{56} ; \end{aligned}$$

cf. Eqs. (4.1) and (5.1).

Theorem 3 *The same as Theorem* 1, *but replacing "classical" with "quantum."*

Theorem 4 *The same as Theorem* 2, *but replacing "classical" with "quantum."*

Proof of Theorem 4 is a straightforward generalization of the construction used in Sect. 6.

Proof of Theorem 3. Having $\operatorname{Tr}(1_k(K)W_{12}) = p'_k > 0$, consider the state W'_k on H_2 obtained from W_{12} by postselection for $K = k$ on H_1; that is, W'_k is defined by the equality

$$\operatorname{Tr}(XW'_k) = \frac{1}{p'_k} \operatorname{Tr}((1_k(K) \otimes X)W_{12})$$

holding for any observable X on H_2. Similarly introduce W''_l on H_5. The conditional probabilities p_{kl}^{ab} may be computed via conditional states W'_k, W''_l:

$$p_{kl}^{ab} = \operatorname{Tr}((1_a(A) \otimes 1_b(B))(W'_k \otimes W_{34} \otimes W''_l)) , \tag{8.5}$$

the trace being taken on $H_2 \otimes H_3 \otimes H_4 \otimes H_5$. (Eq. (8.5) can be proved similarly to Eq. (5.4).) Take some new Hilbert spaces G_2, G_5 (these are quantum counterparts of Ω_1, Ω_3 used in the proof of Theorem 1), and represent each W'_k by a vector $\psi'_k \in G_2 \otimes H_2$:

$$\operatorname{Tr}(XW'_k) = \langle \psi'_k | 1 \otimes X | \psi'_k \rangle$$

for any observable X on H_2. Fix some unit vector $\psi' \in G_2 \otimes H_2$, and choose unitary operators U'_k on $G_2 \otimes H_2$ such that $U'_k \psi' = \psi'_k$. The same holds for ψ''_l, ψ'', U''_l. (These operators are quantum counterparts of ξ'_k, ξ''_l used in the proof of Theorem 1.) We have

$$p_{kl}^{ab} = \operatorname{Tr}(\,(1_a(A) \otimes 1_b(B))\,(|\psi'_k\rangle\langle\psi'_k| \otimes W_{34} \otimes |\psi''_l\rangle\langle\psi''_l|)\,);$$

here and henceforth A is transferred from $H_2 \otimes H_3$ to $G_2 \otimes H_2 \otimes H_3$ by means of tensor multiplication by identity on G_2. Substituting $|\psi'_k\rangle\langle\psi'_k| = U'_k|\psi'\rangle\langle\psi'|U'^*_k$, we obtain

$$\begin{aligned} p_{kl}^{ab} &= \operatorname{Tr}(\,(1_a(A) \otimes 1_b(B))\,(U'_k \otimes U''_l)\,(|\psi'\rangle\langle\psi'| \otimes W_{34} \otimes |\psi''\rangle\langle\psi''|)\,(U'_k \otimes U''_l)^*\,) \\ &= \operatorname{Tr}((1_a(A_k) \otimes 1_b(B_l))W) , \end{aligned}$$

where

$$A_k = U_k'^* A U_k' , \quad B_l = U_l''^* B U_l'' ,$$
$$W = |\psi'\rangle\langle\psi'| \otimes W_{34} \otimes |\psi''\rangle\langle\psi''| ,$$

q.e.d.

So, the two frameworks are also equivalent for quantum theory. Returning to the two-valued case, can we obtain a necessary and sufficient condition for a quadruple (A_1, A_2, A_3, A_4) of two-valued random variables to be a quantum two-block factor, as defined in Sect. 7? Yes, it can be done similarly to the classical case considered above, provided that a necessary and sufficient condition is available for an old-type quantum probability set. Such a condition was indeed obtained [8], though not in the form of explicit inequalities, but in a form free of operators in Hilbert spaces. Unfortunately, the condition is too cumbersome to be reproduced here.

The case of a sequence $(A_1, \ldots, A_n)$ with $n > 4$ has not yet been investigated.

9 A connection to statistical physics

Consider a system of classical statistical physics with a finite-range interaction. Divide it into a chain of subsystems such that the n-th subsystem interacts only with its two adjacent subsystems, the $(n-1)$-th and the $(n+1)$-th. The Hamilton function may be written as

$$H(\lambda_1, \ldots, \lambda_N) = \sum_{n=1}^{N} H_n(\lambda_n) + \sum_{n=1}^{N-1} H_{n,n+1}(\lambda_n, \lambda_{n+1}) ; \tag{9.1}$$

here λ_n denotes the state of the n-th subsystem, which may have any number of discrete and/or continuous components. Suppose that the decomposition of H is made so that

