Gull's theorem revisited

R. D. Gill[*] and D. Karakozak
Mathematical Institute, Leiden University

3 December, 2020
Version 3

Abstract

Steve Gull, in unpublished work [14] available on his Cambridge University homepage, has outlined a proof of Bell's theorem using Fourier theory. Gull's philosophy is that Bell's theorem can be seen as a no-go theorem for a project in distributed computing (with classical, not quantum, computers!). We present his argument, correcting misprints and filling gaps. In his argument, there were two completely separated computers in the network. We need three in order to fill all the gaps in his proof: a third computer supplies a stream of random numbers to the two computers, which represent the two measurement stations in Bell's work. At the end of the day, one can imagine that computer replaced by a cloned, virtual computer, generating the same pseudo-random numbers within each of Alice and Bob's computers. Gull's proof then just needs a third step: writing an expectation as the expectation of a conditional expectation, given the hidden variables.

1 Introduction

This paper is about a simple programming challenge. Simple to state, that is. Can you set up two computers which will each separately receive a stream of angles and each output a stream of completely random plus or minus one's, such that the correlation between the outputs, given the inputs, is minus the cosine of the difference between the angles? The answer is "no". Many correlation functions can be created, since after all the two computers can have an identical pseudorandom generator built in (with the same seed and the same parameters) and hence can create correlated pseudo-randomness. But they cannot create the negative cosine. That is seen as a unique signature of quantum entanglement.

---

[*]corresponding author; `gill@math.leidenuniv.nl`

I give a novel probabilistic proof of this result, using also some basic results from Fourier analysis. Do not fear, no quantum mechanics is involved.

The problem does come from a fundamental question in quantum physics, going back to the debates between Einstein and Bohr, the EPR paradox, and Bell's (1964) theorem [2]. It is also very relevant to recent experimental progress such as the "Delft Bell experiment" of Hensen et al. (2015a, b) reported in *Nature*, [15, 16].

Gull (2016) [14], in overhead sheets which he used in several talks, outlined a novel connection to Fourier analysis. Much of the text in the first subsection below is also part of the paper Gill (2020d) [13]. In that paper, Gull's theorem is of side interest, and is mentioned, but not proven. That led to the necessity to clarify the status of Gull's proof.

1.1   Bell's theorem as a theorem of distributed computing

We would like to paraphrase what is now called Bell's theorem as the metaphysical statement that quantum mechanics (QM) is incompatible with local realism (LR). More precisely, and following B. Tsirelson's wonderful `Citizendium.org` article [22], Bell's theorem states that conventional quantum mechanics is a mathematical structure incompatible with the conjunction of three mathematical properties: *relativistic local causality* (commonly abbreviated to "locality"), *counterfactual definiteness* ("realism") and *no-conspiracy* ("freedom"). By conventional quantum mechanics, we mean: quantum mechanics including the Born rule, but with a minimum of further interpretational baggage. Whether the physicist likes to think of probabilities in a Bayesian or in a frequentist sense is up to them. In Many Worlds interpretations (and some other approaches), the Born rule is argued to follow from the deterministic (unitary evolution) part of the theory. But anyway, everyone agrees that it is there.

Bell himself used the phrase "my theorem" to refer to his *inequalities*: first his (Bell, 1964) three correlations inequality [2], and later what is now called the Bell-CHSH (Clauser, Horne, Shimony, Holt 1969) four correlations inequality [5]. We would rather see those inequalities as simple probabilistic *lemmas* used in two alternative, conventional, proofs of the same *theorem* (the incompatibility of QM and LR). Actually the original proof involves four correlations too, since it also builds on perfect anti-correlation in the singlet correlations, when equal settings are used. Moreover, [2] also has a small section in which Bell shows that his conclusions still hold if all those correlations are only approximately true.

The CHSH inequality is, indeed, a very simple result in elementary probability theory. It is a direct consequence of the Boole inequality stating that the probability of the disjunction (the logical "or") of several propositions is bounded by the sum of their individual probabilities. Something a whole lot stronger is

Fine's (1982) theorem [7], that the satisfaction of all eight one-sided CHSH inequalities together with the four no-signalling equalities is necessary and sufficient for a local hidden variables theory to explain the sixteen conditional probabilities $p(x, y \mid a, b)$ of pairs of binary outcomes x, y, given pairs of binary settings a, b, in a Bell-CHSH-type experiment. No-signalling is the statement that Alice can't see from her statistics, what Bob is doing: with typical statisticians' and physicists' abuse of notation[1], $p(x \mid a, b)$ doesn't depend on b, and similarly for Bob, $p(y \mid a, b)$ doesn't depend on a.

The following remarks form a digression into the prehistory of Bell inequalities. With quite some imagination, a version of Fine's theorem can be recognised in Boole's (1853) monumental work [4], namely his book "*An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*". See the end of Section 11 of Chapter XIX of [4], where the general method of that chapter is applied to Example 7, Case III of Chapter XVIII. Boole derives the conditions on three probabilities p, q and r of three events which must hold in order that a probability space exists on which those three events can be defined with precisely those three probabilities, given certain logical relations between those three events. In modern terminology let the events be $A = \{X = Y\}$, $B = \{Y = Z\}$, $C = \{Z = X\}$ where X, Y and Z are three random variables. It's impossible that two of the events are true but the third is false. Boole derives the 6 linear inequalities involving p, q and r whose simultaneous validity is necessary and sufficient for such a probability space to exist. Boole's proof uses the arithmetic of Boolean logic, taking expectations of elementary linear identities concerning indicator functions (0-1 valued random variables), just as Bell did.

A similar exploration of necessary and sufficient conditions for a probability space to exist supporting a collection of probability assignments to various composite events was given by Vorob'ev (1982) [23]. Vorob'ev starts his impressive but nowadays pretty much forgotten paper with a little three variable example. He has no further concrete examples, just general theorems; no reference to Bell or to Boole. The fact that Bell's inequalities have a precursor in Boole's work was first mentioned (though without giving a precise reference) by Itamar Pitowsky in a number of publications in the early 1980's. This led several later authors to accuse Bell of carelessness and even to suggest plagiarism because Bell does not refer to Boole. At least one can't blame Bell (1964) for not citing Vorob'ev's paper. All that this really shows is that Bell's theorem elicits very strong emotions, both positive and negative. And that lots of physicists and even mathematicians don't know much probability theory.

---

[1]The abuse of notation is that which actual function you are talking about depends on the names you use for the variables on which it depends. This abuse is nowadays reinforced by being built into some computer languages

Having cleared up terminology and attribution, we continue towards the contribution of Gull to the story.

In a Bell-CHSH type experiment we have two locations or labs, in which two experimenters Alice and Bob can each input a binary setting to a device, which then generates a binary output. (The setting corresponds to an intended choice of an angle, but only two angles are considered in each wing of the experiment). This is repeated, say N times. We will talk about a *run of N trials*. The settings are externally generated, perhaps by tossing coins or performing some other auxiliary experiment. The N trials of Alice and Bob are somehow synchronised; exactly how, does not matter for the purposes of this paper, but in real experiments the synchronisation is taken care of using clocks, and the spatial-temporal arrangement of the two labs is such that there is no way a signal carrying Alice's nth setting, sent just before it is inserted into her device, could reach Bob's lab before his device has generated its nth outcome, even if it were transmitted at the speed of light.

In actual fact, these experiments involve measurements of the "spin" of "quantum spin-half particles" (electrons, for instance); or alternatively, measurements of the polarization of photons in the plane opposite to their directions of travel. The two settings, both of Alice and Bob, correspond to what are intended to be two *directions* (spin) or *orientations* (polarization), usually in the plane, but conceivably in three-dimensional space. (Polarization doesn't only have an orientation – horizontal or vertical with respect to any direction in the plane – but also an ellipticity, which can moreover be clockwise or anti-clockwise). Talking about spin: in the so-called singlet state of two maximally entangled spin-half quantum systems, one can conceivably measure each subsystem in any 3D direction whatsoever, and the resulting pair of ± 1-valued outcomes (X_a, Y_b) would have the "correlation" $\mathbb{E}X_a Y_b = -a \cdot b$. Marginally, they would be completely random, $\mathbb{E}X_a = \mathbb{E}Y_b = 0$. The quantum physics set-up is often called an EPR-B experiment: the Einstein, Podolsky, Rosen (1935) thought experiment, transferred to spin by Bohm and Aharonov (1957).