$$H_{n,n+1}(\lambda_n, \lambda_{n+1}) \geq 0$$

for all n, λ_n, and λ_{n+1}. The partition function is

$$\begin{aligned} Z &= \int \exp(-\beta H(\lambda_1, \ldots, \lambda_N))\, d\lambda_1 \ldots d\lambda_N \\ &= Z_1 \ldots Z_N \int \exp\left(-\beta \sum_{n=1}^{N-1} H_{n,n+1}(\lambda_n, \lambda_{n+1})\right) \mu_1(d\lambda_1) \ldots \mu_N(d\lambda_N) ; \end{aligned} \tag{9.2}$$

here $\beta = (kT)^{-1}$ is the inverse temperature, Z_n is the partition function for the n-th subsystem released from the interaction with its neighbors, and μ_n is the corresponding Gibbs measure:

$$\begin{aligned} Z_n &= \int \exp(-\beta H_n(\lambda_n))\, d\lambda_n , \\ \mu_n(d\lambda_n) &= Z_n^{-1} \exp(-\beta H_n(\lambda_n))\, d\lambda_n . \end{aligned} \tag{9.3}$$

A connection to two-block factors considered in Sect. 7 may be established as follows. Let us treat $\lambda_1, \dots, \lambda_N$ as independent random variables with distributions $\mu_1, \dots, \mu_N$, respectively. Introduce additional random variables $\theta_1, \dots, \theta_{N-1}$ distributed uniformly on $[0,1]$ and such that all $2N-1$ variables $(\lambda_1, \theta_1, \dots, \lambda_{N-1}, \theta_{N-1}, \lambda_N)$ are independent. Then random variables

$$A_n = \begin{cases} 1, & \text{when } \theta_n < \exp(-\beta H_{n,n+1}(\lambda_n, \lambda_{n+1})), \\ 0, & \text{otherwise}, \end{cases}$$

form a two-block factor, and their product averages to a ratio of partition functions:

$$\langle A_1 \dots A_{N-1} \rangle = \frac{Z}{Z_1 \dots Z_N} .$$

Moreover, each product of some A_n averages to some ratio of partition functions. To be more specific, we restrict ourselves to a homogeneous one-dimensional lattice system with pair interaction:

$$\begin{aligned} &Z = Z^K(1) \int \exp\left(-\beta \sum_{k=1}^{K-1} h_2(x_k, x_{k+1})\right) \mu(dx_1) \dots \mu(dx_K) , \\ &Z(1) = \int \exp(-\beta h_1(x))\, dx , \\ &\mu(dx) = (1/Z(1)) \exp(-\beta h_1(x))\, dx , \\ &h_2(x_k, x_{k+1}) \geq 0 \text{ always.} \end{aligned}$$

Divide the system into $N = 5$ subsystems:

$$\begin{aligned} &\lambda_1 = (x_1, \dots, x_{k_1}), \quad \lambda_2 = (x_{k_1+1}, \dots, x_{k_1+k_2}), \ \dots, \ \lambda_5 = (x_{k_1+k_2+k_3+k_4+1}, \dots, x_K) ; \\ &H_1(\lambda_1) = h_1(x_1) + h_2(x_1, x_2) + h_1(x_2) + \dots + h_2(x_{k_1-1}, x_{k_1}) + h_1(x_{k_1}) \text{ and so on;} \\ &H_{1,2}(\lambda_1, \lambda_2) = h_2(x_{k_1}, x_{k_1+1}) \text{ and so on.} \end{aligned}$$

Denote by $Z(k)$ the partition function for a k-element subsystem:

$$Z(k) = Z^k(1) \int \exp\left(-\beta \sum_{i=1}^{k-1} h_2(x_i, x_{i+1})\right) \mu(dx_1) \dots \mu(dx_k) .$$

Now Z_n introduced before appear to be

$$Z_1 = Z(k_1) , \dots, \ Z_5 = Z(k_5) ,$$

$k_1 + \dots + k_5 = K$. For the above two-block factor (A_1, A_2, A_3, A_4) we have

$$\langle A_1 A_2 A_3 A_4 \rangle = \frac{Z(k_1 + k_2 + k_3 + k_4 + k_5)}{Z(k_1) Z(k_2) Z(k_3) Z(k_4) Z(k_5)} ,$$

and similarly for other connected products; for example,

$$\langle A_2 \rangle = \frac{Z(k_2 + k_3)}{Z(k_2) Z(k_3)} ; \quad \langle A_2 A_3 \rangle = \frac{Z(k_2 + k_3 + k_4)}{Z(k_2) Z(k_3) Z(k_4)} .$$