When translated to the polarization example, this joint probability distribution of two binary variables is often called Malus' law. We will call it "the singlet correlations" and stick to the spin-half example. Moreover, we will restrict attention to spin measured in directions in the plane. The archetypical example (though itself only a thought experiment) of such an experiment would involve two Stern-Gerlach devices and is a basic example in many quantum physics texts. Present day experiments use completely different physical systems. Nowadays, anyone can buy time on a quantum computer "in the cloud" and do the experiment themselves, on a two-qubit quantum computer. One can perhaps also look forward to a future quantum internet, connecting two one-qubit quantum computers.

Now, Bell was actually interested in what one nowadays calls (possibly "stochastic") "local hidden variables theories" (LHV). According to such a theory, the statistics predicted by quantum mechanics, and observed in experiments, are merely the reflection of a more classical underlying theory of an essentially deterministic and local nature. There might be local randomness, i.e., by definition, randomness inside the internal structure of invididual particles, completely independent of the local randomness elsewhere. Think of "pseudo-random" processes going on inside particles. Think of modelling the whole of nature as a huge stochastic cellular automaton. Mathematically, one might characterise such theories as claiming the mathematical existence of a classical probability space on which are defined a large collection of random variables X_a and Y_b for all directions a and b in the plane, such that each pair (X_a, Y_b) has got the previously described joint probability distribution. So the question addressed by this paper is: can such a probability space exist? The answer is well-known to be "no", and that is, for us, Bell's theorem. We are interested here in a different way to prove it, from the usual proofs.

The underlying probability space is usually called Λ instead of Ω, and the elementary outcomes $\lambda \in \Lambda$ stand for the configuration of all the particles involved in the whole combined set-up of a source connected to two distant detectors, which are fed the settings a and b from outside. Thus $X_a(\lambda)$ stands *within the mathematical model* for the outcome which Alice would theoretically see if she used the setting a, even if she actually used another, or even if the whole system was destroyed before the outcomes of the actually chosen settings were registered. There is no claim that these variables exist in reality, whatever that means. We are talking about the mathematical existence of a model with certain mathematical properties.

In a sequence of trials, one would suppose that for each trial there is some kind of resetting of apparatus, so that at the nth trial we see the outcomes corresponding to $\lambda = \lambda_n$, where the sequence $\lambda_1, \lambda_2, \ldots$, are independent draws from the same probability measure on the same probability space Λ. Now suppose we could come up with such a theory, and indeed come up with a (classical) Monte-Carlo computer simulation of that theory, on a classical PC. Then we could do the following. Simulate N outcomes of the hidden variable λ, and simply write them into two computer programs as N constants defined in the preamble to the programs. More conveniently, if they were simulated by a pseudo random number generator (RNG), then we could write the constants used in the generator, and an initial seed, as just a few constants, and reproduce the RNG itself inside both programs. The programs are to be run on two computers thought of as belonging to Alice and Bob. The two programs are started. They both set up a dialogue (a loop). Initially, n is set to 1. Alice's computer prints the message "Alice, this is trial number $n = \ldots$.

5

Please input an angle." Alice's computer then waits for Alice to type an angle and hit the "enter" key. Bob's computer does exactly the same thing, repeatedly asking Bob for an angle.

If, on her nth trial, Alice submits the angle a, then the program on her computer evaluates and outputs $X_a(\lambda_n) = \pm 1$, increments n by one, and the dialogue is repeated. Alice's computer doesn't need Bob's angle for this – locality! Bob's doesn't need Alice's.

We have argued that if we could implement a local hidden variables theory in *one* computer program, then we could simulate the singlet correlations derived from one run of many trials on *two completely separate* computers, each running its own program, and each receiving its own stream of inputs (settings) and generating its own stream of outputs.

To summarize: for us, a local hidden variables theory, even a stochastic local hidden variables theory, is just a pair of functions $A(a, \lambda)$, $B(b, \lambda)$ taking values ± 1, where a and b are directions in the plane, together with a probability distribution over the third variable λ, which can lie in any space of any complexity whatever, and which represents everything which determines the final measured outcomes throughout the whole combined system of a source, transmission lines, and two measurement stations, including (local, pseudo-) randomness there. In the theory, the outputs are a deterministic function of the inputs. When implemented as two computer programs, even more is true: *given the programs*, the nth output of either computer depends *only* on n and on the nth input setting angle on that computer; not on the earlier input angles. If one reruns either program with an identical input stream, the output stream will be the same, too. In order to get different outputs, one would have to change the constants defined in the program. As we mentioned before, but want to emphasize again, we envisage that an identical stream of instances of the hidden variable λ is generated on both computers by the same RNG initialised by the same seed; that seed is a constant fixed in the start of the programs.

The previous discussion was quiet lengthy, but is needed to clear up questions about Gull's assumptions. Actually, quite a lot of careful thought lies behind them.

Gull [14] posed the problem: write those computer programs ... or prove that they do not exist. He gave a sketch of a rather pretty proof that such programs could not exist using Fourier theory, and that is what we will turn to next. We will call the statement that such programs do not exist *Gull's theorem.*

We will go through Gull's outline proof, but will run into a difficulty at the last step. However it can be fixed. In Gull's argument, there are, after some preparations, two completely separated computers running completely deterministically. We need three, networked computers: a third computer supplies a stream

of random numbers to the two computers which represent the two measurement stations in Bell's theorem. At the end of the day, one can imagine that computer replaced by a cloned, virtual computer, generating the same pseudo-random numbers within each of Alice and Bob's computers. Gull's proof then just needs a third step: writing a grand expectation over all randomness as the expectation of a conditional expectation, given the hidden variables.

A recently posted `stackexchange` discussion [1]) also attempts to decode Gull's proof, but in our opinion is also incomplete, getting stuck at the same point as we did.

In a final section we will show how *Gull's theorem* (with completely separated, deterministically operating computers) can also be proven using a Bell theorem proof variant due to the first author of the present paper, designed specifically for fighting Bell deniers by challenging them to implement their theory as a networked computer experiment. The trick is to use externally created streams of random binary setting choices and derive martingale properties of a suitable game score, treating the physics implemented inside the computers as completely deterministic. Randomness resides only in the streams of settings used, trial by trial, in the experiment.

The important message of this paper is that *Gull's theorem* is true; it is open to several interpretations, needing different proofs, of course; a version certainly can be proven using Fourier theory.

We have referred to the question of computer simulation of Bell experiments, and to Gull's proof of Bell's theorem, in recent papers Gill (2020a, 2020b, 2020c). It is good that any doubts as to the validity of Gull's claim can be dispelled.

2 Gull's theorem: a distributed computer simulation of the singlet correlations is impossible

The reader needs to recall the standard theory of Fourier series, whereby one approximates a square integrable complex valued function on the circle by an infinite sum of complex Fourier coefficients times periodic exponential functions. A useful introductory resource (freely available on the internet) is the book by Schoenstadt (1992) [20]. In particular one needs Section 2.6, The Complex Form of the Fourier Series; Section 6.2, Convolution and Fourier Transforms; and Section 6.9, Correlation. However, the book does not go into measure theoretic niceties. At the very least, an assumption of measurability needs to be made, in order that integrals are well defined (or instead of Lebesgue integrals, one has to use the notion of Radon integrals). The Fourier series transform converts a square integrable complex valued function to a square summable doubly infinite sequence of coefficients. With

appropriate normalisations, one has an isomorphism between the Hilbert spaces $L^2_{\mathbb{C}}(-\pi, \pi)$ and $\ell^2_{\mathbb{C}}(\mathbb{Z})$. Said in another way, the periodic complex exponential functions on the circle, appropriately normalised, form a complete orthonormal basis of the function space of equivalence classes of all square integrable functions on the circle.

The transformation *in each direction* is called by Gull a Fourier transform, and in an abstract sense, this is perfectly appropriate terminology. The fact that a convolution of functions maps to a product of Fourier series is very well known. Less well known is that a similar result holds for the correlation between two functions.

The literature from analysis is full of results giving stronger forms of convergence of Fourier series to the approximated function than the L^2 convergence coming from the point of view just explained. It seems to us that that is where one must look in order to complete a proof. Moreover, as we mentioned, measurability has to be an added assumption.