Substituting them into Eq. (7.4–7.5) we obtain Bell-type inequalities for partition functions! It is natural to suppose that these are irrelevant consequences of some relevant, hopefully simpler inequalities. It should be noted, however, that the inequalities obtained are the best possible for the general inhomogeneous framework given by Eqs. (9.1)–(9.2). This statement follows from the following two facts. First, inequalities (7.5) are the best possible for two-valued two-block factors, as was shown in Sect. 8. Second, any two-block factor valued in $\{0,1\}$, corresponds (exactly or as a limit) to a system described by Eqs. (9.1)–(9.2), as will be shown now. Let $A_n = f_n(\lambda_{n-1}, \lambda_n) \in \{0,1\}$ with independent λ_n, as in Eq. (7.2). Each λ_n runs over a probability space which may be chosen as $[0,1]$ with Lebesgue measure, or equally well, may be identified with the phase space of a physical system, equipped with its Gibbs measure μ_n, as in Eq. (9.3). Introducing an interaction as

$$H_{n,n+1}(\lambda_n, \lambda_{n+1}) = c(1 - f_{n+1}(\lambda_n, \lambda_{n+1}))$$

with a constant c, we obtain

$$\lim_{c\to+\infty} \int \exp\left(-\beta \sum_{n=1}^{N-1} H_{n,n+1}(\lambda_n, \lambda_{n+1})\right) \mu_1(d\lambda_1)\dots\mu_N(d\lambda_N)$$
$$= \int_{\{f_{n+1}(\lambda_n,\lambda_{n+1})=1 \text{ for all } n\}} \mu_1(d\lambda_1)\dots\mu_N(d\lambda_N) = \langle A_1 \dots A_N\rangle \,.$$

The above connection between partition functions and two-block factors remains valid for the homogeneous case. Hence, any inequality for stationary two-valued two-block factors gives an inequality for homogeneous partition functions. However, the inequalities for stationary two-block factors, obtained in Ref. [4], concern only the special case $\langle A_1A_2A_3\rangle = 0$, which is uninteresting for statistical physics. The inequalities obtained in Ref. [6] are non-applicable here, since they require at least five different values for each A_k. The inequalities obtained in Ref. [5] can be applied, giving

$$Z(3k) \le Z^{3/2}(2k)\,, \quad \text{when } Z(2k) \ge \frac{1}{2}Z^2(k)\,,$$

and

$$Z(3k) \le (Z^2(k) - Z(2k))^{3/2} + 2Z(k)Z(2k) - Z^3(k)\,, \quad \text{when } Z(2k) \le \frac{1}{2}Z^2(k)\,.$$

But the inequalities of Ref. [5] do not distinguish two-block factors from one-dependent processes, so, they are not Bell-type inequalities.

10 Independence, free will, conspiracy, and all that

It was not evident from the very beginning, but now it is understood [9, 10, 11], that no experimental test of a Bell-type inequality can be interpreted without assuming some independencies. The traditional framework assumes that the choice of k, l is not correlated with the state of the system. Moreover, no idea of a testable local causality is possible without assuming the statistical independence (or maybe a weak dependence) of observers from observed systems before observations. See Refs. [9, 10, 11] for discussions involving free will, conspiracy, and all that.

The important but implicit premise of independence becomes explicit in the proposed new framework. Figs. 2(b) and 4 show clearly the two kinds of assumptions: on dynamical laws (no faster-than-light propagation) and on initial conditions (no initial correlations). The former is directly connected to the causal structure of space-time. The latter is much more vague. Can it be grounded on the causal structure of the space-time, too? An attempt was made in Ref. [11]: two telescopes pointing at opposite sides of the sky were used as sources of independent random events. If we believe that the sky contains no mirrors or other optical devices at least up to the third minute after the Big Bang, is it enough for becoming free of statistical physics in the argumentation? The answer depends on the accepted cosmology. A Friedman-like scenario leads to non-intersecting past cones, while an inflation scenario does not.

11 Acknowledgment

This research was supported by the Rashi Foundation.

References

[1] Bell, J.S. 1964; Physics (N.Y.) **1**, 195.

[2] Clauser, J.F., Horne, M.A., Shimony, A., and Holt, R.A. 1969; Phys. Rev. Lett. **23**, 880.

[3] Cirel'son (=Tsirelson), B.S. 1980; Lett. Math. Phys. **4**, 93.

[4] Aaronson, J., Gilat, D., Keane, M., and de Valk, V. 1989; Ann. Probab. **17**, 128.

[5] Gandolfi, A., Keane, M., and de Valk, V. 1989; Probab. Th. Rel. Fields **80**, 475.

[6] Matúš, F. 1993; *On two-block-factor sequences and one-dependence,* J. Theor. Probab., to be published.

[7] Froissart, M. 1981; Nuovo Cimento B **64**, 241.
Mermin, N.D., and Schwarz, G.M. 1982; Found. Phys. **12**, 101.
Fine, A. 1982; Phys. Rev. Lett. **48**, 291.

[8] Tsirelson, B.S. 1987; J. of Soviet Math. **36**, 557.

[9] Bell, J.S., Shimony, A., Horne, M.A., and Clauser, J.F. 1985; Dialectica **39**, 85 (republished from Epistem. Letters 1976–78).

[10] Peres, A., and Zurek, W.H. 1982; Am. J. Phys. **50**, 807.

[11] Peres, A. 1986; Found. Phys. **16**, 573.