One could avoid all analytic niceties by insisting that the experimenters only use settings of, say, whole numbers of degrees. Of course, when one selects a whole number of degrees on a machine with a digital interface, what goes on inside the machine will likely involve angles which are never precisely the angles which the user asked for. But that doesn't matter. We are talking about thought experiments and about mathematical models. If we only look at whole numbers of degrees, we could try to rewrite Gull's proof using discrete Fourier transforms. One could say that this is a Bell-CHSH experiment in which Alice and Bob each have 360 different settings at their disposal.

We will not develop that approach in this short paper, but do plan to explore it in the near future. We will just present Gull's argument as it stands, and see where the argument needs fixing.

According to the singlet correlations, if Alice and Bob submit the same sequence of inputs, then their output streams are exactly opposite to one another. Recall that the outcomes are functions only of the trial number and the respective setting angles. Let us denote them by $A_n(\theta)$ and $B_n(\phi)$. Then we must have the following equality between functions, defined, say, on the interval $(-\pi, \pi]$: $B_n = -A_n$.

Next, consider a sequence of N trials in which Alice uses angles, picked uniformly at random from that interval, and Bob uses the same sequence of angles, shifted clockwise by the amount θ. The expectation value of the average of the products of their outcomes is

$$\rho_N(\theta) \;=\; -\frac{1}{N}\sum_{n=1}^{N} \frac{1}{2\pi} \int\limits_{u \in (-\pi, \pi]} A_n(u) A_n(u - \theta)\, \mathrm{d}u. \tag{1}$$

We want this "expected sample correlation function" ρ_N to converge (pointwise)

8

to the negative cosine, the correlation function ρ defined by

$$\rho(\theta) \ := \ -\cos(\theta) \ = \ -\tfrac{1}{2}e^{i\theta} - \tfrac{1}{2}e^{-i\theta} \tag{2}$$

as N converges to infinity.

Notice, we already took the expectation value with respect to the randomness of the sequence of setting pairs. We actually want convergence almost surely: in a single infinitely long experiment we already want to see the negative cosine law. For this we could use the strong law of large numbers in the case of non identically distributed, but bounded, random variables. That would tell us that if ρ_N converges to a limit, then the "measured correlation" will converge also, with probability one. Gull does not enter into any of these subtleties. Moreover, at some point, he drops the index n from his notation.

Suppose A_n has Fourier series $(\widetilde{A}_n(k) : k \in \mathbb{Z})$, i.e., with

$$\widetilde{A}_n(k) \ = \ \frac{1}{2\pi} \int\limits_{u \in (-\pi, \pi]} e^{-iku} A_n(u) \, \mathrm{d}u, \tag{3}$$

$$A_n(u) \ = \ \sum_{k \in \mathbb{Z}} \widetilde{A}_n(k) e^{iku}. \tag{4}$$

This requires some regularity of each function A_n. But assuming (4), we can substitute it into (1), twice, giving us a triple summation over n, k, and k' (say), with coefficient $\widetilde{A}_n(k)\widetilde{A}_n(k')$, and a single integration over u of $e^{-iku}e^{-ik'(u-\theta)}/2\pi$. The integral over the circle of $e^{-(k+k')u}/2\pi$ is zero unless $k+k' = 0$, when it equals 1. We obtain the summation over n and k of $\widetilde{A}_n(k)\widetilde{A}_n(-k)e^{-ik\theta}$. From (3), because the function A_n is real, we obtain

$$\rho_N(\theta) \ = \ -\frac{1}{N} \sum_{n=1}^{N} \sum_{k \in \mathbb{Z}} |\widetilde{A}_n(k)|^2 e^{-ik\theta}.$$

Now, we already saw the Fourier series for the negative cosine: it just has two nonzero coefficients, obviously; those for $k = \pm 1$. This tells us that for all $k \neq \pm 1$, the limit as N tends to infinity of $\sum_{n=1}^{N} |\widetilde{A}_n(k)|^2/N$ is zero. We know that the real functions A_n square to $+1$. Yet only the $k = \pm 1$ terms in their Fourier expansions play asymptotically a significant role in their expansions. Yet on average, and in L^2 sense, they approach the negative cosine. Intuitively, this must be impossible. But we do not presently have a proof.

The problem is all those different functions A_n. Gull in his notes is clearly thinking of there just being one. A possible solution is to introduce a third computer; one could consider it to represent the source. It supplies a stream of uniform

random numbers to computers A and B, and now, inside the dialogue loop, the *same* function is called at each iteration, whose inputs consist *only* of the nth angle and the nth random number. Denote a generic random number by $\lambda \in \Lambda$ with probability distribution $\mathbb{P}$. Fix θ and choose, uniformly at random, N setting pairs, an angle θ apart. We can use equation (1) again in which, by definition now,

$$A_n(u) = A(u, \lambda_n). \tag{5}$$

Equation (1) now represents the the sample average of N independent, identically distributed random variables. We have taken the conditional expectation, given the hidden variables $\lambda_1, \ldots, \lambda_n$ of the correlation observed in a run of length N; we have averaged only over the setting choices. The expectation value of this, thus averaging both over hidden variables and over setting pairs at a fixed distance θ apart, then becomes

$$\frac{1}{2\pi} \int\limits_{\lambda \in \Lambda} \int\limits_{u \in (-\pi, \pi]} A(u, \lambda) A(u - \theta, \lambda) \, du \, d\mathbb{P}(\lambda). \tag{6}$$

Substitute (4) into (6), twice, giving us again summation over k, and k' (say), with coefficient $\widetilde{A}(k, \lambda) \widetilde{A}(k', \lambda)$ and a double integral over u of $e^{-iku} e^{-ik'(u-\theta)}/2\pi$, and over λ with respect to the probability measure $\mathbb{P}$. Now, the integral of $e^{-(k+k')u}$ is zero unless $k + k' = 0$. So we obtain

$$\rho(\theta) = \sum_k \mathbb{E}(|A(k)|^2) e^{-ik\theta}. \tag{7}$$

From (2), $\mathbb{E}(|A(k)|^2) = 0$ for all k except $k = \pm 1$. The random measurement functions only have two non-zero Fourier series coefficients, and cannot possibly be discontinuous, which they must be since they only take the values ± 1.

A similar project was independently embarked on by H. Razmi (2005, 2007). His work was criticised by R. Tumulka (2009), but Razmi (2009) fought back. Razmi makes a further physical assumption of invariance under rotations. Gill (2020c) [12] discusses how some invariances may be imposed "for free" since at the end of the day we are going to generate correlations which are invariant under the same rotation of both settings.

2.1 A classical distributed computer simulation of the singlet correlations is impossible

Quite a few years before the work of Gull, Razmi and Tumulka, and deliberately focussed on classical computer simulations of CHSH-style experiments, Gill (2003a, 2003b) produced several martingale based inequalities on a variant of the CHSH

statistic. He noticed that if the denominators of the four numerators in the four sample correlations were replaced by their expectation values (when settings are chosen completely at random) $N/4$, and then the whole statistic is multiplied by $N/4$, then it equals a sum over the N trials of a quantity which, under local realism, and assuming the complete randomness of the *setting* choices only, has expectation value less than $3/4$. So subtract off $(3/4)N$ and one has, under local realism, a supermartingale. The conditional independence of increments of the process (the time variable being N), conditional on the past, are negative. Moreover the increments of the process are bounded, so powerful martingale inequalities give easy exponential inequalities on the probabilities of large deviations upwards.

These inequalities were later improved, and the martingale structure further exploited, to get wonderfully sharp results. In our opinion, the nicest is that presented in the supplementary material of the paper Hensen et al. (2015), published in *Nature*, as the first ever succesful "loophole-free" Bell type experiment. Since those researchers were based in Delft, the Netherlands, and have had much contact with the authors, we (the authors) are particularly proud of that experiment and that probabilistic result.

Consider such an experiment and let us say that the nth trial results in a success, if and only if the two outcomes are equal and the settings are *not* both the setting with label "2", or the two outcomes are opposite and the settings *are* both the setting with label "2". (Recall that measurement outcomes take the values ± 1; the settings will be labelled "1" and "2", and these labels correspond to certain choices of measurement directions, in each of the two wings of the experiment). The quantum engineering is set up so as to ensure a large positive correlation between the outcomes for setting pairs 11, 12 and 21, but a large negative correlation for setting pair 22. Let us denote the total number of successes in a fixed number, N, of trials, by S_N.

Then Hensen et al. (2015a, b), and see also Bierhorst (2015) and Elkouss and Wehner (2016) for further generalisations, show that, for all x,

$$\mathbb{P}(S_N \geq x) \ \leq \ \mathbb{P}(\text{Bin}(N, 3/4) \geq x), \tag{8}$$

where $\text{Bin}(N, p)$ denotes a binomally distributed random variable with parameters N, the number of trials, and success probability, per independent trial, p.

Above we wrote "under the assumption of local realism". Those are physics concepts. The important point here is that a network of two classical PC's both performing a completely deterministic computation, and allowed to communicate over a classical wired connection *between* every trial and the next, does satisfy those assumptions. The theorem applies to a classical distributed computer simulation of the usual quantum optics lab experiment. Time trends and time jumps in the simulated physics, and correlations (dependency) due to use of memory of past

settings (even of the past settings in the other wing in the experiment) do not destroy the theorem. It is driven *solely* by the random choice anew, trial after trial, of one of the four pairs of settings, and such that each computer is only fed its own setting, not that given to the other computer.

Take for instance $N = 10\,000$. Take a critical level of $x = 0.8N$. Local realism says that S_N is stochastically smaller (in the right tail) than the $\text{Bin}(N, 0.75)$ distribution. According to quantum mechanics, and using the optimal pairs of settings and the optimal quantum state, S_N has approximately the $\text{Bin}(N, 0.85)$ distribution. Under those two distributions, the probabilities of outcomes respectively larger and smaller than $0.80N$ are about 10^{-30} and 10^{-40} respectively. These probabilities give an excellent basis for making bets.

Suppose we use the computer programs in this way just to compute the four correlations between the pairs of settings used in the CHSH inequality. This comes down to looking at the correlation curve in just four points. If the programs yield correlations which converge to some limits as N tends to infinity, then (4) shows that those four limits will violate the CHSH inequality, hence cannot be the four points on the correlation function which the program is supposed to be able to compute.

3 Conclusion

Gull's theorem is true but there are subtleties in the proof which need careful resolution. To follow his proof, we must introduce randomness into the setting choices, and express expectation values of empirical correlations as expectation (over the hidden variables) of conditional expectations, given "hidden variables". One must also put certain (reasonable) restrictions on what are the inputs of each of the programs running on each of the two computers, representing the two measurement stations. The program must represent an implementation of a classical local hidden variables model. There must be no use of memory or time. The two separated computers use the same stream of pseudo-random numbers to model the hidden variables. These restrictions can be avoided by using a proof based on martingale ideas and exploiting the logic behind the CHSH inequality.

A proof via Fourier analysis requires regularity assumptions. Possibly this can be avoided by using discrete Fourier transforms.

References

[1] Anonymous (2020). Question posted on "Stackexchange" by "Eran Medan", reply by "Chiral Anomaly".

https://physics.stackexchange.com/questions/547039/
help-understanding-prof-steve-gulls-explanation-of-bells-theorem

[2] J.S. Bell (1964). On the Einstein Podolsky Rosen paradox. *Physics* **1** (3) 195–200.
https://cds.cern.ch/record/111654/files/vol1p195-200_001.pdf

[3] P. Bierhorst (2015). A robust mathematical model for a loophole-free Clauser–Horne experiment. *J. Phys. A: Math. Theor.* **48** 195302 (19pp).
https://arxiv.org/abs/1312.2999

[4] G. Boole (1853) *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities.* London: Macmillan. https://www.gutenberg.org/files/15114/15114-pdf.pdf

[5] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt (1969). Proposed experiment to test local hidden-variable theories, *Physical Review Letters* **23** (15) 880–884. https://doi.org/10.1103/PhysRevLett.23.880

[6] D. Elkouss and S. Wehner (2016). (Nearly) optimal P values for all Bell inequalities¿. *npj Quantum Information* **2** 16026 (8pp).
https://www.nature.com/articles/npjqi201626

[7] A. Fine (1982). Hidden variables, joint probabilities, and the Bell inequalities. *Physical Review Letters* **48** (5) 291–295. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.48.291

[8] R.D. Gill (2003a). Accardi contra Bell: the impossible coupling. pp. 133–154 in: In M. Moore, C. Leger, and S. Froda (Eds.), *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*, Lecture Notes–Monograph series, Hayward, Ca. Institute of Mathematical Statistics. https://arxiv.org/abs/quant-ph/0110137

[9] R.D. Gill (2003b). Time, Finite Statistics, and Bell's Fifth Position. pp. 179–206 in: *Proc. of "Foundations of Probability and Physics - 2"*, Ser. Math. Modelling in Phys., Engin., and Cogn. Sc., vol. 5/2002, Växjö Univ. Press. https://arxiv.org/abs/quant-ph/0301059

[10] R.D. Gill (2020a). Does Geometric Algebra provide a loophole to Bell's Theorem? (with correction note). https://arxiv.org/abs/1203.1504 *Entropy*, **22** (1), 61 (21 pp.); https://doi.org/10.3390/e22010061

[11] R.D. Gill (2020b) *Comment on "Quantum Correlations Are Weaved by the Spinors of the Euclidean Primitives".*
https://www.math.leidenuniv.nl/~gill/Author_tex-v2.pdf

[12] R.D. Gill (2020c). The Triangle Wave Versus the Cosine: How Classical Systems Can Optimally Approximate EPR-B Correlations. *Entropy* **22** (3) 287 (10 pp). `https://doi.org/10.3390/e22030287`

[13] R.D. Gill (2020d). Comment on "Dr. Bertlmann's Socks in a Quaternionic World of Ambidextral Reality". Submitted to *IEEE Access.* `https://arxiv.org/abs/2001.11338`

[14] S. Gull (2016). *Quantum acausality and Bell's theorem.* Four pages pdf of overhead slides. `http://www.mrao.cam.ac.uk/~steve/maxent2009/images/bell.pdf`. From SteveGull's personal webpage with several items of work from past conferences. `http://www.mrao.cam.ac.uk/~steve/maxent2009/`

[15] B. Hensen, H. Bernien, A. Dréau, A. et al. (2015a). Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **52** (6) 682–686 (2015). `https://doi.org/10.1038/nature15759`. See also the authors report of a second experiment, `https://arxiv.org/abs/1603.05705,`

[16] B. Hensen, H. Bernien, A. Dréau, A. et al. (2015b). Supplementary information. (26 pp) `https://static-content.springer.com/esm/art%3A10.1038%2Fnature15759/MediaObjects/41586_2015_BFnature15759_MOESM113_ESM.pdf`

[17] H. Razmi (2005) Is the Clauser–Horne model of Bell's theorem completely stochastic? *Journal of Physics A* **38**, 3661–3664. `https://iopscience.iop.org/article/10.1088/0305-4470/38/16/013`

[18] H. Razmi (2007). A New Proof of Bell's Theorem Based on Fourier Series Analysis. *Annales de la Fondation Louis de Broglie* **32** (1) 69–76. `https://aflb.minesparis.psl.eu/AFLB-321/aflb321m555.pdf`

[19] H. Razmi (2009). Reply to "Comment on A New Proof of Bell's Theorem Based on Fourier Series Analysis" by Roderich Tumulka, *Annales de la Fondation Louis de Broglie* **34** (1) 43–44. `https://aflb.minesparis.psl.eu/AFLB-341/aflb341m641b.pdf`

[20] A.L. Schoenstadt (1992). *An Introduction to Fourier Analysis. Fourier Series, Partial Differential Equations and Fourier Transforms.* Naval Postgraduate School, Monterey, California. `https://faculty.nps.edu/bneta/soln3139.pdf`

[21] R. Tumulka (2009). Comment on "A New Proof of Bell's Theorem Based on Fourier Series Analysis". *Annales de la Fondation Louis de Broglie* **34** (1) 39–41. `https://aflb.minesparis.psl.eu/AFLB-341/aflb341m641.pdf`

[22] B.Tsirelson (2010). Entanglement (physics). *Knowino* internet encyclopedia, originally *Citizendium*. Last edited 2010.
`https://www.tau.ac.il/~tsirel/dump/Static/knowino.org/wiki/`
`Entanglement_(physics).html`

[23] N.N. Vorob'ev, N. N. (1962). Consistent families of measures and their extension. *Theory Probab. Appl.* **7** 147–163.

4   Acknowledgements